Gregor Kemper

Lineare Algebra für Physiker*

Vorlesungsmanuskript†
Technische Universität München

7. Oktober 2022

*Wesentliche Teile dieses Skripts sind in dem Buch "**Lineare Algebra - mit einer Einführung in diskrete Mathematik und Mengenlehre**" enthalten, das bei **Springer Spektrum** erschienen ist. Siehe https://link.springer.com/book/10.1007/978-3-662-63724-1

†Verbesserungsvorschläge und Fehlermeldungen bitte an: kemper@ma.tum.de.

Inhaltsverzeichnis

Kapitel 1
Matrizen und Vektoren

In diesem Kapitel ist K immer ein Körper (also ein Rechenbereich, der die vier Grundrechenarten zulässt, z.B. $K = \mathbb{R}$, $\mathbb{C}$, $\mathbb{Q} \dots$).

Wir führen Matrizen als grundlegenden „Datentyp“ in der linearen Algebra ein.

Definition 1.1 (Matrizen)**.** *Es seien $m, n \in \mathbb{N}_{>0}$ positive natürliche Zahlen. Eine $m \times n$-**Matrix** ist eine „rechteckige Anordnung“*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

mit $a_{i,j} \in K$. Formaler können wir eine $m \times n$-Matrix definieren als eine Abbildung $\{1, \dots, m\} \times \{1, \dots, n\} \to K$ vom kartesischen Produkt der Mengen $\{1, \dots, m\}$ und $\{1, \dots, n\}$ in K, die jedem Paar (i, j) das Körperelement $a_{i,j}$ zuordnet. Mit dieser Definition wird die informelle Definition einer Matrix als rechteckige Anordnung zu einer bloßen Schreibweise „degradiert“.

Das Element $a_{i,j}$ einer Matrix A heißt der (i, j)-te **Eintrag** *von A. Wir benutzen verschiedene Schreibweisen für Matrizen:*

$$A = (a_{i,j})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = (a_{i,j})_{\substack{1 \le i \le m \\ 1 \le j \le n}} = (a_{i,j})_{i,j} = (a_{i,j}),$$

wobei die beiden letzten benutzt werden, wenn m und n aus dem Kontext klar sind. Durch die Definition einer Matrix ergibt sich automatisch der Gleichheitsbegriff von Matrizen: Zwei $m \times n$-Matrizen $A = (a_{i,j})$ und $B = (b_{i,j})$ sind gleich, falls $a_{i,j} = b_{i,j}$ für alle i und j gilt.

Die Menge aller $m \times n$-Matrizen wird mit $K^{m \times n}$ bezeichnet.

Eine $1 \times n$-Matrix $(a_1, \dots, a_n) \in K^{1\times n}$ wird als **Zeilenvektor***, eine $m \times 1$-Matrix $\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \in K^{m\times 1}$ als* **Spaltenvektor** *bezeichnet. Wir schreiben $K^m := K^{m\times 1}$ und nennen dies den* **m-dimensionalen Standardraum***. Die Benennung wird später klar werden, und auch der Grund, weshalb hierbei den Spaltenvektoren trotz der umständlicheren Schreibweise der Vorzug gegeben wird.*

Für $A = (a_{i,j}) \in K^{m\times n}$ und $i \in \{1, \dots, m\}$ ist $(a_{i,1}, \dots, a_{i,n}) \in K^{1\times n}$ die **i-te Zeile** *von A. Für $j \in \{1, \dots, n\}$ ist $\begin{pmatrix} a_{1,j} \\ \vdots \\ a_{m,j} \end{pmatrix} \in K^{m\times 1}$ die* **j-te Spalte** *von A.*

Eine Matrix $A \in K^{m\times n}$ mit $m = n$ heißt **quadratisch***. Für $A = (a_{i,j}) \in K^{m\times n}$ ist $A^T := (a_{j,i}) \in K^{n\times m}$ die* **transponierte Matrix***; also z.B.*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Eine quadratische Matrix heißt **symmetrisch***, falls $A^T = A$ gilt.*

Wenn Matrizen und Vektoren im Spiel sind, bezeichnet man Elemente des Körpers K oft als *Skalare*. Obwohl wir von Zeilen- und Spaltenvektoren gesprochen haben, ist die korrekte Antwort auf die Frage „Was ist ein Vektor?": „ein Element eines Vektorraums" (siehe Kapitel 4).

Beispiel 1.2. Die folgenden Beispiele sollen die Relevanz von Matrizen demonstrieren.

(1) $\mathbb{R}^2 = \{\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \mid x, y \in \mathbb{R}\}$ lässt sich als Ebene veranschaulichen.
(2) Wenn $S_1, \dots, S_n$ Städte sind und $d_{i,j}$ die Entfernung zwischen S_i und S_j bezeichnet, dann ist $D = (d_{i,j}) \in \mathbb{R}^{n\times n}$ (unter sinnvollen Annahmen) symmetrisch. D ist die *Distanzmatrix*.
(3) $Z_1, \dots, Z_n$ seien Zustände, in denen sich ein gewisses System befinden kann. $P_{i,j}$ sei die Wahrscheinlichkeit, dass das System von dem Zustand Z_i in den Zustand Z_j übergeht. (Man spricht auch von der *Übergangswahrscheinlichkeit*.) Die $P_{i,j}$ werden zusammengefasst in der quadratischen Matrix
$$T := (P_{i,j}) \in \mathbb{R}^{n\times n},$$
die auch *Übergangs-* oder *Transitionsmatrix* genannt wird. Für jedes $i \in \{1, \dots, n\}$ gilt $P_{i,1} + \dots + P_{i,n} = 1$, d.h. die i-te Zeilensumme ist 1. Allgemein nennt man eine Matrix in $\mathbb{R}^{n\times n}$ (zeilen-)**stochastisch**, falls alle Zeilensummen 1 und alle Einträge nicht-negativ sind. T ist also stochastisch. Falls das System schrittweise seinen Zustand ändert, wie berechnet sich dann die Wahrscheinlichkeit, in k Schritten vom Zustand Z_i nach Z_j

zu gelangen? Diese Frage werden wir in Beispiel 1.5 nach der Definition des Matrixprodukts klären. ◁

Definition 1.3. *Wir definieren nun die Summe und das Produkt von Matrizen.*

(a) *Für $A = (a_{i,j}) \in K^{m\times n}$ und $B = (b_{i,j}) \in K^{m\times n}$ ist die Summe $A \boxplus B \in K^{m\times n}$ definiert durch $A \boxplus B = (c_{i,j})$ mit $c_{i,j} := a_{i,j} + b_{i,j}$. Man spricht auch von* komponentenweiser *Addition. Man kann Matrizen nur dann addieren, wenn ihre Spalten- und Zeilenzahl (m und n) übereinstimmen. Wir haben „$\boxplus$“ für die Unterscheidung von der Addition im Körper K verwendet. Ab jetzt werden wir aber immer $A + B$ schreiben.*

(b) *Für $A = (a_{i,j}) \in K^{m\times n}$ und $B = (b_{i,j}) \in K^{n\times l}$ ist das Produkt $A \boxdot B \in K^{m\times l}$ definiert durch $A \boxdot B = (c_{i,j})$ mit*

$$c_{i,j} := \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

Das Produkt ist also nicht *komponentenweise definiert. Es ist nur definiert, wenn die Spaltenzahl von A mit der Zeilenzahl von B übereinstimmt. Im weiteren werden wir $A \cdot B$ oder AB statt $A \boxdot B$ schreiben. Ein wichtiger Spezialfall ist das Produkt einer Matrix $A = (a_{i,j}) \in K^{m\times n}$ mit einem Spaltenvektor $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$:*

$$A \cdot v = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in K^m \quad \text{mit} \quad y_i = \sum_{j=1}^{n} a_{i,j} x_j.$$

(c) *Wir definieren noch das Produkt einer Matrix mit einem Skalar: Für $A = (a_{i,j}) \in K^{m\times n}$ und $s \in K$ ist das Produkt $s \cdot A$ definiert durch $s \cdot A = (c_{i,j}) \in K^{m\times n}$ mit $c_{i,j} := s \cdot a_{i,j}$. Die Matrix wird also komponentenweise mit s multipliziert.*

Beispiel 1.4. Ein paar Beispiele zum Matrixprodukt:

(1)

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 0 \cdot 1 + 1 \cdot 0 & 1 \cdot 1 + 0 \cdot 2 + 1 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 1 + 2 \cdot 0 & 0 \cdot 1 + 1 \cdot 2 + 2 \cdot 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 4 \end{pmatrix}.$$

(2) Für $x_1, x_2 \in K$ ist

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

Mit $\varphi \in \mathbb{R}$ ist

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos(\varphi)x_1 - \sin(\varphi)x_2 \\ \sin(\varphi)x_1 + \cos(\varphi)x_2 \end{pmatrix},$$

was einer Drehung des Vektors $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ um den Winkel φ bedeutet.

(3) Für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in K^n$ liegt der transponierte Vektor v^T in $K^{1\times n}$, also ist $v^T \cdot w \in K^{1\times 1}$. Es gilt

$$v^T \cdot w = \sum_{i=1}^{n} x_i y_i.$$

Dies werden wir in Kapitel 12 als das (Standard-)Skalarprodukt einführen. ◁

Die Definitionen der Summe von Matrizen und des Produkts einer Matrix und eines Skalars sind nicht weiter spannend. Die Definition des Produktes ist jedoch komplizierter und damit interessanter. Wir fragen uns, warum das Produkt so definiert wurde, und nicht etwa auch komponentenweise. Es gibt hierfür mehrere Gründe, und den wichtigsten werden wir in Kapitel 7 kennenlernen. Ein Hinweis, weshalb die Definition sinnvoll ist, finden sich jedoch schon in folgendem Beispiel.

Beispiel 1.5. Wie in Beispiel 1.2(3) sei $P_{i,j}$ die Wahrscheinlichkeit, dass ein System vom Zustand Z_i in den Zustand Z_j übergeht, und $T = (P_{i,j}) \in \mathbb{R}^{n\times n}$ sei die Transitionsmatrix. Was ist die Wahrscheinlichkeit, in zwei Schritten von Z_i nach Z_j zu gelangen? Hierfür ist jedes Z_k als Zwischenschritt möglich. Wir nehmen an, dass die Schritte unabhängig voneinander sind. Dann ist die Wahrscheinlichkeit, von Z_i über Z_k nach Z_j zu gelangen, gleich $P_{i,k} \cdot P_{k,j}$. Insgesamt ergibt sich also die Wahrscheinlichkeit $\sum_{k=1}^{n} P_{i,k}P_{k,j}$ für den Übergang von Z_i nach Z_j in zwei Schritten. Dies ist genau der (i,j)-te Eintrag des Produkts $T \cdot T = T^2$, also ist T^2 die entsprechende Transitionsmatrix. Weiter ergibt sich die Matrix der Wahrscheinlichkeiten, in drei Schritten von Z_i nach Z_j zu gelangen, als T^3, u.s.w. Hier findet das Produkt bzw. die Potenzen von Transitionsmatrizen also eine natürliche Interpretation. ◁

Satz 1.6. *Für Matrizen gelten die folgenden Regeln.*

(a) Für alle $A = (a_{i,j}), B, C \in K^{m\times n}$ *gelten:*

(1) $(A + B) + C = A + (B + C)$,
(2) $A + B = B + A$,
(3) für $\mathbf{0} := \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \in K^{m\times n}$ *(die* Nullmatrix*) gilt* $A + \mathbf{0} = A$
(4) und für $-A := (-a_{i,j})$ *gilt* $-A + A = \mathbf{0}$.

Mit anderen Worten, $K^{m\times n}$ bildet zusammen mit der Addition eine abelsche Gruppe (siehe Kapitel 3, wo der Gruppenbegriff eingeführt wird).

(b) Für alle $A, B \in K^{m\times n}$ und $s, s' \in K$ gelten:

(1) $s \cdot (A + B) = s \cdot A + s \cdot B$,
(2) $(s + s') \cdot A = s \cdot A + s' \cdot A$,
(3) $s \cdot (s' \cdot A) = (ss') \cdot A$,
(4) $1 \cdot A = A$.

(c) Seien A, B, C Matrizen, so dass jeweils die unten gebildeten Summen und Produkte definiert sind. Dann gelten:

(1) $(A \cdot B) \cdot C = A \cdot (B \cdot C)$,
(2) $A \cdot (B + C) = A \cdot B + A \cdot C$,
(3) $(A + B) \cdot C = A \cdot C + B \cdot C$
(4) und für

$$I_n := \begin{pmatrix} 1 & 0 & & \cdots & 0 \\ 0 & 1 & & & \vdots \\ & & \ddots & & \\ \vdots & & & 1 & 0 \\ 0 & \cdots & & 0 & 1 \end{pmatrix} \in K^{n\times n}$$

(die **Einheitsmatrix***) gelten $I_n \cdot A = A$ und $B \cdot I_n = B$.*

Beweis. Der Beweis lässt sich durch direktes Nachrechnen führen. Nur für den Teil (c1) ist die Rechnung etwas umfangreicher. □

Anmerkung 1.7. (a) Im Allgemeinen ist die Formel $A \cdot B = B \cdot A$ *falsch*! Zum Beispiel gilt

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{aber} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \tag{1.1}$$

Dieses Beispiel kann man auf beliebige $n\times n$-Matrizen ($n \geq 2$) ausdehnen.

(b) Aus Satz 1.6(a) und (c) folgt, dass $K^{n\times n}$ ein Ring mit Eins ist. Das Eins-Element ist I_n. Für $n \geq 2$ ist dieser Ring nicht kommutativ. ◁

Kapitel 2
Lineare Gleichungssysteme

In diesem Kapitel untersuchen wir Gleichungssysteme von der Art

$$\begin{aligned} x_1 \quad &+ 2x_3 + x_4 = -3 \\ 2x_1 \quad &+ 4x_3 - 2x_4 = 2 \\ x_2 \quad & - x_4 = 2 \\ x_1 \quad &+ 2x_3 + 2x_4 = -5. \end{aligned} \tag{2.1}$$

Wir können dies umformulieren in eine einzige Matrixgleichung:

$$A \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 2 \\ 2 \\ -5 \end{pmatrix} \qquad \text{mit} \qquad A = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 2 & 0 & 4 & -2 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 2 & 2 \end{pmatrix}.$$

Auch in diesem Kapitel steht K immer für einen Körper.

Definition 2.1. *Eine Gleichung der Form $A \cdot x = b$ mit $A \in K^{m \times n}$ und $b \in K^m$ heißt ein* **lineares Gleichungssystem** *(kurz: LGS). Die* **Lösungsmenge** *ist die Menge aller $x \in K^n$, die die Gleichung erfüllen. Das LGS heißt* **homogen***, falls $b = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$, sonst* **inhomogen***. Die Matrix A heißt die* **Koeffizientenmatrix** *des linearen Gleichungssystems. Man bildet auch die Matrix $(A|b) \in K^{m \times (n+1)}$, indem man den Vektor b als $(n+1)$-te Spalte an A anheftet. Diese heißt die* **erweiterte Koeffizientenmatrix***, und sie kodiert die gesamte Information des LGS. (Dabei ist die Linie vor der Spalte b nur eine schreibtechnische Hilfe und hat keine mathematische Bedeutung.)*

Unser Ziel ist es, einen Algorithmus zur Bestimmung der Lösungsmenge eines LGS zu entwickeln. Hierfür definieren wir zunächst einige Manipulationen, die auf Matrizen allgemein und im besonderen auf die erweiterte Koeffizientenmatrix eines LGS angewandt werden können. Diese Manipulationen heißen **elementare Zeilenoperationen** und gliedern sich in drei Typen:

Typ I: Vertauschen zweier Zeilen;
Typ II: Multiplizieren einer Zeile mit einem Skalar $s \in K \setminus \{0\}$;
Typ III: Addieren des s-fachen einer Zeile zu einer anderen, wobei $s \in K$.

Es ist unmittelbar klar, dass das Anwenden von elementaren Zeilenoperationen auf die erweiterte Koeffizientenmatrix eines LGS die Lösungsmenge unverändert lässt. Wir können ein LGS also mit diesen Operationen manipulieren mit dem Ziel, es auf eine so einfache Gestalt zu bringen, dass man die Lösungsmenge direkt ablesen kann. Die angestrebte Gestalt ist die *Zeilenstufenform* gemäß der folgenden Definition.

Definition 2.2. *Es sei $A \in K^{m \times n}$. Wir sagen, dass A in* **Zeilenstufenform** *ist, falls gelten:*

(a) Beginnt eine Zeile mit k Nullen, so stehen unter diesen Nullen lauter weitere Nullen.
(b) Unter dem ersten Eintrag $\neq 0$ einer jeden Zeile (falls diese nicht nur aus Nullen besteht) stehen lauter Nullen.

Wir sagen, dass A in **strenger Zeilenstufenform** *ist, falls zusätzlich gilt:*

(c) Über dem ersten Eintrag $\neq 0$ einer jeden Zeile (falls diese nicht nur aus Nullen besteht) stehen lauter Nullen.

Beispiel 2.3. Zur Illustration mögen folgende Beispiele dienen:

(1) Die Matrix $\begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ ist *nicht* in Zeilenstufenform.
(2) Die Matrix $\begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ ist *nicht* in Zeilenstufenform.
(3) Die Matrix $\begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$ ist in Zeilenstufenform, aber nicht in strenger Zeilenstufenform.
(4) Die Matrix $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$ ist in strenger Zeilenstufenform. ◁

Beispiel 2.4. Wir wenden elementare Zeilenoperationen auf die erweiterte Koeffizientenmatrix des LGS (2.1) an mit dem Ziel, die Matrix auf strenge Zeilenstufenform zu bringen.

$$
\begin{aligned}
&\left(\begin{array}{cccc|c} 1&0&2&1&-3\\ 2&0&4&-2&2\\ 0&1&0&-1&2\\ 1&0&2&2&-5 \end{array}\right)\begin{array}{c} \\ -2\\ \\ -1 \end{array} \xrightarrow[\text{Typ III}]{}
\left(\begin{array}{cccc|c} 1&0&2&1&-3\\ 0&0&0&-4&8\\ 0&1&0&-1&2\\ 0&0&0&1&-2 \end{array}\right) \xrightarrow[\text{Typ I}]{}\\[1ex]
&\left(\begin{array}{cccc|c} 1&0&2&1&-3\\ 0&1&0&-1&2\\ 0&0&0&-4&8\\ 0&0&0&1&-2 \end{array}\right)\begin{array}{c} \\ \\ \cdot(-\frac{1}{4})\\ \\ \end{array} \xrightarrow[\text{Typ II}]{}
\left(\begin{array}{cccc|c} 1&0&2&1&-3\\ 0&1&0&-1&2\\ 0&0&0&1&-2\\ 0&0&0&1&-2 \end{array}\right)\begin{array}{c} \\ \\ \\ -1 \end{array} \xrightarrow[\text{Typ III}]{}\\[1ex]
&\left(\begin{array}{cccc|c} 1&0&2&1&-3\\ 0&1&0&-1&2\\ 0&0&0&1&-2\\ 0&0&0&0&0 \end{array}\right)\begin{array}{c} -1\\ 1\\ \\ \\ \end{array} \xrightarrow[\text{Typ III}]{}
\left(\begin{array}{cccc|c} 1&0&2&0&-1\\ 0&1&0&0&0\\ 0&0&0&1&-2\\ 0&0&0&0&0 \end{array}\right)
\end{aligned}
$$

Hierbei haben wir jeweils gekennzeichnet, wie wir von einer Matrix zur nächsten gekommen sind. Dies ist sehr zu empfehlen, damit die Rechnung nachvollziehbar und Fehler korrigierbar sind. ◁

Nun können wir das Verfahren formalisieren. Wir erhalten den berühmten Gauß-Algorithmus.

Algorithmus 2.5 (Gauß)**.**

Eingabe: Eine Matrix $A \in K^{m \times n}$.
Ausgabe: Eine Matrix $B \in K^{m \times n}$ in (strenger) Zeilenstufenform, die aus A durch elementare Zeilenoperationen hervorgeht.

(1) Setze $B := A$.
(2) B sei bis zur r-ten Zeile in Zeilenstufenform, d.h. (a) und (b) aus Definition 2.2 seien bis zur r-ten Zeile erfüllt. (Hierbei ist $r = 0$ möglich!)
(3) Falls $r = m$, so ist B in Zeilenstufenform. Falls strenge Zeilenstufenform gewünscht ist, gehe zu (8).
(4) Suche den am weitesten links stehenden Eintrag $\neq 0$ von B unterhalb der r-ten Zeile. (Falls es mehrere solche Einträge gibt, wähle einen aus.)
(5) Bringe diesen Eintrag in die $(r+1)$-te Zeile (Operation Typ I).
(6) Erzeuge unterhalb dieses Eintrags lauter Nullen (Operationen Typ III, optional auch II).
(7) Gehe zu (2).
(8) Bringe B auf strenge Zeilenstufenform (Operationen Typ III).

Der Gaußalgorithmus ist der wichtigste Algorithmus der linearen Algebra. Wir werden noch sehen, dass er für viele rechnerische Aufgaben eingesetzt wird. Wir haben ihn im Zusammenhang mit linearen Gleichungssystemen eingeführt. Da wir bereits gesehen haben, dass sich bei elementaren Zeilenoperationen die Lösungsmenge nicht ändert, müssen wir uns nur noch überzeugen, dass wir anhand einer (strengen) Zeilenstufenform des Systems die Lösungsmenge besonders leicht ablesen können.

Beispiel 2.6. Wir setzen das Beispiel des in (2.1) gegebenen LGS fort. In Beispiel 2.4 wurde die erweiterte Koeffizientenmatrix auf strenge Zeilenstufenform gebracht, wodurch wir das äquivalente LGS

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -2 \\ 0 \end{pmatrix}$$

erhalten. In ausführlicher Schreibweise liest sich dies als

$$\begin{aligned} x_1 + 2x_3 &= -1, \\ x_2 &= 0, \\ x_4 &= -2. \end{aligned}$$

Die Lösungsmenge lässt sich ablesen:

$$L = \left\{ \begin{pmatrix} -2a-1 \\ 0 \\ a \\ -2 \end{pmatrix} \mid a \in K \right\}.$$

◁

Anmerkung. Der Aufwand für den Gauß-Algorithmus bei Anwendung auf eine Matrix in $K^{n \times n}$ beträgt $O(n^3)$ Körperoperationen. ◁

Jetzt geben wir unser Lösungsverfahren für LGS in formalerer Weise an.

Algorithmus 2.7 (Lösen von LGS)**.**

Eingabe: Ein LGS $A \cdot x = b$ mit $A \in K^{m \times n}$ und $b \in K^m$ (also m Gleichungen mit n Unbekannten).
Ausgabe: Die Lösungsmenge L.

(1) Bringe die erweiterte Koeffizientenmatrix $(A|b) \in K^{m \times (n+1)}$ auf strenge Zeilenstufenform. Ab jetzt setzen wir voraus, dass $(A|b)$ bereits in strenger Zeilenstufenform ist.
(2) Es sei r die Anzahl der Zeilen, die mindestens einen Eintag $\neq 0$ haben. Für $i = 1, \ldots, r$ sei $j_i \in \{1, \ldots, n+1\}$ die Position (= Spalte), in der der erste Eintrag $\neq 0$ der i-ten Zeile steht.
(3) Falls $j_r = n+1$, so ist das LGS unlösbar, also $L = \emptyset$. (Die r-te Zeile lautet dann nämlich $(0 \cdots 0 | b_r)$ mit $b_r \neq 0$, was der Gleichung $0 = b_r$ entspricht.)
(4) Andernfalls seien $k_1, \ldots, k_{n-r}$ diejenigen Zahlen in $\{1, \ldots, n\}$, die nicht eines der j_i sind. Also $\{1, \ldots, n\} \setminus \{j_1, \ldots, j_r\} = \{k_1, \ldots, k_{n-r}\}$.
(5) Die Lösungsmenge ist

$$L = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \,\middle|\, x_{k_1}, \ldots, x_{k_{n-r}} \in K \text{ beliebig}, \right.$$
$$\left. x_{j_i} = a_{i,j_i}^{-1} \cdot \Big(b_i - \sum_{j=1}^{n-r} a_{i,k_j} \cdot x_{k_j} \Big) \text{ für } i = 1, \ldots, r \right\}. \quad (2.2)$$

Die Lösungsmenge wird also parametrisiert durch die „freien“ Variablen x_{k_i}, während die x_{j_i} von diesen abhängig sind.

Es ist fast unmöglich, sich die Formel (2.2) zu merken, und noch unmöglicher, sie tatsächlich anzuwenden, es sei denn, man ist ein Computer und kein Mensch. Man ist also weiterhin darauf angewiesen, die Lösungsmenge eines LGS anhand der strengen Zeilenstufenform mit Hilfe von mathematisch-handwerklichen Grundfertigkeiten abzulesen.

Bei homogenen LGS wird die Erweiterungsspalte b weggelassen, und der Fall der Unlösbarkeit tritt nie ein.

Bei LGS können drei „Hauptfälle“ eintreten:

(1) Unlösbarkeit: $L = \emptyset \iff j_r = n + 1$.

(2) Eindeutige Lösbarkeit: $|L| = 1 \iff r = n$ und $j_r = n$. In diesem Fall gilt automatisch $j_i = i$ für alle i, und die strenge Zeilenstufenform hat die übersichtliche Gestalt

$$\left(\begin{array}{ccccc|c} a_{1,1} & 0 & & \cdots & 0 & b_1 \\ 0 & a_{2,2} & & & \vdots & \vdots \\ & & \ddots & & \vdots & \vdots \\ \vdots & & & a_{n-1,n-1} & 0 & b_{n-1} \\ 0 & \cdots & & 0 & a_{n,n} & b_n \\ 0 & \cdots & & \cdots & 0 & 0 \\ \vdots & & & & \vdots & \vdots \\ 0 & \cdots & & \cdots & 0 & 0 \end{array} \right).$$

Die (einzige) Lösung ergibt sich dann als $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1/a_{1,1} \\ \vdots \\ b_n/a_{n,n} \end{pmatrix}$.

(3) Uneindeutige Lösbarkeit: $|L| > 1 \iff r < n$ und $j_r \neq n+1$. Dann hat die Lösungsmenge $n - r$ freie Parameter. Insbesondere folgt $|L| = \infty$, falls K unendlich viele Elemente hat (der Standardfall).

Allein aus der Anzahl der Gleichungen und der Unbekannten kann man nicht auf den Eintritt eines der Hauptfälle schließen. Als Einziges lässt sich sagen, dass eindeutige Lösbarkeit nur dann eintreten kann, wenn mindestens so viele Gleichungen wie Unbekannte vorhanden sind.

Die Zahl r aus Algorithmus 2.7 spielt eine wichtige Rolle. Daher geben wir ihr einen Namen.

Definition 2.8. *Es sei $A \in K^{m\times n}$, und $A' \in K^{m\times n}$ sei eine Matrix in Zeilenstufenform, die durch elementare Zeilenoperationen aus A hervorgegangen ist. Dann ist der* **Rang** *von A die Anzahl r der Zeilen in A', die mindestens einen Eintrag $\neq 0$ haben. Wir schreiben $r =: \mathrm{rg}(A)$.*

Eine quadratische Matrix $A \in K^{n\times n}$ heißt **regulär***, falls $rg(A) = n$.*

Das Problem bei dieser Definition ist, dass es verschiedene Matrizen A' gibt, die in Zeilenstufenform und durch elementare Zeilenoperationen aus A hervorgegangen sind. Aber (bisher) ist nicht klar, dass all diese A' dieselbe Anzahl von Zeilen $\neq 0$ haben. Nur wenn dies klar ist, ist $\mathrm{rg}(A)$ eindeutig definiert. Wir werden dies im Kapitel 6 nachtragen.

Wir sehen sofort, dass für die Einheits- und Nullmatrix gelten: $\mathrm{rg}(I_n) = n$ (also regulär) und $\mathrm{rg}(\mathbf{0}) = 0$. Außerdem gilt für $A \in K^{m\times n}$: $\mathrm{rg}(A) \leq \min\{m, n\}$. Unser Lösbarkeitskriterium für LGS können wir nun so formulieren:

Satz 2.9. *Ein LGS $A \cdot x = b$ ist genau dann lösbar, wenn die Koeffizientenmatrix A denselben Rang hat wie die erweiterte Koeffizientenmatrix $(A|b)$.*

Kapitel 3
Gruppen

Definition 3.1. *Eine* **Gruppe** *ist eine Menge G zusammen mit einer Abbildung $p\colon G \times G \to G$ (die wir* **Produkt** *nennen und für die wir die Schreibweise $p(a,b) = a \cdot b = ab$ verwenden), so dass die folgenden Axiome gelten:*

$$\forall\ a,b,c \in G: \quad (a \cdot b) \cdot c = a \cdot (b \cdot c), \tag{AG}$$

$$\exists\ e \in G: \quad \forall\ a \in G: \quad e \cdot a = a, \tag{NE}$$

$$\forall\ a \in G: \quad \exists\ a' \in G: \quad a' \cdot a = e. \tag{IE}$$

(Hierbei ist (IE) eigentlich eine weitere Eigenschaft von e.)

Eine Gruppe G heißt **kommutativ** *(oder auch abelsch), falls außerdem gilt:*

$$\forall\ a,b \in G: \quad a \cdot b = b \cdot a. \tag{KG}$$

Bevor wir Beispiele von Gruppen anschauen, beweisen wir das folgende Resultat:

Satz 3.2. *Für jede Gruppe G gelten:*

(a) Es gibt genau ein $e \in G$, das (NE) erfüllt. Dieses e heißt das **neutrale Element** *von G.*
(b) Für jedes $a \in G$ gibt es genau ein $a' \in G$, das (IE) erfüllt. Dieses a' heißt das **inverse Element** *zu a und wird mit $a' = a^{-1}$ bezeichnet.*
(c) Für jedes $a \in G$ gelten

$$ae = a \quad \text{und} \quad aa^{-1} = e.$$

Beweis. Wir beginnen mit (c). Für $a \in G$ gibt es wegen (IE) $a' \in G$ mit $a'a = e$ und $a'' \in G$ mit $a''a' = e$. Es folgt

$$\begin{aligned} aa' &\underset{\text{(NE)}}{=} e(aa') \underset{\text{(IE)}}{=} (a''a')(aa') \underset{\text{(AG)}}{=} a''\left(a'(aa')\right) \\ &\underset{\text{(AG)}}{=} a''\left((a'a)a'\right) \underset{\text{(IE)}}{=} a''(ea') \underset{\text{(NE)}}{=} a''a' \underset{\text{(IE)}}{=} e, \end{aligned} \tag{3.1}$$

und weiter
$$ae \underset{(\mathrm{IE})}{=} a(a'a) \underset{(\mathrm{AG})}{=} (aa')a \underset{(3.1)}{=} ea \underset{(\mathrm{NE})}{=} a. \tag{3.2}$$

Damit ist (c) nachgewiesen. Zum Beweis von (a) sei $\tilde{e} \in G$ ein weiteres Element, das (NE) erfüllt. Dann folgt

$$\tilde{e} \underset{(3.2)}{=} \tilde{e}e \underset{(\mathrm{NE})}{=} e,$$

was die behauptete Eindeutigkeit liefert. Zum Beweis von (b) sei $\tilde{a} \in G$ ein weiteres Element mit $\tilde{a}a = e$. Dann folgt

$$\tilde{a} \underset{(3.2)}{=} \tilde{a}e \underset{(3.1)}{=} \tilde{a}(aa') \underset{(\mathrm{AG})}{=} (\tilde{a}a)a' = ea' \underset{(\mathrm{NE})}{=} a'.$$

Dies schließt den Beweis ab. □

Beispiel 3.3. (1) Die Mengen $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$ zusammen mit der gewöhnlichen Addition als Produkt sind kommutative Gruppen mit 0 als neutralem Element.

(2) Die Mengen $\mathbb{Q}\setminus\{0\}$ und $\mathbb{R}\setminus\{0\}$ zusammen mit dem gewöhnlichen Produkt sind kommutative Gruppen mit 1 als neutralem Element.

(3) Die Menge $\mathbb{Z} \setminus \{0\}$ mit dem gewöhnlichen Produkt ist keine Gruppe, da (IE) verletzt ist. Aber $\{1, -1\} \subseteq \mathbb{Z}$ ist mit dem gewöhnlichen Produkt eine kommutative Gruppe.

(4) Auf der Menge
$$G = \left\{(a_1, a_2) \in \mathbb{R}^2 \mid a_1 \neq 0\right\}$$
definieren wir ein Produkt durch
$$(a_1, a_2) \cdot (b_1, b_2) = (a_1b_1, a_1b_2 + a_2),$$
wobei wir in den Formeln die gewöhnliche Addition und Multiplikation von $\mathbb{R}$ verwenden. Für den Nachweis von (AG) nehmen wir $(a_1, a_2), (b_1, b_2), (c_1, c_2) \in G$ und bilden
$$\begin{aligned}\Big((a_1, a_2) \cdot (b_1, b_2)\Big) \cdot (c_1, c_2) &= (a_1b_1, a_1b_2 + a_2) \cdot (c_1, c_2)\\ &= \Big(a_1b_1c_1, a_1b_1c_2 + a_1b_2 + a_2\Big)\end{aligned}$$
und
$$\begin{aligned}(a_1, a_2) \cdot \Big((b_1, b_2) \cdot (c_1, c_2)\Big) &= (a_1, a_2) \cdot (b_1c_1, b_1c_2 + b_2)\\ &= \Big(a_1b_1c_1, a_1(b_1c_2 + b_2) + a_2\Big).\end{aligned}$$
Durch Vergleich erkennt man die Gültigkeit von (AG). Mit $e := (1, 0)$ gilt für alle $(a_1, a_2) \in G$:

$$e \cdot (a_1, a_2) = (a_1, a_2).$$

Außerdem gilt für $(a_1, a_2) \in G$:

$$(a_1^{-1}, -a_1^{-1}a_2) \cdot (a_1, a_2) = (1, 0) = e$$

(wobei a_1^{-1} das reelle Inverse ist). Also ist G eine Gruppe. Ist G kommutativ? Das Beispiel $(1,1) \cdot (2,1) = (2,2)$ und $(2,1) \cdot (1,1) = (2,3)$ zeigt, dass dies nicht der Fall ist.

(5) Die schon in Beispiel 1.4(2) eingeführten Drehmatrizen

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

mit $\varphi \in \mathbb{R}$ bilden (mit dem Matrixprodukt) eine Gruppe, die oft als SO(2) bezeichnet wird.

(6) Die Menge $G = \{e\}$ mit $e \cdot e = e$ bildet eine Gruppe, die *triviale Gruppe*.

(7) Die Menge aller Drehungen, die ein Quadrat in sich selbst überführen, ist mit der Komposition eine Gruppe. Sie hat 4 Elemente. Man nennt G die *Symmetriegruppe* des Quadrates. Auch andere geometrische Objekte haben Symmetriegruppen, ebenso Kristalle oder Moleküle. Beispielsweise hat der Würfel die Symmetriegruppe S_4 (siehe Definition 3.4), die nicht kommutativ ist. ◁

Für eine Gruppe G gelten die folgenden Rechenregeln:

- $\forall\, a \in G: \quad (a^{-1})^{-1} = a$,
- $\forall\, a, b \in G: \quad (ab)^{-1} = b^{-1}a^{-1}$.

Wir verwenden die folgenden Schreibweisen:

- Statt $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ schreiben wir $a \cdot b \cdot c$, und entsprechend $a \cdot b \cdot c \cdot d$ und so weiter.
- Für $n \in \mathbb{N}_{>0}$: $a^n = \underbrace{a \cdots a}_{n \text{ mal}}$, $a^0 = e$ und $a^{-n} = (a^n)^{-1}$.
- Kommutative Gruppen schreiben wir oft *additiv*: Statt $a \cdot b$ schreiben wir $a + b$. In diesem Fall schreiben wir 0 für das neutrale Element und $-a$ für das inverse Element von $a \in G$.

Das für uns wichtigste Beispiel einer Gruppe ist die symmetrische Gruppe, die wir nun einführen.

Definition 3.4. *Für eine Menge A wird*

$$S_A := \{f \colon A \to A \mid f \text{ ist bijektiv}\}$$

durch $f \cdot g := f \circ g$ (Komposition) eine Gruppe. (Die Gültigkeit von (AG) ist klar, die Identität ist das neutrale Element, und zu $f \in S_A$ ist die Umkehrabbildung das inverse Element.) S_A heißt die **symmetrische Gruppe** *auf A. Die Elemente von S_A heißen* **Permutationen**. *Besonders wichtig ist der*

Fall $A = \{1, \dots, n\}$ mit $n \in \mathbb{N}$. Hier schreiben wir S_n statt S_A und sprechen von der symmetrischen Gruppe auf n Ziffern.

Beispiel 3.5. (1) Für $n = 2$ ist

$$S_n = \{\mathrm{id}, a\}$$

mit $a(1) = 2$ und $a(2) = 1$. Es gilt $a^2 = \mathrm{id}$. S_2 ist kommutativ.

(2) Die S_3 hat 6 Elemente, denn es gibt $6 = 3!$ bijektive Abbildungen $\{1, 2, 3\} \to \{1, 2, 3\}$. Wir benutzen folgende Schreibweise: $(1, 2, 3)$ steht für die Permutation aus S_3 mit $1 \mapsto 2 \mapsto 3 \mapsto 1$, und $(1, 2)$ steht für die Permutation mit $1 \mapsto 2 \mapsto 1$ und $3 \mapsto 3$ (und entsprechend für andere Ziffern). Dann gilt

$$S_3 = \left\{ \mathrm{id}, \underbrace{(1, 2, 3)}_{=:a}, (3, 2, 1), \underbrace{(1, 2)}_{=:b}, (1, 3), (2, 3) \right\}.$$

Es gilt

$$a \cdot b = (1, 3),$$

aber

$$b \cdot a = (2, 3).$$

(Man beachte, dass man für die Bildung von $a \cdot b$ zuerst b und dann a ausführen muss.) S_3 ist also nicht kommutativ. ◁

Das obige Beispiel zeigt, dass S_n für $n > 2$ nicht kommutativ ist. Es gilt allgemein

$$|S_n| = n!.$$

Kapitel 4
Vektorräume

Matrizen sind die Hauptobjekte der „handwerklichen" linearen Algebra. Die Hauptobjekte der strukturellen linearen Algebra sind Vektorräume, die wir in diesem Kapitel definieren. Wie früher sei K durchweg ein Körper.

Definition 4.1. *Ein K-***Vektorraum** *(auch: Vektorraum über K) ist eine Menge V zusammen mit zwei Abbildungen $\boxplus\colon V \times V \to V$, $(v, w) \mapsto v \boxplus w$ und $\boxdot\colon K \times V \to V$, $(a, v) \mapsto a \boxdot v$, so dass folgende Axiome gelten:*

(1) V ist mit $\boxplus$ als Verknüpfung eine kommutative Gruppe.
(2) Für alle $a \in K$ und $v, w \in V$ gilt

$$a \boxdot (v \boxplus w) = a \boxdot v \boxplus a \boxdot w$$

(mit der Konvention Punkt vor Strich).
(3) Für alle $a, b \in K$ und $v \in V$ gilt

$$(a + b) \boxdot v = a \boxdot v \boxplus b \boxdot v.$$

(4) Für alle $a, b \in K$ und $v \in V$ gilt

$$(a \cdot b) \boxdot v = a \boxdot (b \boxdot v).$$

(5) Für alle $v \in V$ gilt

$$1 \boxdot v = v.$$

Die Elemente eines Vektorraums heißen **Vektoren**. *Wir haben die Symbole „$\boxplus$" und „$\boxdot$" für die Unterscheidung von der Addition und Multiplikation im Körper K verwendet. Ab jetzt werden wir immer $v + w$ für $v \boxplus w$ und $a \cdot v$ oder av für $a \boxdot v$ schreiben.*

Beispiel 4.2. (1) Wegen Satz 1.6(a) und (b) ist $K^{m \times n}$ ein K-Vektorraum.
(2) Insbesondere ist der n-dimensionale Standardraum K^n ein K-Vektorraum.
(3) K selbst ist ein K-Vektorraum (mit der Addition und Multiplikation von K). Dies ist der Spezialfall $n = 1$ aus (2).

(4) $V = \{0\}$ (kommutative Gruppe mit nur einem Element 0) wird mit $a \cdot 0 := 0$ für $a \in K$ ein K-Vektorraum. Dieser Vektorraum heißt der **Nullraum**.
(5) $\mathbb{C}$ (der Körper der komplexen Zahlen) ist ein $\mathbb{R}$-Vektorraum; $\mathbb{R}$ ist ein $\mathbb{Q}$-Vektorraum.
(6) Der Polynomring

$$K[x] := \left\{a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0 \mid n \in \mathbb{N},\ a_i \in K\right\}$$

ist ein K-Vektorraum (mit der üblichen Polynomaddition und dem üblichen Produkt einer Konstanten aus K und eines Polynoms).
(7) Für (festes) $d \in \mathbb{N}_0$ ist $\{f \in K[x] \mid \deg(f) < d\}$ ein K-Vektorraum. (Hierbei bezeichnet $\deg(f)$ den Grad des Polynoms f, wobei das Nullpolynom per Konvention den Grad $-\infty$ erhält.)
(8) M sei irgendeine Menge und

$$V := \mathrm{Abb}(M, K) := \{f\colon M \to K \mid f \text{ Abbildung}\}.$$

Für $f, g \in V$ und $a \in K$ definieren wir $f \boxplus g$ und $a \boxdot f \in V$ durch

$$f \boxplus g\colon M \to K,\ x \mapsto f(x) + g(x) \quad \text{und} \quad a \boxdot f\colon M \to K,\ x \mapsto a \cdot f(x).$$

(Man sagt auch, dass die Summe von Funktionen und das skalare Vielfache einer Funktion *punktweise* definiert werden.) Durch stures Nachrechen sieht man, dass V ein K-Vektorraum ist. Der Nullvektor ist die sogenannte *Nullfunktion* f_0, definiert durch $f_0(x) = 0$ für alle $x \in M$.
(9) Gegenbeispiel: Es sei V eine kommutative Gruppe mit neutralem Element 0, aber $V \neq \{0\}$. Wir setzen $a \boxdot v := 0$ für alle $a \in K$ und $v \in V$. Dann sind die Axiome (1) bis (4) in Definition 4.1 erfüllt, aber (5) nicht. Der mögliche Verdacht, dass (5) überflüssig sein könnte, erweist sich also als unbegründet. $\triangleleft$

Aus den Vektorraumaxiomen ergeben sich ein paar Rechenregeln:

Proposition 4.3. *Es seien V ein K-Vektorraum und $a \in K$, $v \in V$. Dann gelten:*

(a) $a \cdot 0 = 0$ und $0 \cdot v = 0$ (in der ersten Gleichung bezeichnet die linke 0 den Nullvektor, in der zweiten das Nullelement von K);
(b) $(-a) \cdot v = a \cdot (-v) = -(a \cdot v)$;
(c) aus $a \cdot v = 0$ folgt $a = 0$ oder $v = 0$.

Beweis. Wir verwenden nur die Vektorraum- (und Körper-)Axiome.

(a) Es gelten

$$a \cdot 0 \underset{(1)}{=} a \cdot 0 + a \cdot 0 - (a \cdot 0) \underset{(2)}{=} a \cdot (0 + 0) - (a \cdot 0) \underset{(1)}{=} a \cdot 0 - (a \cdot 0) \underset{(1)}{=} 0$$

und

$$0 \cdot v \underset{(1)}{=} 0 \cdot v + 0 \cdot v - (0 \cdot v) \underset{(3)}{=} (0+0) \cdot v - (0 \cdot v) = 0 \cdot v - (0 \cdot v) \underset{(1)}{=} 0.$$

(b) Es gelten

$$(-a)v \underset{(1)}{=} (-a)v + av - (av) \underset{(3)}{=} (-a+a)v - (av) = 0v - (av) \underset{(a)}{=} -(av)$$

und

$$a(-v) \underset{(1)}{=} a(-v) + av - (av) \underset{(2)}{=} a(-v+v) - (av) \underset{(1)}{=} a0 - (av) \underset{(a)}{=} -(av).$$

(c) Es sei $a \cdot v = 0$ aber $a \neq 0$. Dann folgt

$$v \underset{(5)}{=} 1 \cdot v = (a^{-1}a) \cdot v \underset{(4)}{=} a^{-1} \cdot (av) = a^{-1} \cdot 0 \underset{(a)}{=} 0.$$

□

Definition 4.4. *Sei V ein K-Vektorraum. Eine Teilmenge $U \subseteq V$ heißt ein* **Unterraum** *(auch: Untervektorraum, Teilraum), falls gelten:*

(1) $U \neq \emptyset$;
(2) Für $v, w \in U$ ist auch $v + w \in U$;
(3) Für $a \in K$ und $v \in U$ gilt $a \cdot v \in U$.

Aus der Definition folgt sofort:

- Jeder Unterraum enthält den Nullvektor.
- Mit den Operationen „+“ und „·“ von V wird ein Unterraum U selbst ein K-Vektorraum.

Beispiel 4.5. (1) $V = \mathbb{R}^2$ ist ein Vektorraum über $K = \mathbb{R}$. Jede Gerade durch den Nullpunkt ist ein Unterraum. Formaler: Wähle $v \in V$. Dann ist $K \cdot v := \{a \cdot v \mid a \in K\} \subseteq V$ ein Unterraum. Dies gilt sogar für jeden Vektorraum V und $v \in V$. Geraden im $\mathbb{R}^2$, die nicht durch den Nullpunkt gehen, sind keine Unterräume.
(2) $U = \{0\}$ und V selbst sind Unterräume eines Vektorraums V.
(3) Sei $V = K[x]$ der Polynomring und $d \in \mathbb{N}_0$ fest. Dann ist

$$U = \{f \in V \mid \deg(f) < d\} \subseteq V$$

ein Unterraum (siehe Beispiel 4.2(6) und (7)).
(4) Sei M eine Menge und $V = \mathrm{Abb}(M, K)$ (siehe Beispiel 4.2(8)). Wähle $x \in M$ fest. Dann ist

$$U := \{f \in V \mid f(x) = 0\} \subseteq V$$

ein Unterraum. (Die Bedingung $f(x) = 1$ würde aber nicht zu einem Unterraum führen!)

(5) Ein besonders wichtiges Beispiel: Die Lösungsmenge eines homogenen LGS $A \cdot x = 0$ mit $A \in K^{m \times n}$ ist ein Unterraum von K^n.
(6) Die Vereinigungsmenge zweier Geraden $U_1, U_2 \subseteq \mathbb{R}^2$ durch den Nullpunkt ist kein Unterraum (es sei denn $U_1 = U_2$). ◁

Das letzte Beispiel zeigt, dass Vereinigungen von Unterräumen im Allgemeinen keine Unterräume sind. Die folgende Proposition beschäftigt sich mit Schnitten von Unterräumen.

Proposition 4.6. *Es seien V ein K-Vektorraum und $U_1, U_2 \subseteq V$ Unterräume. Dann gelten:*

(a) $U_1 \cap U_2 \subseteq V$ ist ein Unterraum.
(b) $U_1 + U_2 := \{v + w \mid v \in U_1,\ w \in U_2\} \subseteq V$ ist ein Unterraum.
(c) Ist $\mathcal{M} \neq \emptyset$ eine nicht-leere Menge, deren Elemente Unterräume von V sind, so ist auch der Schnitt

$$\bigcap_{U \in \mathcal{M}} U \subseteq V$$

ein Unterraum.

Beweis. Wir müssen nur (b) und (c) zeigen, da (a) ein Spezialfall von (c) ist.

(b) Es gilt $U_1 + U_2 \neq \emptyset$. Seien $v + w$ und $v' + w'$ Elemente von $U_1 + U_2$ mit $v, v' \in U_1$, $w, w' \in U_2$. Dann folgt

$$(v + w) + (v' + w') = (v + v') + (w + w') \in U_1 + U_2,$$

und für $a \in K$ folgt $a \cdot (v + w) = av + aw \in U_1 + U_2$. Also ist $U_1 + U_2$ ein Unterraum.
(c) Wir schreiben $W := \bigcap_{U \in \mathcal{M}} U$. Für alle $U \in \mathcal{M}$ gilt $0 \in U$, also $0 \in W$. Weiter gilt für $v, w \in W$, dass v und w in allen $U \in \mathcal{M}$ liegen. Damit auch $v + w \in U$ für alle $U \in \mathcal{M}$, also $v + w \in W$. Ebenso folgt $a \cdot v \in W$ für $a \in K$ und $v \in W$. damit ist gezeigt, dass W ein Unterraum ist. □

Der Unterraum $U_1 + U_2$ aus Proposition 4.6(b) heißt der **Summenraum** von U_1 und U_2. Proposition 4.6(c) drückt man manchmal aus, indem man sagt, dass die Menge der Unterräume eines Vektorraums ein *durchschnittsabgeschlossenes System* bilden. Diese Aussage macht die folgende Definition möglich.

Definition 4.7. *Es seien V ein K-Vektorraum und $S \subseteq V$ eine Teilmenge. (Wir setzen nicht voraus, dass S ein Unterraum ist.) Wir betrachten die Menge $\mathcal{M} := \{U \subseteq V \mid U \text{ ist ein Unterraum und } S \subseteq U\}$ und bilden*

$$\langle S \rangle := \bigcap_{U \in \mathcal{M}} U. \tag{4.1}$$

$\langle S\rangle$ heißt der von S **erzeugte Unterraum** *(auch: aufgespannter Unterraum, Erzeugnis) von V. Falls $S = \{v_1, \ldots, v_n\}$ endlich ist, schreiben wir $\langle S\rangle$ auch als $\langle v_1, \ldots, v_n\rangle$.*

$\langle S\rangle$ ist der kleinste Unterraum von V, der S (als Teilmenge) enthält. Genauer: Jeder Unterraum von V, der S enthält, enthält auch $\langle S\rangle$.

Die obige Definition ist konzeptionell elegant. Sie wirft jedoch die Frage auf, wie sich der von S erzeugte Unterraum explizit beschreiben lässt. Dieser Frage wenden wir uns jetzt und zu Beginn des folgenden Kapitels zu.

Beispiel 4.8. (1) Sei $v \in V$ ein Vektor. Wie sieht $\langle v\rangle$ aus? Die Antwort lautet: $\langle v\rangle = K \cdot v = \{a \cdot v \mid a \in K\}$. Denn $K \cdot v$ ist ein Unterraum, der v enthält, und andererseits ist $K \cdot v$ in jedem Unterraum U mit $v \in U$ enthalten.

(2) Noch einfacher ist der Fall $S = \emptyset$: $\langle \emptyset\rangle = \{0\}$, der Nullraum. ◁

Wir betrachten nun den Fall, dass S die Vereinigung zweier Unterräume ist.

Satz 4.9. *Es seien V ein K-Vektorraum, U_1 und U_2 Unterräume und $S := U_1 \cup U_2$. Dann gilt*

$$\langle S\rangle = U_1 + U_2.$$

Beweis. Nach Proposition 4.6(b) ist $U_1 + U_2$ ein Unterraum. Außerdem liegt jedes $v \in U_1$ (als $v+0$) und jedes $w \in U_2$ (als $0+w$) in U_1+U_2. U_1+U_2 ist also einer der Räume U, die in (4.1) zum Schnitt kommen, also $\langle S\rangle \subseteq U_1 + U_2$.

Umgekehrt sei $U \subseteq V$ ein Unterraum mit $S \subseteq U$. Für $v \in U_1$ und $w \in U_2$ folgt dann $v + w \in U$, also $U_1 + U_2 \subseteq U$. Wegen (4.1) impliziert dies $U_1 + U_2 \subseteq \langle S\rangle$. □

Beispiel 4.10. Es seien $U_1, U_2 \subseteq \mathbb{R}^3$ zwei verschiedene Geraden durch den Nullpunkt. Dann ist $U_1 + U_2$ eine Ebene. ◁

Kapitel 5
Linearkombinationen

Auch in diesem Kapitel ist K stets ein Körper. Falls nichts anderes gesagt wird, ist V ein K-Vektorraum. Um eine allgemeingültige Antwort auf die Frage nach einer expliziten Beschreibung des erzeugten Unterraums $\langle S\rangle$ einer Teilmenge $S \subseteq V$ zu geben, benötigen wir eine Definition.

Definition 5.1. *(a) Es seien $v_1, \ldots, v_n \in V$ Vektoren. Ein Vektor $v \in V$ heißt* **Linearkombination** *von $v_1, \ldots, v_n$, falls es Skalare $a_1, \ldots, a_n \in K$ gibt mit*

$$v = a_1 v_1 + \cdots + a_n v_n.$$

(b) Sei $S \subseteq V$ eine Teilmenge. Ein Vektor $v \in V$ heißt **Linearkombination** *von S, falls es $n \in \mathbb{N}$ und $v_1, \ldots, v_n \in S$ gibt, so dass v eine Linearkombination von $v_1, \ldots, v_n$ ist. Falls $S = \emptyset$, so sagen wir, dass der Nullvektor 0 (die einzige) Linearkombination von S ist. (0 wird als leere Summe aufgefasst.)*

Es ist klar, dass die Teile (a) und (b) der Definition für endliche Mengen $S = \{v_1, \ldots, v_n\}$ übereinstimmen. In (b) geht man über endliche Auswahlen von Vektoren, da es in der linearen Algebra nur endliche Summen gibt (ebenso wie in der Analysis, in der man Grenzwerte von endlichen Teilsummen betrachtet). Nun beantworten wir die Frage nach dem erzeugten Unterraum.

Satz 5.2. *Für eine Teilmenge $S \subseteq V$ ist der erzeugte Unterraum $\langle S\rangle$ die Menge aller Linearkombinationen von S:*

$$\langle S\rangle = \{v \in V \mid v \text{ ist Linearkombination von } S\}.$$

Insbesondere gilt für $v_1, \ldots, v_n \in V$:

$$\langle v_1, \ldots, v_n\rangle = \left\{ \sum_{i=1}^{n} a_i v_i \mid a_1, \ldots, a_n \in K \right\}.$$

Beweis. Es sei $W \subseteq V$ die Menge aller Linearkombinationen von S. Es gilt $0 \in W$. Da die Summe zweier Linearkombinationen und ein skalares Vielfaches einer Linearkombination wieder Linearkombinationen sind, folgt, dass W ein Unterraum ist. Außerdem liegt jedes $v \in S$ in W. Damit ist W einer der Unterräume U, die in (4.1) zum Schnitt kommen. Es folgt $\langle S \rangle \subseteq W$.

Andererseits sei $U \subseteq V$ ein Unterraum mit $S \subseteq U$. Für $v_1, \ldots, v_n \in S$ und $a_1, \ldots, a_n \in K$ liegen dann alle v_i in U und damit auch $\sum_{i=1}^n a_i v_i$. Also enthält U alle Linearkombinationen von S, d.h. $W \subseteq U$. Dies impliziert $W \subseteq \langle S \rangle$, und der Beweis ist abgeschlossen. $\square$

Beispiel 5.3. Wir betrachten $V = \mathbb{R}^3$.

(1) Für $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ ist

$$\langle v_1, v_2 \rangle = \left\{ a_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \mid a_1, a_2 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} \mid a_1, a_2 \in \mathbb{R} \right\}.$$

Die Vektoren $v_1' = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$ und $v_2' = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ erzeugen den selben Unterraum.

(2) Für $v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ und $v_2 = \begin{pmatrix} -2 \\ -2 \\ 0 \end{pmatrix}$ ist

$$\langle v_1, v_2 \rangle = \left\{ \begin{pmatrix} a_1 - 2a_2 \\ a_1 - 2a_2 \\ 0 \end{pmatrix} \mid a_1, a_2 \in \mathbb{R} \right\} = \langle v_1 \rangle.$$

(3) Das homogene LGS $A \cdot x = 0$ mit $A = \left(\begin{smallmatrix} 1 & 0 & 2 & 1 \\ 2 & 0 & 4 & -2 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 2 & 2 \end{smallmatrix}\right)$ (siehe (2.1)). Aus der Lösung des inhomogenen LGS in Beispiel 2.6 geht die Lösungsmenge

$$L = \left\{ \begin{pmatrix} -2a \\ 0 \\ a \\ 0 \end{pmatrix} \mid a \in \mathbb{R} \right\} = \left\langle \begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle$$

hervor.

(4)

$$K^3 = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle.$$

Dies lässt sich auf K^n verallgemeinern.

(5) Ein Beispiel für ein unendliches Erzeugendensystem: Der Polynomring $V = K[x]$ wird (als K-Vektorraum) erzeugt von $S = \{x^i \mid i \in \mathbb{N}_0\} = \{1, x, x^2, \dots\}$. $\triangleleft$

Die folgende Proposition ist entscheidend für den Nachweis, dass unsere Definition von Rang einer Matrix A nicht von der Wahl der Matrix A', die aus A durch elementare Zeilenoperationen hervorgegangen ist, abhängt.

Proposition 5.4. *Es seien $A, A' \in K^{m\times n}$, wobei A' durch elementare Zeilenoperationen aus A hervorgegangen ist. Dann erzeugen die Zeilen von A denselben Unterraum von $K^{1\times n}$ wie die Zeilen von A'.*

Beweis. Wir müssen zeigen, dass elementare Zeilenoperationen den von den Zeilen $v_1, \dots, v_m$ erzeugten Raum U nicht ändern.

Typ I: Offenbar ändert sich U nicht.
Typ II: ebenso.
Typ III: Nach Umnummerieren der Zeilen ersetzt die Operation v_1 durch $v_1 + sv_2$, $s \in K$. Die neuen Zeilen erzeugen

$$\langle v_1 + sv_2, v_2, \dots, v_n\rangle = \left\{ a_1(v_1 + sv_2) + \sum_{i=2}^{m} a_i v_i \mid a_i \in K \right\} = U,$$

also auch hier keine Änderung. $\square$

Bei Beispiel 5.3(1),(3),(4) und (5) fällt auf, dass jeder Vektor aus dem erzeugten Unterraum *eindeutig* als Linearkombination darstellbar ist, d.h. es gibt nur eine Wahl für die Koeffizienten a_i. Beim Beispiel 5.3(2) ist dies nicht der Fall. Diese Beobachtung gibt Anlass zu folgender Definition.

Definition 5.5. *(a) Vektoren $v_1, \dots, v_n \in V$ heißen* **linear unabhängig**, *falls für alle $a_1, \dots, a_n$ folgende Implikation gilt:*

$$a_1v_1 + \dots + a_nv_n = 0 \quad \Rightarrow \quad a_1 = 0,\ a_2 = 0, \dots, a_n = 0.$$

*Gleichbedeutend damit ist: Für jede Linearkombination $v \in \langle v_1, \dots, v_n\rangle$ gibt es eindeutig bestimmte $a_1, \dots, a_n \in K$ mit $v = \sum_{i=1}^n a_i v_i$ („*eindeutige Darstellungseigenschaft*“). Der Beweis, dass lineare Unabhängigkeit und die eindeutige Darstellungseigenschaft gleichbedeutend sind, sei dem Leser überlassen. Die Vektoren $v_1, \dots, v_n$ heißen* **linear abhängig**, *falls sie nicht linear unabhängig sind. Wir betonen, dass es sich hierbei nicht um Eigenschaften von einzelnen Vektoren handelt (außer im Fall $n = 1$), sondern um Eigenschaften eines „Ensembles“ von Vektoren.*

(b) Eine Teilmenge $S \subseteq V$ heißt **linear unabhängig**, *falls für alle $n \in \mathbb{N}$ und alle paarweise verschiedenen $v_1, \dots, v_n \in S$ gilt, dass $v_1, \dots, v_n$ linear unabhängig ist. Andernfalls heißt S* **linear abhängig**. *$S = \emptyset$ ist (per definitionem) linear unabhängig.*

Beispiel 5.6. (1) Seien $V = \mathbb{R}^2$, $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Wir testen auf lineare Unabhängigkeit. Es gelte also $a_1 v_1 + a_2 v_2 = 0$ mit $a_1, a_2 \in \mathbb{R}$. Hieraus ergibt sich das homogene LGS $a_1 + a_2 = 0$, $a_1 - a_2 = 0$. Die einzige Lösung ist $a_1 = a_2 = 0$, also sind v_1, v_2 linear unabhängig.

(2) Nun betrachten wir $v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ und $v_2 = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix} \in \mathbb{R}^3$. Wenn wir wie oben auf lineare Unabhängigkeit testen, erhalten wir das homogene LGS $a_1 + 2a_2 = 0$, $-a_1 - 2a_2 = 0$, $0 = 0$, das (unter anderen) die nicht-triviale Lösung $a_1 = 2$, $a_2 = -1$ hat. Es folgt $2v_1 - v_2 = 0$, also sind v_1, v_2 linear abhängig.

(3) Es seien $V = K[x]$ und $S = \{x^i \mid i \in \mathbb{N}_0\}$. Wir behaupten, dass S linear unabhängig ist. Zum Nachweis nehmen wir beliebige, paarweise verschiedene $x^{i_1}, \ldots, x^{i_n} \in S$ und setzen $\sum_{j=1}^n a_j x^{i_j} = 0$ mit $a_j \in K$ voraus. Hieraus folgt (mit dem üblichen Identitätsbegriff für Polynome) direkt, dass $a_j = 0$ für alle j. Also ist S linear unabhängig.

(4) Der Fall $n = 1$: Ein einzelner Vektor $v \in V$ ist genau dann linear unabhängig, wenn $v \neq 0$. Dies folgt aus Proposition 4.3(c). $\triangleleft$

Für Vektoren $v_1, \ldots, v_n \in K^m$ haben wir folgenden Test auf lineare Unabhängigkeit: Man bilde die Matrix $A := (v_1|v_2|\cdots|v_n) \in K^{m\times n}$ mit den v_i als Spalten. (Die senkrechten Linien sollen nur der Verdeutlichung dienen.) Dann gilt:

$$\boxed{v_1, \ldots, v_n \text{ sind linear unabhängig} \quad \Longleftrightarrow \quad \operatorname{rg}(A) = n.}$$

Begründung: Die v_i sind genau dann linear unabhängig, wenn das homogene LGS $A \cdot x = 0$ als einzige Lösung den Nullvektor hat (siehe auch Beispiel 5.6(1) und (2)). Nach (2) auf Seite 15 und Definition 2.8 trifft dies genau dann ein, wenn $\operatorname{rg}(A) = n$. Wegen $\operatorname{rg}(A) \leq \min\{m, n\}$ (siehe nach Definition 2.8) folgt aus unserem Test sofort, dass im K^m höchstens m Vektoren linear unabhängig sein können. Hat man mehr als m Vektoren, so sind diese automatisch linear abhängig. Im folgenden Kapitel wird diese Beobachtung verallgemeinert.

Kapitel 6
Basen

In diesem Kapitel führen wir zwei zentrale Begriffe der linearen Algebra ein: Basis und Dimension. Wie zuvor bezeichnet K immer einen Körper und V einen Vektorraum.

Definition 6.1. *Es sei $S \subseteq V$ eine Teilmenge.*

(a) S heißt ein **Erzeugendensystem** *von V, falls $\langle S \rangle = V$.*
(b) S heißt eine **Basis** *von V, falls S ein linear unabhängiges Erzeugendensystem von V ist. Anders gesagt: S ist Basis, falls jedes $v \in V$ in eindeutiger Weise als Linearkombination von S darstellbar ist.*

Beispiel 6.2. (1) Die Vektoren

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

bilden eine Basis von K^3.

(2) Auch die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \; v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

bilden eine Basis von K^3. Wir sehen also, dass ein Vektorraum mehrere Basen haben kann. (In der Tat haben „fast alle“ Vektorräume „sehr viele“ verschiedene Basen.)

(3) In Verallgemeinerung von (1) sei

$$e_i := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \text{ (i-te Position)} \in K^n.$$

Dann ist $S = \{e_1, \dots, e_n\}$ eine Basis von K^n. S heißt die **Standardbasis** des K^n.

(4) Für $V = K[x]$ ist $S = \{x^i \mid i \in \mathbb{N}_0\}$ eine Basis. Dies geht aus Beispiel 5.3(5) und aus Beispiel 5.6(3) hervor. Wir haben es hier mit einer unendlichen Basis zu tun.

(5) Der Lösungsraum

$$L = \left\langle \begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle$$

des homogenen LGS aus Beispiel 5.3(3) hat die Basis $S = \left\{ \begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}$.

(6) Allgemeiner sei $A \cdot x = 0$ ein homogenes LGS. Es seien $k_1, \dots, k_{n-r}$ die im Lösungsverfahren 2.7(4) bestimmten Indizes. Für $j = 1, \dots, n-r$ sei v_j der durch (2.2) gewonnene Lösungsvektor mit $x_{k_j} = 1$ und $x_{k_i} = 0$ für $i \neq j$. Dann ist $\{v_1, \dots, v_{n-r}\}$ eine Basis des Lösungsraums L. Die Erzeugereigenschaft ergibt sich direkt aus (2.2), und diese Gleichung zeigt außerdem, dass die k_j-te Koordinate von $\sum_{i=1}^{n-r} a_i v_i$ (mit $a_i \in K$) genau a_j ist, woraus die lineare Unabhängigkeit folgt. Dieses Beispiel ist von allgemeiner Bedeutung. Es zeigt, wie man eine Basis des Lösungsraums eines homogenen LGS gewinnt.

(7) Der Nullraum $V = \{0\}$ hat die leere Menge $S = \emptyset$ als Basis. Dies ist einer der exotischen Fälle, in denen es nur eine Basis gibt. $\triangleleft$

Wir geben nun zwei (zur Definition alternative) Charakterisierungen von Basen an.

Satz 6.3. *Für eine Teilmenge $S \subseteq V$ sind äquivalent:*

(a) S ist eine Basis von V.

(b) S ist eine maximal linear unabhängige Teilmenge von V (d.h. S ist linear unabhängig, aber für jedes $v \in V \setminus S$ wird $S \cup \{v\}$ linear abhängig).

(c) S ist ein minimales Erzeugendensystem von V (d.h. $V = \langle S \rangle$, aber für alle $v \in S$ ist $S \setminus \{v\}$ kein Erzeugendensystem).

Beweis. Wir beginnen mit der Implikation „(a) $\Rightarrow$ (b)“. Sei also S eine Basis von V. Dann ist S linear unabhängig, es ist also nur die Maximalität zu zeigen. Hierzu sei $v \in V \setminus S$. Da S ein Erzeugendensystem ist, gibt es $v_1, \ldots, v_n \in S$ und $a_1, \ldots, a_n \in K$ mit

$$v = \sum_{i=1}^{n} a_i v_i,$$

also

$$(-1) \cdot v + \sum_{i=1}^{n} a_i v_i = 0.$$

Dies zeigt, dass $\{v, v_1, \ldots, v_n\}$ linear abhängig ist, also auch $S \cup \{v\}$.

Nun zeigen wir „(b) $\Rightarrow$ (c)“. Es sei also S maximal linear unabhängig. Wir zeigen zunächst, dass S ein Erzeugendensystem ist. Hierzu sei $v \in V$. Falls $v \in S$, so gilt auch $v \in \langle S \rangle$, und wir sind fertig. Wir dürfen also $v \notin S$ annehmen. Nach Voraussetzung ist $S \cup \{v\}$ linear abhängig, also gibt es $v_1, \ldots, v_n \in S$ und $a, a_1, \ldots, a_n \in K$, die nicht alle 0 sind, so dass

$$av + \sum_{i=1}^{n} a_i v_i = 0.$$

(Selbst falls v in einer solchen Darstellung des Nullvektors nicht vorkäme, könnten wir es „künstlich“ durch $a := 0$ hinzufügen.) Falls $a = 0$, so wären $v_1, \ldots, v_n$ linear abhängig, im Widerspruch zur linearen Unabhängigkeit von S. Es folgt $a \neq 0$, also

$$v = -\sum_{i=1}^{n} a^{-1} a_i v_i \in \langle S \rangle.$$

Nun ist noch die Minimalität von S als Erzeugendensystem zu zeigen. Hierzu sei $v \in S$. Falls $S \setminus \{v\}$ ein Erzeugendensystem wäre, dann gäbe es insbesondere $v_1, \ldots, v_n \in S \setminus \{v\}$ und $a_1, \ldots, a_n \in K$ mit

$$v = \sum_{i=1}^{n} a_i v_i.$$

Hieraus folgt $(-1) \cdot v + \sum_{i=1}^{n} a_i v_i = 0$, im Widerspruch zur linearen Unabhängigkeit von S. Also ist S tatsächlich eine minimales Erzeugendensystem.

Schließlich zeigen wir „(c) $\Rightarrow$ (a)“. Es sei also S ein minimales Erzeugendensystem. Wir müssen die lineare Unabhängigkeit von S zeigen. Es seien also $v_1, \ldots, v_n \in S$ paarweise verschieden und $a_1, \ldots, a_n \in K$ mit $\sum_{i=1}^{n} a_i v_i = 0$. Wir nehmen an, dass nicht alle a_i Null sind. Durch Umnummerieren können wir $a_1 \neq 0$ erreichen. Es folgt

$$v_1 = \sum_{i=2}^{n} -a_1^{-1} a_i v_i \in \langle S' \rangle$$

mit $S' := S \setminus \{v_1\}$. Alle Elemente von S liegen also in $\langle S' \rangle$, also $V = \langle S' \rangle$, im Widerspruch zur Minimalität von S. Somit ist S linear unabhängig. □

Hat überhaupt jeder Vektorraum eine Basis? Aus dem obigen Satz können wir sofort eine Teilantwort als Folgerung ziehen.

Korollar 6.4. *Falls V ein endliches Erzeugendensystem besitzt, so hat V auch eine Basis.*

Beweis. Unter allen endlichen Erzeugendensystemen kann man eines mit minimaler Elementanzahl auswählen. Dieses ist dann auch minimal im Sinne von Satz 6.3(c), also nach Satz 6.3 eine Basis. □

Es gilt aber noch mehr:

Satz 6.5 (Basissatz)**.** *Jeder Vektorraum hat eine Basis.*

Der Beweis dieses Satzes benutzt das Zornsche Lemma. Wir lassen den Beweis weg. Wer mehr über das Zornsche Lemma erfahren möchte, kann Wikipedia konsultieren. In dieser Vorlesung wird Satz 6.5 nicht verwendet.

Beispiel 6.6. Es sei M eine unendliche Menge und $V = \mathrm{Abb}(M, K)$. Für V ist keine Basis bekannt, auch wenn Satz 6.5 die Existenz zusichert! Auch in Spezialfällen oder für viele interessante Unterräume ist keine Basis bekannt. Beipielsweise ist keine Basis für den Vektorraum der konvergenten reellen Folgen bekannt.

Für jedes $x \in M$ kann man die Abbildung $\delta_x \in V$ mit $\delta_x(y) = 1$ für $y = x$, 0 sonst, betrachten. Dann ist $S := \{\delta_x \mid x \in M\}$ linear unabhängig. S ist jedoch keine Erzeugendensystem, da es in der linearen Algebra keine unendlichen Summen gibt. ◁

Wir haben gesehen, dass ein Vektorraum (sehr viele) verschiedene Basen haben kann. Unser nächstes Ziel ist der Nachweis, dass alle Basen gleich viele Elemente haben (sofern sie endlich sind). Der Schlüssel hierzu ist das folgende Lemma.

Lemma 6.7. *Es seien $E \subseteq V$ ein endliches Erzeugendensystem und $U \subseteq V$ eine linear unabhängige Menge. Dann gilt für die Elementanzahlen:*

$$|U| \leq |E|.$$

Beweis. Als Teilmenge einer endlichen Menge ist auch $E \setminus U$ endlich. Wir benutzen Induktion nach $|E \setminus U|$. Wir schreiben $E = \{v_1, \ldots, v_n\}$ mit $v_1, \ldots, v_n$ paarweise verschieden.

1. Fall: $U \subseteq E$. Dann ist automatisch $|U| \leq |E|$, also nichts zu zeigen.

2. Fall: Es gibt ein $v \in U \setminus E$. Wir werden ein „Austauschargument“ benutzen und einen Vektor von E durch v ersetzen. Dies funktioniert folgendermaßen: Wegen $V = \langle E \rangle$ existieren $a_1, \ldots, a_n \in K$ mit

$$v = a_1 v_1 + \cdots + a_n v_n. \tag{6.1}$$

Wegen $v \notin E$ gilt $v \neq v_i$ für alle i. Es gibt ein i, so dass $v_i \notin U$ und $a_i \neq 0$, denn sonst ergäbe (6.1) die lineare Abhängigkeit von U. Nach Umnummerieren haben wir $v_1 \in E \setminus U$ und $a_1 \neq 0$. Dies zeigt auch, dass der Induktionsanfang ($|E \setminus U| = 0$) automatisch in den 1. Fall fällt. Mit $E' := \{v, v_2, \ldots, v_n\}$ ergibt sich aus (6.1):

$$v_1 = a_1^{-1} \cdot \left(v - \sum_{i=2}^{n} a_i v_i \right) \in \langle E' \rangle.$$

Hieraus folgt, dass auch E' ein Erzeugendensystem ist. Nach Definition von E' gilt $|E' \setminus U| = |E \setminus U| - 1$. Induktion liefert also $|U| \leq |E'|$. Wieder nach Definition gilt $|E'| = |E|$, und es folgt die Behauptung. □

Korollar 6.8. *Falls V ein endliches Erzeugendensystem hat, so sind alle Basen von V endlich und haben gleich viele Elemente.*

Beweis. B_1 und B_2 seien Basen von V. Da B_1 und B_2 linear unabhängig sind, liefert Lemma 6.7 $|B_1| < \infty$ und $|B_2| < \infty$. Weiter liefert Lemma 6.7 mit $U = B_1$ und $E = B_2$: $|B_1| \leq |B_2|$. Nach Rollenvertauschung erhalten wir ebenso $|B_2| \leq |B_1|$, also Gleichheit. □

Nun können wir einen der wichtigsten Begriffe der linearen Algebra definieren.

Definition 6.9. *Falls V ein endliches Erzeugendensystem hat, so ist die* **Dimension** *von V die Elementanzahl einer (und damit jeder) Basis von V. Wir schreiben $\dim(V)$ für die Dimension von V. Falls V kein endliches Erzeugendensystem hat, schreiben wir $\dim(V) := \infty$, um diesen Sachverhalt auszudrücken. Im ersten Fall heißt V* **endlich-dimensional***, im zweiten* **unendlich-dimensional**.

Beispiel 6.10. (1) Der Standardraum K^n hat die Dimension n. Damit ist auch die Bezeichnung „n-dimensionaler Standardraum“ aufgeklärt.
(2) Der Lösungsraum des homogenenen LGS $A \cdot x = 0$ aus Beispiel 5.3(3) hat die Dimension 1 (siehe Beispiel 6.2(5)).
(3) Der Nullraum $V = \{0\}$ hat die Dimension 0.
(4) Für $V = K[x]$ gilt $\dim(V) = \infty$. Hier können wir eine unendliche Basis angeben (siehe Beispiel 6.2(4)). Ist M eine unendliche Menge, so gilt auch $\dim(\mathrm{Abb}(M, K)) = \infty$. Wir können zwar keine Basis angeben, aber doch eine unendliche linear unabhängige Menge (siehe Beispiel 6.6), so dass $\mathrm{Abb}(M, K)$ nach Lemma 6.7 nicht endlich erzeugt sein kann. ◁

Aus Beispiel 6.2(6) gewinnen wir:

Proposition 6.11. *Es sei* $A \cdot x = 0$ *ein homogenes LGS mit* $A \in K^{m \times n}$*. Dann gilt für die Lösungsmenge* L*:*

$$\dim(L) = n - \operatorname{rg}(A).$$

Wie kann man eine Basis eines Unterraums $U \subseteq K^n$ finden? Wir nehmen an, U sei durch erzeugende Vektoren $v_1, \ldots, v_m$ gegeben. Dann bilden wir die Matrix $A \in K^{m \times n}$ mit den v_i als *Zeilen*. Nun bringen wir A mit dem Gauß-Algorithmus auf Zeilenstufenform. Dann bilden diejenigen Zeilen der Zeilenstufenform, die nicht komplett aus Nullen bestehen, eine Basis von U. *Begründung*: Nach Proposition 5.4 wird U von den Zeilen der Zeilenstufenform erzeugt, also auch durch die Zeilen $\neq 0$. Außerdem sieht man sofort, dass die Zeilen $\neq 0$ einer Matrix in Zeilenstufenform immer linear unabhängig sind.

Es folgt insbesondere: $\dim(U) = \operatorname{rg}(A)$. Damit haben wir bewiesen:

Proposition 6.12. *Der Rang einer Matrix* $A \in K^{m \times n}$ *ist die Dimension des von den Zeilen aufgespannten Unterraums von* $K^{1 \times n}$*.*

Hiermit haben wir für den Rang eine nicht-prozedurale Charakterisierung gefunden. Hierdurch ist die Lücke, die sich durch Definition 2.8 ergeben hat, geschlossen. Eine weitere Charakterisierung des Rangs ist bereits in Proposition 6.11 enthalten. Auch diese zeigt die eindeutige Bestimmtheit des Rangs.

Wir ziehen noch ein paar weitere Folgerungen aus Lemma 6.7. Die erste ermöglicht in vielen Fällen, die Basiseigenschaft zu verifizieren oder zu falsifizieren.

Korollar 6.13. *Es seien* $v_1, \ldots, v_n \in V$ *paarweise verschieden und* $S = \{v_1, \ldots, v_n\}$*. Dann gelten:*

(a) *S ist eine Basis von V* $\Longleftrightarrow \dim(V) = n$ *und S ist linear unabhängig* $\Longleftrightarrow \dim(V) = n$ *und* $V = \langle S \rangle$*.*
(b) *Falls* $n < \dim(V)$*, so folgt* $V \neq \langle S \rangle$*.*
(c) *Falls* $n > \dim(V)$*, so ist S linear abhängig.*

Beweis. (a) Falls S eine Basis ist, so folgt aus Korollar 6.8 und Definition 6.1, dass $\dim(V) = n$, $V = \langle S \rangle$, und dass S linear unabhängig ist. Ist umgekehrt $\dim(V) = n$ und S linear unabhängig, so folgt aus Lemma 6.7, dass S maximal linear unabhängig ist, also ist S nach Satz 6.3 eine Basis. Falls $\dim(V) = n$ und $V = \langle S \rangle$, so folgt aus Lemma 6.7, dass S ein minimales Erzeugendensystem ist, also ist S nach Satz 6.3 eine Basis.

(b) Falls $n < \dim(V)$, so gibt es eine linear unabhängige Menge $U \subseteq V$ mit $|S| < |U|$. Nach Lemma 6.7 kann S kein Erzeugendensystem sein.

(c) Falls $n > \dim(V)$, so gibt es eine Basis $B \subseteq V$ mit $|B| < |S|$. Nach Lemma 6.7 kann S nicht linear unabhängig sein. $\square$

Korollar 6.14. *Es sei V endlich-dimensional und $S \subseteq V$ linear unabhängig. Dann gibt es eine Basis B von V mit $S \subseteq B$. B heißt eine* Basisergänzung *von S.*

Beweis. Wegen Lemma 6.7 hat jede linear unabhängige Menge in V höchstens $\dim(V)$ Elemente. Wir können also innerhalb der linear unabhängigen Mengen, die S enthalten, eine aussuchen, die maximale Elementanzahl hat. Diese Menge B ist dann eine maximal linear unabhängige Menge, also nach Satz 6.3 eine Basis. $\square$

Beispiel 6.15. Der Vektor $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \mathbb{R}^2$ ist linear unabhängig. Er lässt sich durch $v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{R}^2$ zu einer Basis von $\mathbb{R}^2$ ergänzen. $\triangleleft$

Anmerkung. Korollar 6.14 gilt auch, wenn V unendlich-dimensional ist (ohne Beweis). $\triangleleft$

Korollar 6.16. *Es sei $U \subseteq V$ ein Unterraum. Dann gelten:*

(a) $\dim(U) \leq \dim(V)$.
(b) Falls $\dim(U) = \dim(V) < \infty$*, so folgt $U = V$.*

Beweis. Im Fall $\dim(V) = \infty$ ist nichts zu zeigen, wir können also $\dim(V) < \infty$ annehmen. Nach Korollar 6.14 können wir jede linear unabhängige Teilmenge $S \subseteq U$ zu einer Basis B von V ergänzen, also

$$|S| \leq |B| = \dim(V).$$

Wegen dieser „globalen Beschränktheit“ gibt es eine maximal linear unabhängige Teilmenge $S_{\max} \subseteq U$, die wegen Satz 6.3 eine Basis ist. Auch für $S_{\max}$ gilt $|S_{\max}| \leq \dim(V)$. Dies ergibt (a). Falls $\dim(U) = \dim(V)$, so folgt $S_{\max} = B$, also $U = \langle S_{\max} \rangle = \langle B \rangle = V$. $\square$

Kapitel 7
Lineare Abbildungen

Auch in diesem Kapitel sei K ein Körper. Weiter seien V und W zwei K-Vektorräume (über demselben Körper K!).

Definition 7.1. *Eine Abbildung φ: $V \to W$ heißt* **linear**, *falls gelten:*

(1) Für alle $v, v' \in V$: $\varphi(v + v') = \varphi(v) + \varphi(v')$. (Hierbei ist das „+“ auf der linken Seite das von V, das auf der rechten das von W.)
(2) Für alle $v \in V$ und $a \in K$: $\varphi(a \cdot v) = a \cdot \varphi(v)$.

Insbesondere bildet eine lineare Abbildung den Nullvektor von V auf den Nullvektor von W ab.

Beispiel 7.2. (1) Sei $A \in K^{m\times n}$. Dann ist

$$\varphi_A\colon K^n \to K^m,\ v \mapsto A \cdot v$$

eine lineare Abbildung. Dies ist einer der wichtigsten Typen von linearen Abbildungen. Die Bezeichnung φ_A werden wir in Zukunft weiter benutzen.
(2) Die *Nullabbildung* $V \to W$, $v \mapsto 0$ ist linear.
(3) Die folgenden geometrisch definierten Abbildungen $\mathbb{R}^2 \to \mathbb{R}^2$ sind linear: Drehungen um den Nullpunkt, Streckungen mit dem Nullpunkt als Zentrum und Spiegelungen an einer durch den Nullpunkt gehenden Geraden. Drehungen um Punkte $\neq 0$ und Verschiebungen sind *nicht* linear.
(4) Für $V = \mathbb{R}[x]$ ist

$$\varphi\colon V \to V,\ f \mapsto f' \quad \text{(Ableitung)}$$

linear. Ebenso ist $\psi\colon V \to \mathbb{R},\ f \mapsto f(1)$ linear.
(5) Für $V = K^n$ und $i \in \{1, \ldots, n\}$ ist

$$\pi_i\colon V \to K,\ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_i$$

linear. Man bezeichnet π_i als das i-te *Koordinatenfunktional*.

(6) Es sei M eine Menge und $x_1, \dots, x_n \in M$ irgendwelche (fest gewählten) Elemente. Dann ist

$$\varphi\colon V := \mathrm{Abb}(M, K) \to K^n,\ f \mapsto \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}$$

linear. ◁

Sind $\varphi, \psi\colon V \to W$ linear, so gilt dies auch für

$$\varphi + \psi\colon V \to W,\ v \mapsto \varphi(v) + \psi(v).$$

Außerdem ist für ein $a \in K$ auch

$$a \cdot \varphi\colon V \to W,\ v \mapsto a \cdot \varphi(v)$$

linear. Dies bedeutet, dass die Menge $\mathrm{Hom}(V, W)$ aller linearer Abbildungen $V \to W$ einen K-Vektorraum bildet.

Weiter gilt: Sind $\varphi\colon V \to W$ und $\psi\colon W \to U$ (mit U ein weiterer K-Vektorraum) linear, so gilt dies auch für das Kompositum (= „Hintereinanderausführung") $\psi \circ \varphi\colon V \to U$. Damit wird $\mathrm{Hom}(V, V)$ sogar zu einem Ring. (Wir werden sehen, dass dieser für $\dim(V) \geq 2$ nicht-kommutativ ist.)

Definition 7.3. *Es sei $\varphi\colon V \to W$ linear. Der* **Kern** *von φ ist die Menge*

$$\mathrm{Kern}(\varphi) := \{v \in V \mid \varphi(v) = 0\} \subseteq V.$$

Das **Bild** *von φ ist*

$$\mathrm{Bild}(\varphi) := \varphi(V) = \{\varphi(v) \mid v \in V\} \subseteq W.$$

Satz 7.4. *Es sei $\varphi\colon V \to W$ eine lineare Abbildung.*

(a) $\mathrm{Kern}(\varphi) \subseteq V$ *ist ein Unterraum.*
(b) $\mathrm{Bild}(\varphi) \subseteq W$ *ist ein Unterraum.*
(c) Es gilt die Äquivalenz:

$$\varphi \textit{ ist injektiv} \quad \Longleftrightarrow \quad \mathrm{Kern}(\varphi) = \{0\}.$$

Beweis. (a) Der Nullvektor von V ist in $\mathrm{Kern}(\varphi)$ enthalten. Für $v, v' \in \mathrm{Kern}(\varphi)$ gilt $\varphi(v + v') = \varphi(v) + \varphi(v') = 0$, also $v + v' \in \mathrm{Kern}(\varphi)$. Weiter gilt für $v \in \mathrm{Kern}(\varphi)$ und $a \in K$: $\varphi(a \cdot v) = a \cdot \varphi(v) = a \cdot 0 = 0$, also $a \cdot v \in \mathrm{Kern}(\varphi)$. Insgesamt folgt (a).

(b) folgt durch einfaches Nachrechnen.

(c) Zunächst sei φ injektiv. Für $v \in \mathrm{Kern}(\varphi)$ gilt $\varphi(v) = 0 = \varphi(0)$, also $v = 0$. Da umgekehrt $0 \in \mathrm{Kern}(\varphi)$, folgt $\mathrm{Kern}(\varphi) = \{0\}$.

Nun setzen wir umgekehrt $\mathrm{Kern}(\varphi) = \{0\}$ voraus. Für den Injektivitätsnachweis seien $v, v' \in V$ mit $\varphi(v) = \varphi(v')$. Hieraus folgt

$$\varphi(v - v') = \varphi(v) - \varphi(v') = 0,$$

also $v - v' \in \mathrm{Kern}(\varphi)$. Nach Voraussetzung erhalten wir $v - v' = 0$, also $v = v'$. Damit ist gezeigt, dass φ injektiv ist. □

Beispiel 7.5. (1) Sei $A \in K^{m\times n}$. Dann ist $\mathrm{Kern}(\varphi_A)$ die Lösungsmenge des homogenen LGS $A \cdot x = 0$. Also: φ_A ist injektiv $\Longleftrightarrow \mathrm{rg}(A) = n$.

(2) Sei $V = \mathbb{R}[x]$ und $\varphi\colon V \to V,\ f \mapsto f'$ (Ableitung). $\mathrm{Kern}(\varphi)$ ist die Menge aller konstanter Polynome. (Wie wir wissen) ist φ nicht injektiv. Es gilt $\mathrm{Bild}(\varphi) = V$. ◁

Definition 7.6. *Eine lineare Abbildung $\varphi\colon V \to W$ heißt* **Isomorphismus**, *falls φ bijektiv ist. Dann ist auch die Umkehrabbildung $\varphi^{-1}\colon W \to V$ ein Isomorphismus. V und W heißen* **isomorph**, *falls es einen Isomorphismus $V \to W$ gibt. Notation: $V \cong W$.*

Betrachten wir einen K-Vektorraum V mit $n = \dim(V) < \infty$. Nachdem wir eine Basis $B = \{v_1, \ldots, v_n\}$ von V gewählt haben, können wir die lineare Abbildung

$$\varphi\colon K^n \to V,\quad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \mapsto \sum_{i=1}^{n} a_i v_i$$

definieren. Die lineare Unabhängigkeit von B liefert $\mathrm{Kern}(\varphi) = \{0\}$, also ist φ nach Satz 7.4(c) injektiv. Da B ein Erzeugendensystem ist, folgt die Surjektivität von φ. Also ist φ ein Isomorphismus. Die Umkehrabbildung ist dadurch gegeben, dass jedem $v \in V$ sein **Koordinatenvektor** bezüglich B zugewiesen wird, also der eindeutig bestimmte Vektor $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in K^n$ mit $v = \sum_{i=1}^{n} a_i v_i$. Wir haben bewiesen:

Satz 7.7. *Es sei $n := \dim(V) < \infty$. Dann gilt*

$$V \cong K^n.$$

Beispiel 7.8. $V = \{f \in K[x] \mid \deg(f) < 3\} \cong K^3$. Ein Isomorphismus wird gegeben durch

$$\varphi\colon K^3 \to V,\quad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \mapsto a_1 + a_2 x + a_3 x^2.$$

◁

Der Isomorphismus aus Satz 7.7 kann immer erst nach Wahl einer Basis angegeben werden. Man spricht auch von einem *nicht kanonischen* Isomorphismus. Satz 7.7 besagt, dass man sich beim Studium von endlich-dimensionalen Vektorräumen immer auf den Fall $V = K^n$ zurückziehen kann.

Satz 7.9 (Dimensionssatz für lineare Abbildungen)**.** *Sei $\varphi\colon V \to W$ linear. Dann gilt:*

$$\dim(V) = \dim(\mathrm{Kern}(\varphi)) + \dim(\mathrm{Bild}(\varphi)).$$

Beweis. Wir betrachten nur den Fall, dass $\mathrm{Kern}(\varphi)$ und $\mathrm{Bild}(\varphi)$ endlich-dimensional sind. (Der allgemeine Fall geht genauso, benötigt aber aufwändigere Notation.) Es seien $\{w_1, \dots, w_n\}$ eine Basis von $\mathrm{Bild}(\varphi)$ und $\{v_1, \dots, v_m\}$ eine Basis von $\mathrm{Kern}(\varphi)$. Wir können $v_1', \dots, v_n' \in V$ wählen mit $\varphi(v_i') = w_i$. Behauptung: $B := \{v_1, \dots, v_m, v_1', \dots, v_n'\}$ ist eine Basis von V.

Zum Nachweis der linearen Unabhängigkeit sei

$$a_1v_1 + \cdots + a_mv_m + b_1v_1' + \cdots + b_nv_n' = 0 \tag{7.1}$$

mit $a_i, b_i \in K$. Anwendung von φ auf (7.1) liefert:

$$0 = \varphi(0) = \sum_{i=1}^{m} a_i\varphi(v_i) + \sum_{i=1}^{n} b_i\varphi(v_i') = \sum_{i=1}^{n} b_iw_i.$$

Wegen der linearen Unabhängigkeit der w_i liefert dies $b_1 = \cdots = b_n = 0$. Nun folgt aus (7.1)

$$a_1v_1 + \cdots + a_mv_m = 0,$$

also auch $a_1 = \cdots = a_m = 0$.

Für den Nachweis, dass B ein Erzeugendensystem ist, sei $v \in V$ beliebig. Wegen $\varphi(v) \in \mathrm{Bild}(\varphi)$ können wir v schreiben als $\varphi(v) = \sum_{i=1}^n b_iw_i$ mit $b_i \in K$. Mit $\widetilde{v} := v - \sum_{i=1}^n b_iv_i'$ folgt

$$\varphi(\widetilde{v}) = \varphi(v) - \sum_{i=1}^{n} b_i\varphi(v_i') = \varphi(v) - \sum_{i=1}^{n} b_iw_i = 0,$$

also $\widetilde{v} \in \mathrm{Kern}(\varphi)$. Damit gibt es $a_1, \dots, a_m \in K$, so dass

$$\widetilde{v} = a_1v_1 + \cdots + a_mv_m.$$

Insgesamt erhalten wir

$$v = \widetilde{v} + \sum_{i=1}^{n} b_iv_i' = \sum_{i=1}^{m} a_iv_i + \sum_{i=1}^{n} b_iv_i',$$

also $v \in \langle B\rangle$.

Wir haben nachgewiesen, das B eine Basis von V ist, also $\dim(V) = |B| = m + n = \dim(\mathrm{Kern}(\varphi)) + \dim(\mathrm{Bild}(\varphi))$. □

Wir betrachten jetzt eine durch eine Matrix $A \in K^{m\times n}$ gegebene lineare Abbildung $\varphi_A\colon K^n \to K^m$, $v \mapsto A\cdot v$. Nach Proposition 6.11 hat $\mathrm{Kern}(\varphi_A)$ die Dimension $n - \mathrm{rg}(A)$. Satz 7.9 liefert $n = \dim(\mathrm{Kern}(\varphi_A)) + \dim(\mathrm{Bild}(\varphi_A))$, also folgt $\dim(\mathrm{Bild}(\varphi_A)) = \mathrm{rg}(A)$. Was ist $\mathrm{Bild}(\varphi_A)$? Das Bild besteht genau aus allen Linearkombinationen der Spalten von A. Damit haben wir bewiesen:

Korollar 7.10. *Der Rang einer Matrix $A \in K^{m\times n}$ ist die Dimension des von den Spalten aufgespannten Unterraums von K^m.*

Der Vergleich mit Proposition 6.12 ist besonders interessant! Die durch Proposition 6.12 und Korollar 7.10 gegebenen Interpretationen des Rangs laufen unter der Merkregel

$$\boxed{\text{„Zeilenrang“} = \text{„Spaltenrang“}.}$$

Korollar 7.11. *Es gelte $\dim(V) = \dim(W) < \infty$, und $\varphi\colon V \to W$ sei eine lineare Abbildung. Dann sind äquivalent:*

(a) φ ist ein Isomorphismus.
(b) φ ist injektiv.
(c) φ ist surjektiv.

Beweis. Es wird behauptet, dass in der betrachteten Situation Injektivität und Surjektivität von φ äquivalent sind. Nach Satz 7.4(c) ist Injektivität gleichbedeutend mit $\mathrm{Kern}(\varphi) = \{0\}$, also mit $\dim(\mathrm{Kern}(\varphi)) = 0$. Wegen Satz 7.9 ist $\dim(\mathrm{Bild}(\varphi)) = \dim(V) - \dim(\mathrm{Kern}(\varphi)) = \dim(W) - \dim(\mathrm{Kern}(\varphi))$. Also ist φ genau dann injektiv, wenn $\dim(\mathrm{Bild}(\varphi)) = \dim(W)$. Dies ist wegen Korollar 6.16(b) gleichbedeutend mit $\mathrm{Bild}(\varphi) = W$, also mit der Surjektivität von φ. □

Es sei $A \in K^{n\times n}$. Wenn wir Korollar 7.11 auf φ_A anwenden, erhalten wir

$$\varphi_A \text{ ist ein Isomorphismus} \iff \mathrm{rg}(A) = n.$$

In diesem Fall liefert das folgende Verfahren eine inverse Matrix $B \in K^{n\times n}$ mit $AB = I_n$.

(1) Bilde die „erweiterte“ Matrix $(A|I_n) \in K^{n\times(2n)}$ durch Anhängen einer Einheitsmatrix.
(2) Führe diese (mit dem Gauß-Algorithmus) über in strenge Zeilenstufenform, so dass zusätzlich in jeder Zeile $\neq 0$ der erste Eintrag $\neq 0$ eine 1 ist.
(3) 1. Fall: Die Zeilenstufenform hat die Gestalt $(I_n|B)$ mit $B \in K^{n\times n}$: Dann gilt $AB = I_n$, und wir sind fertig.
2. Fall: Die Zeilenstufenform hat eine andere Gestalt: Dann ist $\mathrm{rg}(A) < n$, also gibt es kein $B \in K^{n\times n}$ mit $AB = I_n$.

Die Korrektheit des Algorithmus begründen wir wie folgt: Es werden simultan die LGSe $A \cdot x = e_i$ (i-ter Standardbasisvektor) gelöst. Der erste Fall ist der Fall eindeutiger Lösbarkeit. Dann sind die Spalten von B jeweils die Lösungsvektoren, und es folgt $A \cdot B = I_n$. Bevor wir ein Beispiel bringen, machen wir eine Definition.

Definition 7.12. *Eine quadratische Matrix $A \in K^{n \times n}$ heißt* **invertierbar**, *falls es $B \in K^{n \times n}$ gibt mit $A \cdot B = I_n$. Wie wir später sehen werden, ist B dann eindeutig bestimmt, und es gilt auch $B \cdot A = I_n$. B heißt die* **Inverse** *von A und wird als $B = A^{-1}$ geschrieben.*

Für $A \in K^{n \times n}$ gilt also die Äquivalenz

$$A \text{ invertierbar} \quad \Longleftrightarrow \quad A \text{ regulär.}$$

Beispiel 7.13. Wir möchten die Matrix $A = \begin{pmatrix} 1 & -2 & 0 \\ -1 & 3 & -2 \\ -1 & 2 & -1 \end{pmatrix}$ invertieren. Obiges Verfahren läuft wie folgt ab:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & 0 \\ -1 & 3 & -2 & 0 & 1 & 0 \\ -1 & 2 & -1 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 \end{array}\right) \longrightarrow$$
$$\left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & -2 \\ 0 & 0 & -1 & 1 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 2 & -4 \\ 0 & 1 & 0 & -1 & 1 & -2 \\ 0 & 0 & 1 & -1 & 0 & -1 \end{array}\right),$$

also $A^{-1} = \begin{pmatrix} -1 & 2 & -4 \\ -1 & 1 & -2 \\ -1 & 0 & -1 \end{pmatrix}$. Per Probe-Multiplikation prüft man leicht $A \cdot A^{-1} = A^{-1} \cdot A = I_3$ nach. ◁

Zum Abschluss dieses Kapitels beweisen wir einen Satz, der im folgenden Kapitel eine wichtige Rolle spielen wird.

Satz 7.14 (lineare Fortsetzung)**.** *Es sei $B = \{v_1, \ldots, v_n\}$ eine Basis von V.*

(a) Eine lineare Abbildung $\varphi\colon V \to W$ ist durch die Bilder der Basisvektoren v_i eindeutig bestimmt. Mit anderen Worten: Ist $\psi\colon V \to W$ eine weitere lineare Abbildung mit $\varphi(v_i) = \psi(v_i)$ für alle i, so folgt $\varphi = \psi$.
(b) Seien $w_1, \ldots, w_n \in W$ beliebig. Dann gibt es eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(v_i) = w_i$ für alle i.

Zusammengefasst: Man kann lineare Abbildungen eindeutig definieren, indem man die Bilder der Basisvektoren angibt. Dies nennt man das Prinzip der linearen Fortsetzung.

Beweis. (a) Es gelte $\varphi(v_i) = \psi(v_i)$ für alle i. Sei $v \in V$. Dann gibt es $a_1, \ldots, a_n \in K$ mit $v = \sum_{i=1}^n a_i v_i$, also

$$\varphi(v) = \varphi\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i \varphi(v_i) = \sum_{i=1}^n a_i \psi(v_i) = \psi\left(\sum_{i=1}^n a_i v_i\right) = \psi(v).$$

Dies bedeutet $\varphi = \psi$.

(b) Wir definieren $\varphi \colon V \to W$ folgendermaßen: Für $v \in V$ sei $v = \sum_{i=1}^n a_i v_i$ mit $a_i \in K$. Dann setzen wir

$$\varphi(v) := \sum_{i=1}^n a_i w_i.$$

Die eindeutige Darstellungseigenschaft von B liefert die Wohldefiniertheit von φ. Die Linearität ergibt sich durch einfaches Nachprüfen. Außerdem gilt nach Konstruktion $\varphi(v_i) = w_i$. $\square$

Kapitel 8
Darstellungsmatrizen

In diesem Kapitel seien K ein Körper, V und W endlich-dimensionale K-Vektorräume und $B = \{v_1, \ldots, v_n\}$ bzw. $C = \{w_1, \ldots, w_m\}$ Basen von V bzw. von W. Für das Folgende ist die *Reihenfolge* der Basisvektoren wichtig. Wir könnten dies zum Ausdruck bringen, indem wir als neues mathematisches Objekt eine *geordnete Basis* einführen, etwa als ein Element des n-fachen kartesischen Produkts $V \times \cdots \times V$ (mit gewissen Zusatzeigenschaften). Wir werden aber davon absehen, solchen begrifflichen und notationstechnischen Aufwand zu betreiben.

Nun sei $\varphi\colon V \to W$ eine lineare Abbildung. Für $j \in \{1, \ldots, n\}$ können wir schreiben:

$$\varphi(v_j) = \sum_{i=1}^{m} a_{i,j} w_i$$

mit $a_{i,j} \in K$. Nun bilden wir die Matrix

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \in K^{m \times n}.$$

Die Spalten von A sind also die Koordinatenvektoren der $\varphi(v_i)$.

Definition 8.1. *Die oben definierte Matrix A heißt die* **Darstellungsmatrix** *von φ (bezüglich der Basen B und C). Schreibweise:*

$$A = D_{B,C}(\varphi).$$

Falls $V = W$ gilt, so verwendet man dieselbe Basis $B = C$ und schreibt $D_B(\varphi) \in K^{n \times n}$.

Als Merkregel halten wir fest:

$$\boxed{\text{Spalten der Darstellungsmatrix} \quad \longleftrightarrow \quad \text{Bilder der Basisvektoren}}$$

Es sei angemerkt, dass unsere Schreibweise für Darstellungsmatrizen nicht allgemein gebräuchlich ist.

Wegen Satz 7.14 ist φ durch seine Darstellungsmatrix eindeutig bestimmt, und jede Matrix taucht als Darstellungsmatrix einer linearen Abbildung auf.

Beispiel 8.2. (1) Es sei $V = W = \mathbb{R}^2$ mit Basis $B = \{e_1, e_2\}$, und $\varphi\colon V \to V$ sei eine Drehung um $60°$ nach links. Wir haben

$$\varphi(e_1) = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix} = 1/2 e_1 + \sqrt{3}/2 e_2,$$
$$\varphi(e_2) = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix} = -\sqrt{3}/2 e_1 + 1/2 e_2,$$

also

$$D_B(\varphi) = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}.$$

(2) Es sei $V = \{f \in \mathbb{R}[x] \mid \deg(f) < 3\}$ mit Basis $B = \{1, x, x^2\}$. Für $\varphi\colon V \to V,\ f \mapsto f'$ (Ableitung) erhalten wir

$$\varphi(1) = 0,\ \varphi(x) = 1 \quad \text{und} \quad \varphi(x^2) = 2x,$$

also

$$D_B(\varphi) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

◁

In Beispiel 7.2(1) haben wir mit Hilfe einer Matrix eine lineare Abbildung $K^n \to K^m$ definiert, also bereits eine Zuordnung zwischen Matrizen und linearen Abbildungen hergestellt. Besteht zwischen dieser Zuordnung und Definition 8.1 ein Zusammenhang?

Satz 8.3. *Gegeben seien $V = K^n$ und $W = K^m$ mit den Standardbasen B und C, und eine lineare Abbildung $\varphi\colon V \to W$. Mit $A := D_{B,C}(\varphi)$ gilt dann*

$$\varphi = \varphi_A.$$

Insbesondere sind alle linearen Abbildungen $V \to W$ von der Form φ_A mit $A \in K^{m \times n}$, und A ist die Darstellungsmatrix von φ_A bezüglich der Standardbasen.

Beweis. Wir schreiben $A = (a_{i,j})$ und rechnen nach: Für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{j=1}^n x_j e_j$ ist

$$\varphi(v) = \sum_{j=1}^n x_j \varphi(e_j) = \sum_{j=1}^n \left(x_j \cdot \sum_{i=1}^m a_{i,j} e_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{i,j} x_j \right) e_i = A \cdot v,$$

also $\varphi = \varphi_A$. □

Wir wissen, dass das Kompositum von linearen Abbildungen wieder linear ist. Damit ergibt sich die Frage: Was passiert mit den Darstellungsmatrizen bei Bildung des Kompositums?

Satz 8.4. *Es seien U, V und W endlich-dimensionale K-Vektorräume mit Basen A, B bzw. C, und es seien $\varphi\colon U \to V$ und $\psi\colon V \to W$ lineare Abbildungen. Dann gilt*

$$D_{A,C}(\psi \circ \varphi) = D_{B,C}(\psi) \cdot D_{A,B}(\varphi).$$

Als Merkregel halten wir fest:

$$\boxed{\text{Kompositum von linearen Abbildungen} \quad \longleftrightarrow \quad \text{Matrixprodukt}}$$

Beweis. Wir müssen zunächst Bezeichnungen einführen. Wir schreiben $A = \{u_1, \ldots, u_n\}$, $B = \{v_1, \ldots, v_m\}$, $C = \{w_1, \ldots, w_l\}$ und

$$D_{B,C}(\psi) = (a_{i,j}) \in K^{l \times m}, \ D_{A,B}(\varphi) = (b_{i,j}) \in K^{m \times n}.$$

Für $j \in \{1, \ldots, n\}$ gilt:

$$(\psi \circ \varphi)(u_j) = \psi \left(\sum_{k=1}^m b_{k,j} v_k \right) = \sum_{k=1}^m b_{k,j} \psi(v_k) = \sum_{k=1}^m \left(b_{k,j} \sum_{i=1}^l a_{i,k} w_i \right) = \sum_{i=1}^l \left(\sum_{k=1}^m a_{i,k} b_{k,j} \right) w_i.$$

Aus der Beobachtung, dass im letzten Ausdruck der Koeffizient von w_i genau der (i,j)-te Eintrag des Produkts $D_{B,C}(\psi) \cdot D_{A,B}(\varphi)$ ist, folgt die Behauptung. □

Der obige Satz liefert einen weiteren Beleg dafür, dass unsere Definition des Matrixprodukts sinnvoll war. Man könnte auch sagen: Das Matrixprodukt ist genau so definiert, dass Satz 8.4 gilt.

Kombiniert man den Satz mit Satz 8.3, so erhält man für Matrizen $A \in K^{l\times m}$ und $B \in K^{m\times n}$ (und die dadurch definierten Abbildungen $\varphi_A\colon K^m \to K^l$ und $\varphi_B\colon K^n \to K^m$):

$$\varphi_A \circ \varphi_B = \varphi_{A\cdot B}.$$

Ist insbesondere $A \in K^{n\times n}$ invertierbar, so folgt

$$\varphi_A \circ \varphi_{A^{-1}} = \varphi_{I_n} = \mathrm{id}_{K^n}$$

(die identische Abbildung von K^n), also ist

$$\varphi_{A^{-1}} = \varphi_A^{-1}$$

die *Umkehrabbildung* von φ_A. Hieraus folgt, dass A^{-1} die Darstellungsmatrix von φ_A^{-1} bezüglich der Standardbasis ist, was die *Eindeutigkeit* der inversen Matrix liefert. Außerdem gilt für die Umkehrabbildung auch $\varphi_A^{-1}\circ\varphi_A = \mathrm{id}_{K^n}$, was sich zu

$$A^{-1} \cdot A = I_n$$

übersetzt. Hiermit sind zwei Lücken geschlossen, die in Definition 7.12 entstanden waren.

Definition 8.5. *Die Menge*

$$\mathrm{GL}_n(K) := \left\{A \in K^{n\times n} \mid A \textit{ ist invertierbar}\right\}$$

heißt die **allgemeine lineare Gruppe**. *(Die Buchstaben GL erklären sich aus der englischen Bezeichnung* general linear group.*) Wir wissen, dass* $\mathrm{GL}_n(K)$ *zusammen mit dem Matrixprodukt tatsächlich eine Gruppe bildet.*

Wir wissen, dass Vektorräume verschiedene Basen haben. Was passiert mit der Darstellungsmatrix einer linearen Abbildung $V \to V$, wenn man die Basis von V wechselt?

Es sei $B = \{v_1, \ldots, v_n\}$ eine Basis von V, und $B' = \{v_1', \ldots, v_n'\}$ sei eine weitere Basis. Wir können die „neuen“ Basisvektoren v_j' mit Hilfe der alten ausdrücken:

$$v_j' = \sum_{i=1}^{n} a_{i,j} v_i \tag{8.1}$$

mit $a_{i,j} \in K$. Hieraus können wir die Matrix $S := (a_{i,j}) \in K^{n\times n}$ bilden. S heißt die **Basiswechselmatrix**. Sie beschreibt den Übergang von B zu B'. Man schreibt bisweilen $S =: S_{B,B'}$. Die Basiswechselmatrix wird nach folgender Merkregel gebildet:

$$\boxed{\text{Spalten von } S \quad = \quad \text{Koordinatenvektoren der „neuen“ Basisvektoren}}$$

Man kann auch umgekehrt die v_j mit Hilfe der v_i' ausdrücken: $v_j = \sum_{i=1}^n b_{i,j} v_i'$ mit $b_{i,j} \in K$. Wir setzen $T := (b_{i,j}) \in K^{n \times n}$. Für alle $j \in \{1, \ldots, n\}$ folgt:

$$v_j = \sum_{i=1}^n b_{i,j} \left(\sum_{k=1}^n a_{k,i} v_k \right) = \sum_{k=1}^n \left(\sum_{i=1}^n a_{k,i} b_{i,j} \right) v_k.$$

Den in der rechten Klammer stehenden Ausdruck erkennen wir als den (k,j)-te Eintrag des Matrixprodukts $S \cdot T$. Aus der Gleichung folgt (wegen der linearen Unabhängigkeit von B), dass $S \cdot T = I_n$ gelten muss, also $T = S^{-1}$.

Wir bemerken noch, dass jede invertierbare Matrix $S = (a_{i,j}) \in \mathrm{GL}_n(K)$ einen Basiswechel beschreibt, indem man die neue Basis einfach durch (8.1) definiert.

Wir kehren zurück zu unserer Ausgangsfrage und betrachten eine lineare Abbildung $\varphi \colon V \to V$. Wir schreiben $D_B(\varphi) = (d_{i,j}) \in K^{n \times n}$ und möchten nun $D_{B'}(\varphi)$ bestimmen. Dazu rechnen wir

$$\varphi(v_j') = \varphi \left(\sum_{i=1}^n a_{i,j} v_i \right) = \sum_{i=1}^n a_{i,j} \varphi(v_i) = \sum_{i=1}^n a_{i,j} \left(\sum_{k=1}^n d_{k,i} v_k \right) =$$
$$\sum_{i,k=1}^n d_{k,i} a_{i,j} \left(\sum_{l=1}^n b_{l,k} v_l' \right) = \sum_{l=1}^n \left(\sum_{i,k=1}^n b_{l,k} d_{k,i} a_{i,j} \right) v_l'.$$

Den in der rechten Klammer stehenden Ausdruck erkennen wir als den (l,j)-te Eintrag des Matrixprodukts $T \cdot D_B(\varphi) \cdot S$. Aus der Gleichung folgt, dass dieser Ausdruck andererseits der (l,j)-te Eintrag der Darstellungsmatrix $D_{B'}(\varphi)$ sein muss. Damit haben wir gezeigt:

Satz 8.6. *Es seien B und B' Basen eines endlich-dimensionalen K-Vektorraums V und $S := S_{B,B'}$ die Basiswechselmatrix. Dann gilt für eine lineare Abbildung $\varphi \colon V \to V$:*

$$\boxed{D_{B'}(\varphi) = S^{-1} \cdot D_B(\varphi) \cdot S.}$$

Dieser Satz beantwortet die Frage, was bei Wechsel der Basis mit der Darstellungsmatrix passiert. Für lineare Abbildungen erhalten wir durch ein ganz entsprechendes (aber notationstechnisch aufwändigeres) Argument:

Satz 8.7. *Es seien B, B' endliche Basen von V und C, C' endliche Basen von W. Dann gilt für eine linear Abbildung $\varphi \colon V \to W$:*

$$\boxed{D_{B',C'}(\varphi) = S_{C,C'}^{-1} \cdot D_{B,C}(\varphi) \cdot S_{B,B'}.}$$

Wir nehmen diese beiden Sätze (und die Bemerkung, dass jede invertierbare Matrix einen Basiswechsel vemittelt) zum Anlass für folgende Definition:

Definition 8.8. *(a) Zwei quadratische Matrizen $A, B \in K^{n\times n}$ heißen* **ähnlich**, *falls es $S \in \mathrm{GL}_n(K)$ gibt mit*

$$B = S^{-1}AS.$$

(b) Zwei Matrizen $A, B \in K^{m\times n}$ heißen **äquivalent**, *falls es $S \in \mathrm{GL}_n(K)$ und $T \in \mathrm{GL}_m(K)$ gibt mit*

$$B = T^{-1}AS.$$

Wie man sich leicht überlegt, sind Ähnlichkeit und Äquivalenz Äquivalenzrelationen. Von diesen beiden Begriffen ist die Ähnlichkeit der wichtigere.

Das folgende Beispiel soll einen Hinweis darauf geben, weshalb ein Basiswechsel nützlich sein kann.

Beispiel 8.9. Es seien $V = \mathbb{R}^2$ und $\varphi\colon V \to V$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y \\ x \end{pmatrix}$. Mit der Standardbasis $B = \{e_1, e_2\}$ haben wir

$$D_B(\varphi) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Als neue Basis wählen wir $B' = \{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\}$. Die Basiswechselmatrix und ihre Inverse sind

$$S = S_{B,B'} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{und} \quad S^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Es ergibt sich

$$D_{B'}(\varphi) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Die Darstellungsmatrix $D_{B'}(\varphi)$ beschreibt φ in einfacherer Weise: Der erste Basisvektor wird durch φ festgehalten, der zweite wird „umgeklappt“. ◁

Der Abschnitt sei mit einer eindringlichen Warnung abgeschlossen: Man neigt dazu, Basiswechselmatrizen eine Interpretation als eine Art von Darstellungsmatrizen zu geben, etwa als Matrizen, die eine Basis auf eine andere abbilden. Dies führt immer wieder zu großer Verwirrung. Man sollte sich vor solchen Interpretationen hüten, und den Formalismus des Basiswechsels stattdessen als Rezept, dessen Korrektheit bewiesen wurde, wortwörtlich anwenden.

Kapitel 9
Determinanten

Bevor wir die Determinante definieren, müssen wir uns mit der symmetrischen Gruppe beschäftigen. Wir erinnern an Definition 3.4: Für $n \in \mathbb{N}_{>0}$ ist die **symmetrische Gruppe** definiert als

$$S_n := \{\sigma\colon \{1,\dots,n\} \to \{1,\dots,n\} \mid \sigma \text{ ist bijektiv}\}.$$

Die Elemente von S_n heißen *Permutationen*, und die Verknüpfung ist durch die Komposition gegeben.

Definition 9.1. *Für $\sigma \in S_n$ definieren wir*

- *$w(\sigma)$ als die Anzahl der Paare $(i,j) \in \mathbb{N} \times \mathbb{N}$ mit $1 \le i < j \le n$ aber $\sigma(i) > \sigma(j)$ (solche Paare nennt man auch* Fehlstellen*);*
- *$\operatorname{sgn}(\sigma) := (-1)^{w(\sigma)}$, das* **Vorzeichen** *(oder* **Signum***) von σ.*

Beispiel 9.2. (1) Die Identität $\mathrm{id} \in S_n$ hat keine Fehlstellen, also $\operatorname{sgn}(\mathrm{id}) = 1$.

(2) Es sei $\sigma \in S_n$ gegeben durch $\sigma(1) = 2$, $\sigma(2) = 1$ und $\sigma(i) = i$ für $i > 2$, in Zykelschreibweise (siehe Beispiel 3.5(2)) also $\sigma = (1,2)$. Offenbar ist $(1,2)$ die einzige Fehlstelle von σ, also $\operatorname{sgn}(\sigma) = -1$.

(3) Es seien $1 \le i < j \le n$, und $\sigma \in S_n$ vertausche i und j und halte alle anderen Elemente von $\{1,\dots,n\}$ fest, also $\sigma = (i,j)$. Eine solche Permutation nennt man auch eine *Transposition*. Wir zählen Fehlstellen und kommen auf $w(\sigma) = 2(j-i) - 1$, also $\operatorname{sgn}(\sigma) = -1$. ◁

Die wichtigste Eigenschaft des Vorzeichens ist seine Multiplikativität:

Satz 9.3. *Für $\sigma, \tau \in S_n$ gilt*

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).$$

Beweis. Es seien $x_1, \dots, x_n \in \mathbb{Q}$ paarweise verschiedene rationale Zahlen. Wir behaupten, dass für alle $\sigma \in S_n$ gilt:

$$\operatorname{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{x_{\sigma(i)} - x_{\sigma(j)}}{x_i - x_j}. \tag{9.1}$$

Um dies einzusehen bemerken wir zunächst, dass Zähler und Nenner des Produkts bis auf das Vorzeichen übereinstimmen. Im Zähler tritt aber genau $w(\sigma)$ mal ein $x_k - x_l$ mit $k > l$ auf, während dies im Nenner nie vorkommt. Hieraus ergibt sich (9.1).

Nun setzen wir $y_i := x_{\sigma(i)}$. Ebenso wie die x_i sind auch die y_i paarweise verschieden, also gilt wegen (9.1) für alle $\tau \in S_n$

$$\operatorname{sgn}(\tau) = \prod_{1 \le i < j \le n} \frac{y_{\tau(i)} - y_{\tau(j)}}{y_i - y_j} = \prod_{1 \le i < j \le n} \frac{x_{\sigma\tau(i)} - x_{\sigma\tau(j)}}{x_{\sigma(i)} - x_{\sigma(j)}}.$$

Wir erhalten

$$\operatorname{sgn}(\sigma\tau) = \prod_{1 \le i < j \le n} \frac{x_{\sigma\tau(i)} - x_{\sigma\tau(j)}}{x_i - x_j} = \\ \prod_{1 \le i < j \le n} \frac{x_{\sigma\tau(i)} - x_{\sigma\tau(j)}}{x_{\sigma(i)} - x_{\sigma(j)}} \cdot \prod_{1 \le i < j \le n} \frac{x_{\sigma(i)} - x_{\sigma(j)}}{x_i - x_j} = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma).$$

□

Aus Satz 9.3 folgt direkt, dass $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)^{-1} = \operatorname{sgn}(\sigma)$ für alle $\sigma \in S_n$ gilt.

Nun können wir die Determinante einer quadratischen Matrix definieren. Nebenbei definieren wir auch die weniger wichtige Permanente.

Definition 9.4. *Es sei $A = (a_{i,j}) \in K^{n \times n}$ eine quadratische Matrix.*

(a) Die **Permanente** *von A ist*

$$\operatorname{perm}(A) := \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

(b) Die **Determinante** *von A ist*

$$\det(A) := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Beispiel 9.5. Für $n \le 3$ machen wir Definition 9.4 explizit.

(1) Für $n = 1$ ist $A = (a)$ und

$$\det(A) = \operatorname{perm}(A) = a.$$

(2) Für $n = 2$ ist $S_n = \{\mathrm{id}, \sigma\}$ mit σ aus Beispiel 9.2(2). Wir erhalten

$$\operatorname{perm}\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = a_{1,1}a_{2,2} + a_{1,2}a_{2,1}$$

und

$$\det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

(3) Für $n = 3$ hat die S_n sechs Elemente: die Identität, drei Transpositionen, sowie die „zyklischen“ Permutationen $1 \mapsto 2 \mapsto 3 \mapsto 1$ und $1 \mapsto 3 \mapsto 2 \mapsto 1$. Die zyklischen Permutationen haben Vorzeichen 1. Wir erhalten

$$\det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} = a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2}$$
$$- a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2}.$$

Für die Permanente ist jedes „$-$“ durch „$+$“ zu ersetzen. Es gibt eine graphische Merkeregel für die Determinante einer 3×3-Matrix, die sogenannte *Sarrus-Regel*:

$$\begin{array}{c} \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,1} & a_{3,2} \end{pmatrix} \\ \begin{array}{cccccc} - & - & - & + & + & + \end{array} \end{array}$$

Der Zusammenhang zwischen der obigen Formel und der Graphik dürfte selbsterklärend sein.

(4) Für die Einheitsmatrix I_n gilt: $\det(I_n) = 1$. $\triangleleft$

Ab jetzt behandeln wir nur noch die Determinante, die sehr viel wichtiger ist als die Permanente.

Lemma 9.6. *Sei $A = (a_{i,j}) \in K^{n\times n}$.*

(a) $\det(A^T) = \det(A)$.

(b) Es sei $\sigma \in S_n$. Wir definieren $b_{i,j} := a_{i,\sigma(j)}$ und $B := (b_{i,j}) \in K^{n\times n}$ (d.h. B geht aus A durch Permutation der Spalten gemäß σ hervor). Dann gilt

$$\det(B) = \operatorname{sgn}(\sigma) \cdot \det(A).$$

Entsprechendes gilt für Permutationen der Zeilen. Als Speziallfall bedeutet dies, dass $\det(B) = -\det(A)$, falls B aus A durch Vertauschung zweier Zeilen oder Spalten hervorgeht.

(c) Falls in A zwei Zeilen oder zwei Spalten übereinstimmen, so folgt

$$\det(A) = 0.$$

Beweis. (a) Wir rechnen

$$\det(A^T) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^n a_{\sigma(i),i} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{j=1}^n a_{j,\sigma^{-1}(j)}$$
$$= \sum_{\tau \in S_n} \operatorname{sgn}(\tau^{-1}) \cdot \prod_{j=1}^n a_{j,\tau(j)} = \det(A).$$

(b) Wir rechnen

$$\det(B) = \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \cdot \prod_{i=1}^n b_{i,\tau(i)} = \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \cdot \prod_{i=1}^n a_{i,\sigma\tau(i)}$$
$$= \sum_{\rho \in S_n} \operatorname{sgn}(\sigma^{-1}\rho) \cdot \prod_{i=1}^n a_{i,\rho(i)} = \operatorname{sgn}(\sigma^{-1}) \cdot \det(A),$$

wobei Satz 9.3 für die letzte Gleichheit benutzt wurde. Wegen $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$, also folgt die Behauptung.
Die entsprechende Aussage für Zeilenpermutationen lässt sich durch (a) auf die für Spaltenpermutationen zurückführen.

(c) Wegen (a) ist $\det(A) = 0$ nur für den Fall zweier gleicher Spalten nachzuweisen. Wir nehmen also an, dass es $1 \leq j < k \leq n$ gibt, so dass $a_{i,j} = a_{i,k}$ für alle i gilt. Es sei $\tau \in S_n$ definiert durch $\tau(j) = k$, $\tau(k) = j$, und alle anderen Elemente bleiben fest (siehe Beispiel 9.2(3)). Für alle $i, l \in \{1, \ldots, n\}$ gilt dann
$$a_{i,l} = a_{i,\tau(l)}. \tag{9.2}$$
Wir definieren
$$A_n := \{\sigma \in S_n \mid \operatorname{sgn}(\sigma) = 1\}.$$
(Nebenbei gesagt folgt aus Satz 9.3, dass A_n eine Untergruppe der S_n ist; sie heißt die *alternierende Gruppe*.) Wegen $\operatorname{sgn}(\tau) = -1$ folgt aus Satz 9.3, dass S_n die disjunkte Vereinigung von A_n und $\tau \cdot A_n := \{\tau\sigma \mid \sigma \in A_n\}$ ist:
$$S_n = A_n \mathbin{\dot\cup} \tau \cdot A_n.$$
Nun folgt

$$\det(A) = \sum_{\sigma \in A_n} \left(\operatorname{sgn}(\sigma) \cdot \prod_{i=1}^n a_{i,\sigma(i)} + \operatorname{sgn}(\tau\sigma) \cdot \prod_{i=1}^n a_{i,\tau\sigma(i)} \right)$$
$$= \sum_{\sigma \in A_n} \operatorname{sgn}(\sigma) \cdot \left(\prod_{i=1}^n a_{i,\sigma(i)} - \prod_{i=1}^n a_{i,\tau(\sigma(i))} \right) = 0,$$

wobei (9.2) für die letzte Gleichheit verwendet wurde. □

Der wohl wichtigste Satz über die Determinante ist der folgende.

Satz 9.7 (Determinantenmultiplikationssatz)**.** *Für $A, B \in K^{n \times n}$ gilt*

$$\boxed{\det(A \cdot B) = \det(A) \cdot \det(B).}$$

Beweis. Wie immer schreiben wir $A = (a_{i,j})$ und $B = (b_{i,j})$. Der (i,j)-te Eintrag von $A \cdot B$ ist $\sum_{k=1}^{n} a_{i,k} b_{k,j}$, also

$$\det(A \cdot B) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} \left(\sum_{k=1}^{n} a_{i,k} b_{k,\sigma(i)} \right).$$

Ausmultiplizieren des Produkts und Vertauschung der Summation liefern

$$\begin{aligned}\det(A \cdot B) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot & \sum_{k_1,\ldots,k_n=1}^{n} \prod_{i=1}^{n} \left(a_{i,k_i} b_{k_i,\sigma(i)}\right) \\ &= \sum_{k_1,\ldots,k_n=1}^{n} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^{n} a_{i,k_i} \cdot \prod_{j=1}^{n} b_{k_j,\sigma(j)} = \\ & \qquad \sum_{k_1,\ldots,k_n=1}^{n} \prod_{i=1}^{n} a_{i,k_i} \cdot \det(b_{k_j,l})_{j,l=1,\ldots,n}.\end{aligned}$$

Wegen Lemma 9.6(c) ist $\det(b_{k_j,l})_{j,l=1,\ldots,n}$ nur dann $\neq 0$, wenn die k_j paarweise verschieden sind, d.h. wenn die Abbildung $\{1,\ldots,n\} \to \{1,\ldots,n\}$, $j \mapsto k_j$ eine Permutation ist. Statt über die $k_1,\ldots,k_n$ zu summieren, können wir also auch über die Permutationen $\tau \in S_n$ summieren und erhalten

$$\begin{aligned}\det(A \cdot B) &= \sum_{\tau \in S_n} \prod_{i=1}^{n} a_{i,\tau(i)} \cdot \det(b_{\tau(j),l})_{j,l=1,\ldots,n} \\ &= \sum_{\tau \in S_n} \prod_{i=1}^{n} a_{i,\tau(i)} \cdot \operatorname{sgn}(\tau) \cdot \det(B) = \det(A) \cdot \det(B),\end{aligned}$$

wobei für die zweite Gleichheit Lemma 9.6(b) verwendet wurde. □

Die Determinante ist also multiplikativ. Als Warnung sei hier angemerkt, dass sie nicht additiv ist (außer im Fall $n = 1$)! Wir ziehen eine wichtige Folgerung.

Satz 9.8. *Für $A \in K^{n \times n}$ gilt die Äquivalenz*

$$\boxed{A \text{ ist regulär} \quad \Longleftrightarrow \quad \det(A) \neq 0.}$$

Falls A regulär ist, so folgt

$$\det(A^{-1}) = 1/\det(A).$$

Beweis. Zunächst sei A regulär. Dann gibt es eine Inverse A^{-1}. Nach Satz 9.7 und Beispiel 9.5(4) gilt

$$\det(A^{-1}) \cdot \det(A) = \det(A^{-1} \cdot A) = \det(I_n) = 1,$$

woraus $\det(A) \neq 0$ und $\det(A^{-1}) = 1/\det(A)$ folgen.

Nun nehmen wir an, dass A *nicht* regulär sei. Dann gibt es $v \in K^n \setminus \{0\}$ mit $A \cdot v = 0$. Wir ergänzen v zu einer Basis $\{v = v_1, v_2, \dots, v_n\}$ von K^n. $B := (v_1, \dots, v_n) \in K^{n \times n}$ sei die Matrix mit den v_i als Spalten. Da die Spalten von B linear unabhängig sind, folgt nach Korollar 7.10, dass B regulär ist, also $\det(B) \neq 0$ nach dem bereits Gezeigten. Für den Standardbasisvektor e_1 gilt

$$A \cdot B \cdot e_1 = A \cdot v_1 = A \cdot v = 0.$$

Dies bedeutet, dass die erste Spalte von $A \cdot B$ aus Nullen besteht. Hieraus folgt direkt $\det(A \cdot B) = 0$. Aber Satz 9.7 liefert

$$\det(A \cdot B) = \det(A) \cdot \det(B),$$

Also folgt wegen $\det(B) \neq 0$, dass $\det(A) = 0$ gelten muss. Dies schließt den Beweis ab. □

Für eine quadratische Matrix $A \in K^{n \times n}$ sind damit die folgenden Aussagen äquivalent:

- A ist regulär;
- A ist invertierbar (anders gesagt: $A \in GL_n(K)$);
- die Zeilen von A sind linear unabhängig;
- die Spalten von A sind linear unabhängig;
- die Abbildung φ_A ist injektiv;
- die Abbildung φ_A ist surjektiv;
- das LGS $A \cdot x = 0$ ist eindeutig lösbar.
- für alle $b \in K^n$ ist das LGS $A \cdot x = b$ eindeutig lösbar.
- $\det(A) \neq 0$.

Wir ziehen eine weitere Folgerung aus Satz 9.7.

Korollar 9.9. *Zwei Matrizen $A, B \in K^{n \times n}$ seien ähnlich. Dann gilt*

$$\det(A) = \det(B).$$

Beweis. Wir haben $B = S^{-1}AS$ mit $S \in \mathrm{GL}_n(K)$. Wegen der Sätze 9.7 und 9.8 folgt

$$\det(B) = \det(S)^{-1} \det(A) \det(S) = \det(A).$$

□

Korollar 9.9 hat eine interessante konzeptionelle Interpretation: Ist $\varphi\colon V \to V$ eine lineare Selbstabbildung eines endlich-dimensionalen Vektorraums V, so lässt sich $\det(\varphi)$ nach Wahl einer Basis B von V durch

$$\det(\varphi) := \det\left(D_B(\varphi)\right)$$

definieren. Denn bei einer anderen Basiswahl geht $D_B(\varphi)$ nach Satz 8.6 über in eine ähnliche Matrix.

Definition 9.10. *Die Menge*

$$\mathrm{SL}_n(K) := \left\{A \in K^{n\times n} \mid \det(A) = 1\right\}$$

heißt die **spezielle lineare Gruppe**. *Aus Satz 9.7 folgt, dass* $\mathrm{SL}_n(K)$ *(mit dem Matrizenprodukt) eine Gruppe ist.*

Für den Rest des Kapitels beschäftigen wir uns mit dem effizienten Berechnen der Determinante. Die Definition 9.4 ist explizit, so dass eine direkte Berechnung möglich ist. Sie erfordert jedoch wegen $|S_n| = n!$ etwa $n \cdot n!$ Körperoperationen, ein für große n nicht hinnehmbarer Aufwand. Wir werden ein besseres Verfahren entwickeln.

Satz 9.11. *Es sei* $A = (a_{i,j}) \in K^{n\times n}$ *mit* $n \geq 2$. *Für* $i, j \in \{1, \dots, n\}$ *sei* $A_{i,j} \in K^{(n-1)\times(n-1)}$ *die Matrix, die aus* A *durch Weglassen der* i*-ten Zeile und der* j*-ten Spalte entsteht. Für alle* $i \in \{1, \dots, n\}$ *gilt*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det(A_{i,j}), \tag{9.3}$$

und für alle $j \in \{1, \dots, n\}$ *gilt*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det(A_{i,j}). \tag{9.4}$$

Wir lassen den Beweis weg. Er ist nicht besonders schwer, aber sehr notationslastig. Man beachte, dass die Formeln (9.3) und (9.4) das rekursive Berechnen der Determinante ermöglichen. Die Berechnung gemäß Formel (9.3) wird als *Entwicklung nach der* i*-ten Zeile* bezeichnet, und gemäß (9.4) als *Entwicklung nach der* j*-ten Spalte*. Man kann eine dieser Formeln anwenden und dabei i bzw. j nach Opportunitätsgesichtspunkten auswählen.

Beispiel 9.12. Wir möchten die Determinante von

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}$$

berechnen und entscheiden uns für Entwicklung nach der ersten Zeile. Es ergibt sich

$$\det(A) = 0 \cdot \det\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} - 1 \cdot \det\begin{pmatrix} 3 & 5 \\ 6 & 8 \end{pmatrix} + 2 \cdot \det\begin{pmatrix} 3 & 4 \\ 6 & 7 \end{pmatrix}$$
$$= -(3 \cdot 8 - 6 \cdot 5) + 2 \cdot (3 \cdot 7 - 6 \cdot 4) = 6 - 6 = 0.$$

◁

Aus Satz 9.11 und Lemma 9.6(c) folgt auch die Regel für die sogenannte adjunkte Matrix: Mit $c_{i,j} := (-1)^{i+j} \det(A_{j,i})$ und $C := (c_{i,j}) \in K^{n \times n}$ (dies ist die **adjunkte Matrix**) gilt

$$A \cdot C = C \cdot A = \det(A) \cdot I_n.$$

Auch hierfür werden wir den Beweis nicht führen. Für $A \in GL_n(K)$ erhalten wir die Formel

$$A^{-1} = \frac{1}{\det(A)} \cdot C.$$

Das Berechnen der Inversen nach dieser Formel ist aufwändiger als durch das in Kapitel 7 angegebene Verfahren. Die Formel kann jedoch nützlich sein, wenn in A Parameter vorkommen, oder um die auftretenden Nenner zu kontrollieren.

Beispiel 9.13. Für invertierbare 2×2-Matrizen liest sich die obige Formel als

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Dies lässt sich auch direkt verifizieren. ◁

Wir können schon jetzt die Determinante einiger spezieller Matrizen im „Eilverfahren“ berechnen. Wir führen drei Fälle an. Begründen kann man die Ergebnisse jeweils entweder durch Entwicklung nach einer Zeile oder Spalte, oder indem man direkt mit Definition 9.4 arbeitet.

(1) Für eine *Diagonalmatrix*

$$A = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix}$$

gilt

$$\det(A) = a_1 \cdots a_n.$$

Man schreibt Diagonalmatrizen wie oben auch als

$$A = \operatorname{diag}(a_1, \ldots, a_n).$$

(2) Für eine *obere Dreiecksmatrix*

$$A = \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix} \tag{9.5}$$

gilt

$$\det(A) = a_1 \cdots a_n. \tag{9.6}$$

Zur Erklärung: (9.5) soll andeuten, dass oberhalb der Diagonalen irgendwelche Einträge stehen können, unterhalb aber lauter Nullen. Man könnte eine obere Dreiecksmatrix $A = (a_{i,j}) \in K^{n \times n}$ auch formaler durch die Bedingung $a_{i,j} = 0$ für $i > j$ definieren.

Dasselbe Ergebnis (9.6) gilt auch für untere Dreiecksmatrizen.

(3) Für eine Matrix

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

mit $B \in K^{l \times l}$, $D \in K^{(n-l)\times(n-l)}$ und $C \in K^{l \times (n-l)}$ gilt

$$\det(A) = \det(B) \cdot \det(D).$$

Man sagt auch, dass A *Block-Dreiecksgestalt* hat. Dies lässt sich erweitern auf Matrizen mit mehr als zwei Diagonal-Blöcken.

Nun wenden wir uns dem Berechnen der Determinante einer Matrix, die keine spezielle Gestalt hat, zu. Ziel ist es, auch hierfür den Gauß-Algorithmus einzusetzen. Wir müssen uns also überlegen, welche Auswirkungen elementare Zeilenoperationen auf die Determinante haben. Bei Operationen von Typ I (Vertauschen zweier Zeilen) geht die Antwort aus Lemma 9.6(b) hervor: Die Determinante ändert das Vorzeichen. Für Operationen vom Typ II und (wichtiger!) vom Typ III ist es zweckdienlich, diese als Links-Multiplikation mit gewissen Matrizen zu interpretieren: Multiplikation der i-ten Zeile von A mit einem Skalar $s \neq 0$ entspricht der Multiplikation von A mit der Matrix

$$S = \operatorname{diag}(1, \ldots, 1, s, 1, \ldots, 1),$$

wobei s der i-te Eintrag ist; also $A \to S \cdot A$. Wegen Satz 9.7 und der Regel (1) ergibt sich, dass sich bei einer Operation von Typ II die Determinante mit s multipliziert.

Um Operationen von Typ III zu behandeln, betrachten wir Matrizen $E_{i,j} \in K^{n \times n}$, die per Definition überall Nullen haben außer im (i,j)-ten Eintrag, der 1 ist. Nun sieht man leicht, dass Addition des s-fachen der j-ten Zeile zu der i-ten Zeile einer Multiplikation mit $I_n + s \cdot E_{i,j}$ entspricht: $A \to (I_n + s \cdot E_{i,j}) \cdot A$. Da $I_n + s \cdot E_{i,j}$ eine Dreiecksmatrix ist, folgt aus der Regel (2), dass $\det(I_n + s \cdot E_{i,j}) = 1$ ist, also ändert sich nach Satz 9.7 die Determinante bei Operationen von Typ III nicht. Wir fassen zusammen:

Typ I (Vertauschen zweier Zeilen): Die Determinante ändert das Vorzeichen.

Typ II (Multiplikation einer Zeile mit einem Skalar $s \in K \setminus \{0\}$): Die Determinante multipliziert sich mit s. Als Formel ausgedrückt:

$$\det(\text{neue Matrix}) = s \cdot \det(\text{alte Matrix}).$$

Typ III (Addition des s-fachen einer Zeile zu einer anderen): Die Determinante ändert sich nicht.

Wir bemerken noch, dass Entsprechendes auch für *elementare Spaltenoperationen* gilt.

Nun kann man den Gauß-Algorithmus zum Berechnen von Determinanten verwenden. Die Strategie ist, jeweils eine Spalte (oder Zeile) so weit auszuräumen, dass eine Entwicklung nach dieser Spalte (Zeile) sehr einfach wird. Man kann dabei den Gauß-Algorithmus variieren, denn es kommt nicht darauf an, welche Spalte bzw. Zeile jeweils ausgeräumt wird.

Beispiel 9.14. Wir berechnen (mit nachfolgenden Kommentaren zu den Rechenschritten)

$$\det\begin{pmatrix} 1 & 3 & 4 & 2 \\ 1 & 4 & 2 & 0 \\ 0 & 2 & 1 & 3 \\ 1 & -5 & 0 & -1 \end{pmatrix} \underset{(1)}{=} \det\begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & 1 & -2 & -2 \\ 0 & 2 & 1 & 3 \\ 0 & -8 & -4 & -3 \end{pmatrix} \underset{(2)}{=} 1 \cdot \det\begin{pmatrix} 1 & -2 & -2 \\ 2 & 1 & 3 \\ -8 & -4 & -3 \end{pmatrix}$$

$$\underset{(3)}{=} \det\begin{pmatrix} 5 & 0 & 4 \\ 2 & 1 & 3 \\ 0 & 0 & 9 \end{pmatrix} \underset{(4)}{=} 1 \cdot \det\begin{pmatrix} 5 & 4 \\ 0 & 9 \end{pmatrix} \underset{(5)}{=} 5 \cdot 9 = 45.$$

Hierbei wurden folgende Schritte durchgeführt:

(1) Ausräumen der ersten Spalte durch Addition des (-1)-fachen der ersten Zeile zur zweiten und Addition des (-1)-fachen der ersten Zeile zur vierten;
(2) Entwicklung nach der ersten Spalte;
(3) Ausräumen der zweiten Spalte durch Addition des 2-fachen der zweiten Zeile auf die erste und Addition des 4-fachen der zweiten Zeile auf die dritte (Ausräumen der ersten Spalte wäre ein etwas größerer arithmetischer Aufwand gewesen: Wer möchte schon mit 8 multiplizieren?);
(4) Entwicklung nach der zweiten Spalte;
(5) die Formel für Dreiecksmatrizen (oder die Formel für 2×2-Determinanten). $\triangleleft$

Zum Abschluss des Kapitels geben wir noch eine geometrische Interpretation der Determinante. Für $v_1, v_2 \in \mathbb{R}^2$ ist $|\det(v_1 v_2)|$ der *Flächeninhalt* des Parallelogramms mit den Seiten v_1 und v_2. Dies lässt sich auf n-dimensionale Volumina verallgemeinern. Diese Interpretation ist solange nicht beweisbar, wie wir keinen mathematisch definierten Begriff von Flächeninhalt haben. Flächeninhalte von Parallelogrammen (bzw. deren höher-dimensionalen Verallgemeinerungen) sind besonders wichtig, weil Parallelogramme bei Flächen-Integralen als „infinitessimale“ Flächenelemente auftreten.

Kapitel 10
Eigenwerte

Auch in diesem Kapitel sei K ein Körper.

Definition 10.1. *Sei $A \in K^{n \times n}$ eine quadratische Matrix. Ein $\lambda \in K$ heißt* **Eigenwert** *von A, falls es $v \in K^n \setminus \{0\}$ gibt mit $A \cdot v = \lambda \cdot v$. Ein solcher Vektor v heißt dann ein* **Eigenvektor** *von A (zum Eigenwert λ).*

$$E_\lambda := \{v \in K^n \mid A \cdot v = \lambda \cdot v\}$$

heißt der **Eigenraum** *zum Eigenwert λ. Er besteht aus allen Eigenvektoren und dem Nullvektor.*

Beispiel 10.2. Für $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$ gilt

$$A \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

also ist 1 ein Eigenwert von A und $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ein zugehöriger Eigenvektor. Ein weiterer Eigenwert ist -1, denn

$$A \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = - \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Der Eigenraum zu $\lambda = 1$ ist

$$E_1 = \left\{v \in \mathbb{R}^2 \mid A \cdot v = v\right\} = \left\{v \in \mathbb{R}^2 \mid (A - I_2) \cdot v = 0\right\},$$

also der Lösungsraum des homogenen LGS $(A - I_2) \cdot x = 0$. $A - I_2 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ hat den Rang 1, also folgt nach Proposition 6.11 $\dim(E_1) = 1$. Wir erhalten also

$$E_1 = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle,$$

und mit den gleichen Argumenten

$$E_{-1} = \langle \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle.$$

Insgesamt stellen wir fest, dass $\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \}$ eine Basis aus Eigenvektoren bildet. Die Frage, ob A außer ± 1 noch weitere Eigenwerte hat, werden wir bald beantworten können. ◁

Definition 10.1 lässt sich auf lineare Abbildungen $\varphi: V \to V$ eines K-Vektorraums V übertragen: Man fordert $\varphi(v) = \lambda v$ für ein $v \in V \setminus \{0\}$.

Beispiel 10.3. (1) Auf dem Polynomring $V = \mathbb{R}[x]$ betrachten wir den Ableitungs-Operator $\varphi: V \to V,\ f \mapsto f'$. Wir sehen, dass $\lambda = 0$ der einzige Eigenwert ist. Der Eigenraum besteht aus den konstanten Polynomen. Erweitert man den Ableitungs-Operator auf den Raum $C^\infty(\mathbb{R})$ der unendlich oft differenzierbaren Funktionen $\mathbb{R} \to \mathbb{R}$, so ändert sich die Situation: Für $\lambda \in \mathbb{R}$ ist die Exponentialfunktion $f_\lambda: \mathbb{R} \to \mathbb{R},\ x \mapsto \exp(\lambda x)$ ein Eigenvektor (man spricht in diesem Zusammenhang auch von einer *Eigenfunktion*) zum Eigenwert λ.

(2) Auf dem Vektorraum $V = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ definieren wir eine lineare Abbildung $\varphi: V \to V,\ f \mapsto \tilde{f}$, wobei $\tilde{f}$ gegeben ist durch $\tilde{f}(x) = x \cdot f(x)$. Die Suche nach Eigenwerten und -funktionen bedeutet, dass wir Funktionen f suchen, so dass $x \cdot f(x) = \lambda \cdot f(x)$ für alle $x \in \mathbb{R}$ gilt. Diese (einschneidende) Bedingung wird durch die Funktion δ_λ mit $\delta_\lambda(x) = 1$ für $x = \lambda$ und $\delta_\lambda(x) = 0$ for $x \neq \lambda$ erfüllt. Es treten also alle $\lambda \in \mathbb{R}$ als Eigenwerte auf. Die lineare Abbildung φ lässt sich als „Ortsoperator“ interpretieren. Genau die Funktionen, die am „Ort“ λ „konzentriert“ sind, sind die Eigenfunktionen zu λ.

(3) Der Eigenraum zum Eigenwert $\lambda = 0$ ist nicht anderes als $\mathrm{Kern}(\varphi)$. ◁

Auch für $\lambda \in K$, die nicht Eigenwerte sind, kann man E_λ wie in Definition 10.1 definieren. Es kommt dann der Nullraum heraus.

Im obigen Beispiel haben wir bereits gesehen, dass Eigenräume Unterräume sind. Dies gilt allgemein, denn für $A \in K^{n \times n}$ und $\lambda \in K$ ist E_λ der Lösungsraum des homogenen LGS $(A - \lambda I_n) \cdot x = 0$. Wir halten fest:

Proposition 10.4. *Für $A \in K^{n \times n}$ und $\lambda \in K$ ist $E_\lambda \subseteq K^n$ ein Unterraum.*

Wie kann man Eigenwerte berechnen? Nach Definition ist $\lambda \in K$ genau dann ein Eigenwert, wenn $E_\lambda \neq \{0\}$, d.h. wenn das homogene LGS $(A - \lambda I_n) \cdot x = 0$ nicht eindeutig lösbar ist. Dies ist nach den Ergebnissen von Kapitel 9 äquivalent zu $\det(A - \lambda I_n) = 0$ (siehe Seite 58). Diese Überlegungen nehmen wir zum Anlass für eine Definition.

Definition 10.5. *Sei $A \in K^{n \times n}$ eine quadratische Matrix. Im Polynomring $K[x]$ bilden wir*

$$\chi_A := \det(x \cdot I_n - A).$$

Das so definierte Polynom heißt das **charakteristische Polynom** *von A. Wir bemerken, dass χ_A ein Polynom von Grad n mit höchstem Koeffizient 1 ist.*

Den folgenden Satz haben wir bereits gezeigt.

Satz 10.6. *Die Eigenwerte einer quadratischen Matrix A sind die Nullstellen des charakteristischen Polynoms χ_A.*

Beispiel 10.7. (1) Für $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ gilt

$$\chi_A = \det \begin{pmatrix} x & -1 \\ -1 & x \end{pmatrix} = x^2 - 1,$$

also sind 1 und -1 die (einzigen) Eigenwerte.

(2) Für $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ gilt

$$\chi_A = \det \begin{pmatrix} x & -1 \\ 1 & x \end{pmatrix} = x^2 + 1,$$

also hat A keine Eigenwerte (in $\mathbb{R}$). ◁

Wir erinnern kurz an das Rechnen mit Polynomen und deren Nullstellen. Wir wissen, dass $K[x]$ mit der gewöhnlichen Addition und Multiplikation von Polynomen ein Ring ist. Außerdem haben wir eine *Division mit Rest*: Für $f, g \in K[x]$ mit $g \neq 0$ gibt es Polynome $q, r \in K[x]$ mit

$$f = g \cdot q + r \quad \text{und} \quad \deg(r) < \deg(g).$$

Falls nun $\lambda \in K$ eine Nullstelle eines Polynoms $f \neq 0$ ist, so können wir Division mit Rest durch $g = x - \lambda$ durchführen und erhalten

$$f = (x - \lambda) \cdot q + r$$

mit $\deg(r) < 1$, also ist r ein konstantes Polynom. Einsetzen von $x = \lambda$ liefert $r = r(\lambda) = 0$. Es folgt $f = (x - \lambda) \cdot q$ mit $\deg(q) = \deg(f) - 1$. Man sagt, dass man den *Linearfaktor* $x - \lambda$ von f abspalten kann. Ist $\mu \in K$ nun eine weitere Nullstelle von f mit $\mu \neq \lambda$, so folgt

$$(\mu - \lambda) q(\mu) = f(\mu) = 0,$$

also $q(\mu) = 0$. Ein Induktionsargument nach $n := \deg(f)$ zeigt nun, dass f höchstens n verschiedene Nullstellen hat. Falls $\lambda_1, \ldots, \lambda_r$ die paarweise verschiedenen Nullstellen von f sind, so folgt

$$f = (x - \lambda_1)^{e_1} \cdots (x - \lambda_r)^{e_r} \cdot g,$$

wobei $g \in K[x]$ ein Polynom ohne Nullstellen ist, und $e_i \in \mathbb{N}_{>0}$. Die Zahl e_i heißt die *Vielfachheit* der Nullstelle λ_i. Falls g ein konstantes Polynom ist, so sagen wir, dass f *in Linearfaktoren zerfällt*. Dann ist g automatisch der höchste Koeffizient von f.

Beispiel 10.8. Für das Polynom $f = x^4 + 2x^3 + 2x^2 + 2x + 1 \in \mathbb{R}[x]$ finden wir durch Ausprobieren die Nullstelle $\lambda_1 = -1$. Division mit Rest liefert

$$f = (x + 1) \cdot (x^3 + x^2 + x + 1).$$

Auch das verbleibend Polynom x^3+x^2+x+1 hat die Nullstelle -1. Erneute Division mit Rest liefert

$$f = (x + 1)^2 \cdot (x^2 + 1).$$

Da $x^2 + 1$ keine Nullstelle in $\mathbb{R}$ hat, ist -1 die einzige Nullstelle von f, und die Vielfachheit ist 2. Wenn wir f als ein Polynom über $\mathbb{C}$ auffassen, zerfällt es in Linearfaktoren:

$$f = (x + 1)^2 \cdot (x - i) \cdot (x + i).$$

$\lhd$

Über $\mathbb{R}$ zerfallen also nicht alle Polynome. Wie das obige Beispiel suggeriert, haben wir über $\mathbb{C}$ bessere Chancen. (Zur Erinnerung: $\mathbb{C}$ lässt sich definieren als

$$\mathbb{C} := \{a + b \cdot i \mid a, b \in \mathbb{R}\}$$

mit $i^2 = -1$.) In der Tat gilt der folgende

Satz 10.9 (Fundamentalsatz der Algebra)**.** *Jedes nicht-konstante Polynom $f \in \mathbb{C}[x]$ hat eine Nullstelle in $\mathbb{C}$. Damit zerfällt f in Linearfaktoren.*

Körper, die wie $\mathbb{C}$ die Eigenschaft aus Satz 10.9 haben, bezeichnet man als **algebraisch abgeschlossen**. Wir werden Satz 10.9 nicht beweisen, da dies mit unseren derzeitigen Mitteln nicht möglich ist. Der Beweis benötigt Methoden aus der Funktionentheorie (= komplexe Analysis) oder der Analysis. Eine Beweisvariante benutzt Methoden aus der Algebra und (ein wenig) Analysis. (Ganz ohne Analysis kann man nicht auskommen, da schon die Definition von $\mathbb{R}$ und damit von $\mathbb{C}$ Begriffe aus der Analysis benötigt.)

Korollar 10.10. *Sei $A \in K^{n\times n}$.*

(a) A hat höchstens n Eigenwerte.
(b) Falls K algebraisch abgeschlossen ist (z.B. $K = \mathbb{C}$), so hat A Eigenwerte.

Im Lichte der bisherigen Überlegungen erscheinen die folgenden zwei Definitionen für die Vielfachheit eines Eigenwertes als natürlich.

Definition 10.11. *Es sei $\lambda \in K$ ein Eigenwert einer Matrix $A \in K^{n\times n}$.*

(a) *Die* **algebraische Vielfachheit** $m_a(\lambda)$ *von* λ *ist die Vielfachheit der Nullstelle* λ *im charakteristischen Polynom* χ_A.

(b) *Die* **geometrische Vielfachheit** *von* λ *ist*

$$m_g(\lambda) := \dim(E_\lambda).$$

Beispiel 10.12. (1) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ hat die Eigenwerte 1 und -1 (siehe Beispiel 10.2). Für beide Eigenwerte sind algebraische- und geometrische Vielfachheit gleich 1.

(2) Für $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ gilt

$$\chi_A = \det \begin{pmatrix} x-1 & -1 \\ 0 & x-1 \end{pmatrix} = (x-1)^2$$

(obere Dreiecksmatrix), also ist $\lambda = 1$ der einzige Eigenwert mit algebraische Vielfachheit $m_a(\lambda) = 2$. Zur Ermittlung der geometrischen Vielfachheit bemerken wir, dass

$$A - I_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

den Rang 1 hat, also $m_g(\lambda) = 1$. ◁

Satz 10.13. *Ist* $\lambda \in K$ *ein Eigenwert einer Matrix* $A \in K^{n\times n}$, *so gilt*

$$1 \le m_g(\lambda) \le m_a(\lambda).$$

Beweis. Die erste Ungleichung ist klar, denn für einen Eigenwert gilt $E_\lambda \neq \{0\}$, also $\dim(E_\lambda) \ge 1$.

Zur Beweis der zweiten Ungleichung setzen wir $m := m_g(\lambda)$ und wählen eine Basis $\{v_1, \ldots, v_m\}$ von E_λ. Diese können wir zu einer Basis $B = \{v_1, \ldots, v_n\}$ von K^n ergänzen. Für $1 \le i \le m$ gilt

$$\varphi_A(v_i) = A \cdot v_i = \lambda \cdot v_i,$$

also hat die Darstellungsmatrix von φ_A bzgl. B die Form

$$D_B(\varphi_A) = \left(\begin{array}{ccc|c} \lambda & & 0 & \\ & \ddots & & * \\ 0 & & \lambda & \\ \hline & 0 & & C \end{array}\right) =: D$$

mit $C \in K^{(n-m)\times(n-m)}$. Mit $S := (v_1 \ldots v_n) \in \mathrm{GL}_n(K)$ (die Matrix mit den v_i als Spalten) gilt $S^{-1}AS = D$ (wegen Satz 8.6), also

$$A = SDS^{-1}.$$

Es folgt

$$\chi_A = \det(xI_n - A) = \det\left(xI_n - SDS^{-1}\right) = \det\left(S(xI_n - D)S^{-1}\right) = \det(xI_n - D),$$

wobei die letzte Gleichheit aus Korollar 9.9 folgt. Die Matrix $xI_n - D$ ist jedoch (ebenso wie D selbst) eine obere Dreiecksmatrix. Damit können wir die Determinante ablesen und erhalten

$$\chi_A = (x-\lambda)^m \cdot \chi_C.$$

Also wird χ_A durch $(x-\lambda)^m$ geteilt, und wir schließen $m_a(\lambda) \geq m$, wie behauptet. □

Definition 10.14. *Eine quadratische Matrix $A \in K^{n\times n}$ heißt* **diagonalisierbar**, *falls es eine Basis von K^n bestehend aus Eigenvektoren von A gibt. Gleichbedeutend: A ist ähnlich zu einer Diagonalmatrix.*

Beispiel 10.15. (1) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist diagonalisierbar (siehe Beispiel 10.2).

(2) $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist nicht diagonalisierbar. Es fehlen Eigenwerte (siehe Beispiel 10.7(2)).

(3) $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist nicht diagonalisierbar. Es fehlen Eigenvektoren (siehe Beispiel 10.12(2)). ◁

Wir werden folgendes Kriterium für Diagonalisierbarkeit beweisen. Es besagt, dass die in Beispiel 10.15(2) und (3) aufgetretenen Hindernisse für die Diagonalisierbarkeit tatsächlich die einzig möglichen Hindernisse sind.

Satz 10.16. *Eine Matrix $A \in K^{n\times n}$ ist genau dann diagonalisierbar, wenn beide der folgenden Bedingungen erfüllt sind:*

(a) Das charakteristische Polynom χ_A zerfällt in Linearfaktoren, also

$$\chi_A = \prod_{i=1}^{r} (x - \lambda_i)^{e_i}$$

mit $e_i = m_a(\lambda_i)$.

(b) Für alle Eigenwerte λ_i gilt

$$m_g(\lambda_i) = m_a(\lambda_i).$$

Das folgende Lemma benötigen wir für den Beweis.

Lemma 10.17. *Es seien $\lambda_1, \dots, \lambda_r \in K$ paarweise verschiedene Eigenwerte einer Matrix $A \in K^{n \times n}$. Weiter seien $v_i \in E_{\lambda_i}$ mit*

$$v_1 + \dots + v_r = 0.$$

Dann sind alle $v_i = 0$.

Beweis. Wir benutzen Induktion nach r. Für $r = 1$ ist nichts zu zeigen. Wir können also ab jetzt $r \geq 2$ voraussetzen. Wir rechnen:

$$\sum_{i=1}^{r} \lambda_i v_i = \sum_{i=1}^{r} A \cdot v_i = A \cdot \left(\sum_{i=1}^{r} v_i \right) = A \cdot 0 = 0.$$

Andererseits gilt

$$\sum_{i=1}^{r} \lambda_1 v_i = \lambda_1 \cdot \left(\sum_{i=1}^{r} v_i \right) = 0.$$

Wir subtrahieren beide Gleichungen und erhalten

$$\sum_{i=2}^{r} (\lambda_i - \lambda_1) v_i = 0.$$

Da $(\lambda_i - \lambda_1) v_i$ in E_{λ_i} liegt, liefert die Induktionsvoraussetzung $(\lambda_i - \lambda_1) v_i = 0$ für $i \in \{2, \dots, r\}$. Wegen $\lambda_i \neq \lambda_1$ folgt $v_i = 0$ für $i \in \{2, \dots, r\}$. Nun folgt auch $v_1 = -(v_2 + \dots + v_r) = 0$. □

Beweis von Satz 10.16. Zunächst nehmen wir an, dass A diagonalisierbar ist, es gibt also eine Basis B von K^n aus Eigenvektoren. Sind $\lambda_1, \dots, \lambda_r$ die Eigenwerte von A, so folgt mit $B_i := B \cap E_{\lambda_i}$:

$$n = |B| = \sum_{i=1}^{r} |B_i| \leq \sum_{i=1}^{r} m_g(\lambda_i) \leq \sum_{i=1}^{r} m_a(\lambda_i) \leq \deg(\chi_A) = n,$$

wobei die mittlere Ungleichung aus Satz 10.13 folgt und die letzte aus der Definition der $m_a(\lambda_i)$ als Vielfachheiten der Nullstellen von χ_A folgt. Es muss also überall Gleichheit gelten, und es folgen (a) und (b).

Nun nehmen wir umgekehrt an, dass (a) und (b) gelten. Für $i \in \{1, \dots, r\}$ sei B_i eine Basis des Eigenraums E_{λ_i}. Wir setzen $B := B_1 \cup \dots \cup B_r$. Es ist klar, dass B aus Eigenvektoren besteht. Aus Lemma 10.17 folgt, dass B linear unabhängig ist. Außerdem gilt

$$|B| = \sum_{i=1}^{r} |B_i| = \sum_{i=1}^{r} m_g(\lambda_i) \underset{(b)}{=} \sum_{i=1}^{r} m_a(\lambda_i) \underset{(a)}{=} \deg(\chi_A) = n.$$

Insgesamt folgt mit Korollar 6.13(a), dass B eine Basis von K^n ist. □

Korollar 10.18. *Es sei $A \in K^{n\times n}$. Falls χ_A in Linearfaktoren zerfällt und nur Nullstellen der Vielfachheit 1 hat, so ist A diagonalisierbar.*

Als Anwendung betrachten wir ein physikalisches Beispiel. Wir stellen uns vor, dass zwei gleichschwere Massepunkte mit identischen, masselosen Federn an gegenüberliegenden Wänden verbunden sind, und dass zwischen den Massepunkten eine weitere, andersartige Feder befestigt ist. Man spricht auch von *gekoppelten Schwingern*. Wenn $x_1(t)$ und $x_2(t)$ die Auslenkungen der Massepunkte (gemessen ab er Ruhelage) zur Zeit t bezeichnen, so gelten die Differentialgleichungen

$$\begin{aligned} \ddot{x}_1(t) &= -ax_1(t) - b\left(x_1(t) - x_2(t)\right), \\ \ddot{x}_2(t) &= -ax_2(t) - b\left(x_2(t) - x_1(t)\right), \end{aligned}$$

wobei die Doppelpunkte wie üblich für die zweite Ableitung nach t stehen und die positiven Konstanten a und b von den Federeigenschaften und dem Gewicht der Massepunkte anhängen. In Matrixschreibweise:

$$\begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = \underbrace{\begin{pmatrix} -a-b & b \\ b & -a-b \end{pmatrix}}_{=:A} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Das charakeristische Polynom von A ist

$$\chi_A = \det \begin{pmatrix} x+a+b & -b \\ -b & x+a+b \end{pmatrix} = (x+a+b)^2 - b^2 = (x+a)(x+a+2b).$$

Korollar 10.18 garantiert, dass A diagonalisierbar ist. Die Eigenräume berechnen wir durch Auflösen von homogenen LGS (oder hinschauen):

$$E_{-a} = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle \quad \text{und} \quad E_{-a-2b} = \langle \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle.$$

Mit $S := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ folgt

$$S^{-1}AS = \begin{pmatrix} -a & 0 \\ 0 & -a-2b \end{pmatrix}.$$

Wir setzen $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} := S^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ und erhalten die Differentialgleichung

$$\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} = S^{-1} \begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{pmatrix} = S^{-1}A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S^{-1}AS \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -a & 0 \\ 0 & -a-2b \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Die Diagonalisierung der Matrix hat also dazu geführt, dass wir zwei getrennte Differentialgleichungen für y_1 und y_2 bekommen haben. Diese können wir leicht lösen. Mit $\omega := \sqrt{a}$ und $\tilde{\omega} := \sqrt{a+2b}$ lautet die allgemeine Lösung

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} c_1 \cos(\omega t) + c_2 \sin(\omega t) \\ c_3 \cos(\tilde{\omega} t) + c_4 \sin(\tilde{\omega} t) \end{pmatrix}$$

mit Konstanten c_i. Durch Multiplikation mit S erhalten wir

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = c_1 \begin{pmatrix} \cos(\omega t) \\ \cos(\omega t) \end{pmatrix} + c_2 \begin{pmatrix} \sin(\omega t) \\ \sin(\omega t) \end{pmatrix} + c_3 \begin{pmatrix} \cos(\tilde{\omega} t) \\ -\cos(\tilde{\omega} t) \end{pmatrix} + c_4 \begin{pmatrix} \sin(\tilde{\omega} t) \\ -\sin(\tilde{\omega} t) \end{pmatrix}.$$

Interessant ist die Lösung mit $c_1 = c_3 = 0$ und $c_2 = c_4 = 1$, die (nach ein paar Umformungen)

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = 2 \begin{pmatrix} \cos\left(\frac{\tilde{\omega}-\omega}{2} \cdot t\right) \cdot \sin\left(\frac{\tilde{\omega}+\omega}{2} \cdot t\right) \\ -\sin\left(\frac{\tilde{\omega}-\omega}{2} \cdot t\right) \cdot \cos\left(\frac{\tilde{\omega}+\omega}{2} \cdot t\right) \end{pmatrix}$$

lautet. Diese beschreibt ein periodisches Übertragen der Schwingung von der einen Masse zur anderen und zurück.

Durch Satz 10.16 wird klar: Auch über einem algebraisch abgeschlossenen Körper sind nicht alle quadratischen Matrizen diagonalisierbar. Eine über keinem Körper diagonalisierbare Matrix findet sich in Beispiel 10.15(3). Es gilt aber die folgende Abschwächung:

Satz 10.19. *Es sei K algebraisch abgeschlossen (z.B. $K = \mathbb{C}$) und $A \in K^{n \times n}$. Dann ist A ähnlich zu einer oberen Dreiecksmatrix:*

$$S^{-1}AS = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

mit $S \in \mathrm{GL}_n(K)$. Es gilt $\chi_A = \prod_{i=1}^{n}(x - \lambda_i)$.

Anmerkung. Eine zu einer Dreiecksmatrix ähnliche Matrix nennt man auch **trigonalisierbar**. ◁

Beweis von Satz 10.19. Wir benutzen Induktion nach n. Für $n = 1$ ist nichts zu zeigen, also sei $n > 1$. Nach Voraussetzung hat χ_A eine Nullstelle $\lambda_1 \in K$, also ist λ_1 ein Eigenwert. Wir nehmen einen Eigenvektor v_1 und ergänzen zu einer Basis $\{v_1, \ldots, v_n\}$ von K^n. Wir bilden die Matrix $S = (v_1, \ldots, v_n) \in \mathrm{GL}_n(K)$ mit den v_i als Spalten. Da v_1 ein Eigenvektor zum Eigenwert λ_1 ist, folgt

$$S^{-1}AS = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right) \tag{10.1}$$

mit $B \in K^{(n-1)\times(n-1)}$. (Wie immer gilt die Konvention, dass Sterne für unbekannte Einträge stehen.) Nach Induktion gibt es $T \in \mathrm{GL}_{n-1}(K)$, so dass

$$T^{-1}BT = \begin{pmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

gilt. Nun folgt

$$\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T & \\ 0 & & & \end{array}\right)^{-1} S^{-1}AS \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T & \\ 0 & & & \end{array}\right) =$$

$$\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T^{-1} & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & T & \\ 0 & & & \end{array}\right) = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & \lambda_2 & & * \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_n \end{array}\right).$$

Damit ist gezeigt, dass A ähnlich ist zu einer oberen Dreiecksmatrix. Die Aussage über das charakteristische Polynom folgt daraus, dass ähnlich Matrizen identische charakteristische Polynome haben. □

Anmerkung 10.20. (a) In Wirklichkeit gilt Satz 10.19 unter der allgemeineren Voraussetzung, dass das charakteristische Polynom χ_A in Linearfaktoren zerfällt. Unser Beweis funktioniert auch unter dieser Voraussetzung, denn aus (10.1) folgt $\chi_A = (x - \lambda_1) \cdot \chi_B$, also zerfällt auch χ_B in Linearfaktoren. Außerdem zeigt der Beweis, dass man die Reihenfolge der λ_i nach Belieben wählen kann.

(b) Jede quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, ist auch ähnlich zu einer *unteren* Dreiecksmatrix. Der Beweis läuft analog, oder man kann die Aussage über untere Dreiecksmatrizen auf die über obere Dreiecksmatrizen zurückführen. ◁

Kapitel 11
Die Jordansche Normalform

In diesem Kapitel behandeln wir (ohne Beweise) die Jordansche Normalform, die eine wesentliche Verfeinerung der Trigonalisierung (Satz 10.19) darstellt.

Definition 11.1. *(a) Ein* **Jordan-Block** *(auch genannt:* **Jordan-Kästchen***) der Länge e ist eine Matrix von der Form*

$$J = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & & 1 \\ 0 & & & \lambda \end{pmatrix} \in K^{e \times e},$$

wobei $\lambda \in K$ und $e \in \mathbb{N}_{>0}$. Wir haben also lauter (identische) Einträge λ auf der Diagonalen und lauter Einsen auf der oberen Nebendiagonalen. Ein Jordan-Block der Länge 1 ist eine 1×1-Matrix mit Eintrag λ, also einfach ein Skalar ohne Nebendiagonale.

(b) Eine quadratische Matrix $A \in K^{n \times n}$ heißt in **Jordanscher Normalform***, falls*

$$A = \begin{pmatrix} \boxed{J_1} & & & 0 \\ & \boxed{J_2} & & \\ & & \ddots & \\ 0 & & & \boxed{J_r} \end{pmatrix}$$

eine Block-Diagonalmatrix ist mit Jordan-Blöcken J_i als Diagonalblöcke. Dabei müssen die J_i nicht verschieden sein; es kann auch mehrere Blöcke verschiedener Länge mit dem gleichen Diagonaleintrag λ geben. Haben al-

le Jordan-Blöcke die Länge 1*, so ist* A *eine Diagonalmatrix. Die Jordansche Normalform verallgemeinert also die Diagonalform.*

(c) *Sei* $A \in K^{n \times n}$ *eine quadratische Matrix. Eine Matrix* $B \in K^{n \times n}$ *heißt eine* **Jordansche Normalform** *von* A*, falls* B *in Jordanscher Normalform und ähnlich zu* A *ist.*

Das charakteristische Polynom eines Jordan-Blocks J der Länge e mit Diagonaleinträgen λ ist $(x-\lambda)^e$. Außerdem ist das charakteristische Polynom einer Matrix in Jordanscher Normalform das Produkt der charakteristischen Polynome der Jordan-Blöcke. Also sind die Diagonaleinträge zwangsläufig die Eigenwerte, und jeder Eigenwert λ tritt so oft in der Diagonalen auf, wie seine algebraische Vielfachheit angibt. Der Rang von $J - \lambda I_e$ ist $e-1$. Hieraus sieht man, dass die geometrische Vielfachheit eines Eigenwertes λ gleich der Anzahl der Jordan-Blöcke mit Diagonaleinträgen λ ist.

Beispiel 11.2. (1) Die Matrizen

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

sind in Jordanscher Normalform, die Matrix

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

aber nicht.

(2) Die Matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

hat

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

als Jordansche Normalform, denn mit $S = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ gilt

$$B = S^{-1} A S.$$

◁

Hat eine Matrix eine Jordansche Normalform? Der folgende Satz gibt die Antwort. Er liefert die angündigte Verfeinerung von Satz 10.19.

Satz 11.3. *Eine quadratische Matrix* $A \in K^{n \times n}$ *hat genau dann eine Jordansche Normalform, wenn das charakteristische Polynom* χ_A *in Linearfaktoren zerfällt. Falls* K *algebraisch abgeschlossen ist (z.B.* $K = \mathbb{C}$*), so hat also jede quadratische Matrix eine Jordansche Normalform.*

Wir lassen den (aufwendigen) Beweis weg. Der Satz lässt sich auch so ausdrücken: Zerfällt χ_A in Linearfaktoren, so gibt es eine Basis von K^n, bezüglich der die Darstellungsmatrix von φ_A in Jordanscher Normalform ist.

Es gibt eine Verallgemeinerung der Jordanschen Normalform, die sogenannte *allgemeine Normalform*. Diese existiert auch dann, wenn das charakteristische Polynom nicht in Linerfaktoren zerfällt. Wir werden sie in dieser Vorlesung nicht behandeln.

Nachdem die Existenz einer Jordanschen Normalform geklärt ist, ist die nächste Frage die Eindeutigkeit: Hat eine Matrix $A \in K^{n\times n}$ nur eine Jordansche Normalform? Sicherlich ist die Reihenfolge der Jordan-Blöcke in einer Jordanschen Normalform austauschbar: Dies entspricht einer Umordnung der Basis von K^n, bezüglich der die Darstellungsmatrix von φ_A in Jordanscher Normalform hat. Die bestmögliche Aussage wäre also die Eindeutigkeit der Jordanschen Normalform bis auf Reihenfolge der Jordan-Blöcke, und diese ist tatsächlichh wahr. Genauer gilt:

Satz 11.4. *Es seien $A, B \in K^{n\times n}$ Matrizen mit Jordanschen Normalformen J_A bzw. J_B. Dann sind A und B ähnlich genau dann, wenn J_A und J_B bis auf die Reihenfolge der Jordan-Blöcke übereinstimmen.*

Der Satz besagt, dass die Jordanschen Normalformen (bis auf Reihenfolge der Blöcke) die Ähnlichkeitsklassen von Matrizen „parametrisieren". Dies macht die Jordansche Normalform in theoretischer Hinsicht so interessant. Wir werden später den Beweis des Satzes skizzieren.

Wir wollen uns nun mit praktischen Aspekten beschäftigen. Die erste frage ist, wie man zu einer gegebenen Matrix eine Jordansche Normalform berechnen kann. Zunächst folgt aus der Ähnlichkeit einer Matrix A mit ihrer Jordanschen Normalform B, dass beide Matrizen dieselben Eigenwerte haben. Die Diagonaleinträge der Jordan-Blöcke sind also Eigenwerte von A. Es muss nur noch ermittelt werden, wieviele Jordan-Blöcke welcher Länge zu den Eigenwerten gehören. Dies leistet der folgende Satz:

Satz 11.5. *Es seien $A \in K^{n\times n}$ eine Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, $B \in K^{n\times n}$ eine Jordansche Normalform von A und $\lambda \in K$ ein Eigenwert von A.*

(a) Für $e \in \mathbb{N}_{>0}$ ist die Anzahl der Jordan-Blöcke in B der Länge e mit Diagonaleinträgen λ

$$\operatorname{rg}(A - \lambda I_n)^{e-1} - 2\operatorname{rg}(A - \lambda I_n)^e + \operatorname{rg}(A - \lambda I_n)^{e+1}.$$

(b) Die Gesamtlänge aller Jordan-Blöcke mit Diagonaleinträgen λ ist die algebraische Vielfachheit $m_a(\lambda)$.

(c) Die Anzahl der Jordan-Blöcke mit Diagonaleinträgen λ ist die geometrische Vielfachheit $m_g(\lambda)$.

Die (grobe) Idee für den Beweis ist die folgende: Weil die Potenzen $(A - \lambda I_n)^e$ ähnlich sind zu $(B - \lambda I_n)^e$, kann man die Formel in (a) für B anstelle von A nachprüfen. Weil die Ränge sich für Block-Diagonalmatrizen additiv verhalten, genügt das Überprüfen für einzelne Jordan-Blöcke. Dies läuft auf eine explizite Rechnung heraus. Die Behauptung in (c) ergibt sich durch aufsummieren der Vielfachheiten in (a). Der Teil (b) folgt aus der Gleichheit $\chi_A = \chi_B$ und der Berechnung des charakteristischen Polynoms einer Matrix in Jordanscher Normalform (siehe nach Definition 11.1).

Aus Satz 11.5 folgt auch der Eindeutigkeitssatz (Satz 11.4). Denn sind A und B ähnlich, so stimmen die Ränge in Satz 11.5(a) für A und B überein, also auch die Jordan-Blöcke in den Jordanschen Normalformen. Sind umgekehrt die Jordanschen Normalformen bis auf die Reihenfolge identisch, so sind sie ähnlich, also sind auch A und B ähnlich.

Wir fassen die Methode zur Berechnung der Jordanschen Normalform, die sich aus Satz 11.5 ergibt, zusammen.

Der erste Schritt ist die Berechnung des charakteristischen Polynoms χ_A und das Auffinden der Nullstellen. Wir setzen voraus, dass χ_A in Linearfaktoren zerfällt. Damit sind die Eigenwerte und deren algebraische Vielfachheiten bekannt. Hat ein Eigenwert λ die algebraische Vielfachheit 1, so gibt es zu λ genau einen Jordan-Block der Länge 1, also einen Diagonaleintrag λ in der Jordanschen Normalform ohne Einsen in der Nebendiagonalen. Bei algebraischer Vielfachheit > 1 berechnet man die geometrische Vielfachheit, also $n - \operatorname{rg}(A - \lambda I_n)$. Damit kennt man die Anzahl der Jordan-Blöcke zum Eigenwert λ, womit man zusammen mit der Kenntnis der Gesamtlänge (= algebraische Vielfachheit) häufig schon deren Längen bestimmen kann. Falls das nicht geht, muss man die Ränge der Matrizen $(A - \lambda I_n)^k$ berechnen und daraus die Häufigkeiten der Jordan-Blöcke zu λ der verschiedenen Längen gemäß Satz 11.5(a). Das macht man solange, bis man aufgrund der Kenntnis der Gesamtlänge die Längen aller Jordan-Blöcke zum Eigenwert λ bestimmt hat. Auf diese Art arbeitet man alle Eigenwerte λ ab.

Beispiel 11.6. (1) Wir betrachten die Matrix

$$A = \begin{pmatrix} -3 & -1 & 2 \\ 4 & 1 & -4 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Das charakteristische Polynom ist

$$\begin{aligned}\chi_A &= \det \begin{pmatrix} x+3 & 1 & -2 \\ -4 & x-1 & 4 \\ 0 & 0 & x+1 \end{pmatrix} = (x+1) \cdot \det \begin{pmatrix} x+3 & 1 \\ -4 & x-1 \end{pmatrix} \\ &= (x+1) \cdot (x^2 + 2x + 1) = (x+1)^3,\end{aligned}$$

wobei wir im ersten Schritt nach der dritten Zeile entwickelt haben. Der einzige Eigenwert ist also $\lambda = -1$ mit algebraischer Vielfachheit 3. Der Rang von

$$A + I_3 = \begin{pmatrix} -2 & -1 & 2 \\ 4 & 2 & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

ist 1, also gibt es zwei Jordan-Blöcke. Da die Gesamtlänge 3 ist, müssen sie die Länge 1 und 2 haben, die Jordansche Normalform ist also

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

(2) Wir betrachten die Matrix

$$A = \begin{pmatrix} -3 & -1 & 4 & -3 & -1 \\ 1 & 1 & -1 & 1 & 0 \\ -1 & 0 & 2 & 0 & 0 \\ 4 & 1 & -4 & 5 & 1 \\ -2 & 0 & 2 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{5\times 5}.$$

Das Berechnen des charakteristischen Polynoms ist aufwendig:

$$\begin{aligned}
\chi_A &= \det \begin{pmatrix} x+3 & 1 & -4 & 3 & 1 \\ -1 & x-1 & 1 & -1 & 0 \\ 1 & 0 & x-2 & 0 & 0 \\ -4 & -1 & 4 & x-5 & -1 \\ 2 & 0 & -2 & 2 & x-1 \end{pmatrix} \\
&\underset{(1)}{=} \det \begin{pmatrix} x+3 & 1 & -x^2-x+2 & 3 & 1 \\ -1 & x-1 & x-1 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ -4 & -1 & 4x-4 & x-5 & -1 \\ 2 & 0 & -2x+2 & 2 & x-1 \end{pmatrix} \\
&\underset{(2)}{=} \det \begin{pmatrix} 1 & -x^2-x+2 & 3 & 1 \\ x-1 & x-1 & -1 & 0 \\ -1 & 4x-4 & x-5 & -1 \\ 0 & -2x+2 & 2 & x-1 \end{pmatrix} \\
&\underset{(3)}{=} \det \begin{pmatrix} 0 & -x^2+3x-2 & x-2 & 0 \\ x-1 & x-1 & -1 & 0 \\ -1 & 4x-4 & x-5 & -1 \\ -x+1 & 4x^2-10x+6 & x^2-6x+7 & 0 \end{pmatrix} \\
&\underset{(4)}{=} \det \begin{pmatrix} 0 & -x^2+3x-2 & x-2 \\ x-1 & x-1 & -1 \\ -x+1 & 4x^2-10x+6 & x^2-6x+7 \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&\underset{(5)}{=} \det\begin{pmatrix} 0 & -x^2+3x-2 & x-2 \\ x-1 & x-1 & -1 \\ 0 & 4x^2-9x+5 & x^2-6x+6 \end{pmatrix} \\
&\underset{(6)}{=} -(x-1)\cdot\det\begin{pmatrix} -x^2+3x-2 & x-2 \\ 4x^2-9x+5 & x^2-6x+6 \end{pmatrix} \\
&\underset{(7)}{=} -(x-1)\cdot\det\begin{pmatrix} 0 & x-2 \\ x^3-3x^2+3x-1 & x^2-6x+6 \end{pmatrix} \\
&= (x-1)(x-2)(x^3-3x^2+3x-1) = (x-2)(x-1)^4.
\end{aligned}
$$

Die Schritte waren: (1) Addieren des $(-x+2)$-fachen der ersten Spalte zur dritten, (2) Entwickeln nach der dritten Zeile, (3) Addition der dritten Zeile zur ersten und des $(x-1)$-fachen der dritten Zeile zur letzten, (4) Entwickeln nach der letzten Spalte, (5) Addieren der zweiten Zeile zur dritten, (6) Entwickeln nach der ersten Spalte und (7) Addieren des $(x-1)$-fachen der zweiten Spalte zur ersten.

Der Eigenwert 2 ergibt einen Jordan-Block der Länge 1. Der Eigenwert 1 hat algebraische Vielfachheit 4. Wir berechnen den Rang von $A - I_5$:

$$
\begin{aligned}
\operatorname{rg}(A-I_5) &= \operatorname{rg}\begin{pmatrix} -4 & -1 & 4 & -3 & -1 \\ 1 & 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 4 & 1 & -4 & 4 & 1 \\ -2 & 0 & 2 & -2 & 0 \end{pmatrix} = \operatorname{rg}\begin{pmatrix} 0 & -1 & 0 & 1 & -1 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\
&= \operatorname{rg}\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = 3.
\end{aligned}
$$

Es gibt also $5 - 3 = 2$ Jordan-Blöcke zum Eigenwert 1. Dafür gibt es zwei Möglichkeiten (zwei Blöcke der Länge 2 oder je eines der Länge 1 und 3). Um die Anzahl der Jordan-Blöcke der Länge 1 nach Satz 11.5(a) zu berechnen, brauchen wir den Rang von $(A - I_5)^2$:

$$
\operatorname{rg}\left((A-I_5)^2\right) = \operatorname{rg}\begin{pmatrix} 1 & 1 & -1 & 1 & 1 \\ 1 & 0 & -1 & 1 & 0 \\ 3 & 1 & -3 & 3 & 1 \\ 3 & 0 & -3 & 3 & 0 \\ -2 & 0 & 2 & -2 & 0 \end{pmatrix} = \operatorname{rg}\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = 2.
$$

Die Anzahl der Jordan-Blöcke der Länge 1 ist also $5 - 2\cdot 3 + 2 = 1$. Damit hat A die Jordansche Normalform

$$\begin{pmatrix} 2&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&1&0 \\ 0&0&0&1&1 \\ 0&0&0&0&1 \end{pmatrix}.$$

◁

Oft ist es von Interesse, nicht nur die allgemeine bzw. Jordansche Normalform B einer Matrix $A \in K^{n\times n}$ zu bestimmen, sondern auch eine transformierende Matrix $S \in \mathrm{GL}_n(K)$ mit $B = S^{-1}AS$. Dies ist gleichbedeutend mit der Bestimmung einer Basis von K^n, bezüglich der φ_A die Darstellungsmatrix B hat. Bisweilen wird eine solche Basis eine *Jordan-Basis* genannt.

Im folgenden skizzieren wir die gängige Methode zur Berechnung einer Jordan-Basis. Es wird vorausgesetzt, dass die Jordansche Normalform einer Matrix $A \in K^{n\times n}$ bekannt ist. Eine Jordan-Basis setzt man zusammen aus Vektoren, die durch Anwendung von A gemäß den einzelnen Jordan-Kästchen transformiert werden. Man behandelt die Eigenwerte λ nacheinander. Zu einem Eigenwert λ sucht man zunächst Basisvektoren, die zu den *längsten* Jordan-Kästchen zum Eigenwert λ gehören. Ist deren Länge e, so berechnet man den sogenannten *Hauptraum*

$$E_\lambda^{(e)} := \{v \in K^n \mid (A - \lambda I_n)^e \cdot v = 0\}.$$

Haupträume stellen eine Verallgemeinerung der Eigenräume dar. Man ergänzt nun eine Basis des Unterraums $E_\lambda^{(e-1)}$ zu einer Basis von $E_\lambda^{(e)}$. Die ergänzenden Vektoren bilden die „Keime“ der zu den Jordan-Kästchen gehörenden Basisvektoren. Ist $v \in E_\lambda^{(e)}$ ein solcher, so setzen wir nämlich

$$v_e := v, \quad v_{e-1} := Av_e - \lambda v_e, \quad \ldots, \quad v_1 := Av_2 - \lambda v_2. \tag{11.1}$$

Für $i > 1$ folgt $A \cdot v_i = \lambda \cdot v_i + v_{i-1}$, also genau das Verhalten, das durch ein Jordan-Kästchen beschrieben wird. Aus $v \in E_\lambda^{(e)}$ folgt weiter $Av_1 = \lambda \cdot v_1$, was auch dem Jordan-Kästchen entspricht. Die Vektoren v_i schlägt man der Jordan-Basis zu, und so verfährt man mit allen Vektoren, die eine Basis von $E_\lambda^{(e-1)}$ zu einer von $E_\lambda^{(e)}$ ergänzen. Nun hat man Basisvektoren, die zu den Jordan-Kästchen zum Eigenwert λ mit der maximalen Länge e gehören.

Es geht weiter mit den Basisvektoren zu den Jordan-Kästchen der Länge $e-1$ (falls vorhanden). Um lineare Abhängigkeit mit den schon in der Jordan-Basis befindlichen Vektoren zu vermeiden, muss man Basen von $E_\lambda^{(e-2)}$ und von $(A - \lambda I_n) \cdot E_\lambda^{(e)}$ zu einer Basis von $E_\lambda^{(e-1)}$ ergänzen. Eine Basis von $(A-\lambda I_n)\cdot E_\lambda^{(e)}$ erhält man hierbei aus den „Abkömmlingen“ v_{e-1} gemäß (11.1) der Vektoren aus der Basisergänzung von $E_\lambda^{(e-1)}$ zu $E_\lambda^{(e)}$.

Beispiel 11.7. Zur Illustration der Methode betrachten wir unsere Standardbeispiele.

(1) Wir betrachten wieder

$$A = \begin{pmatrix} -3 & -1 & 2 \\ 4 & 1 & -4 \\ 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Wir wissen, dass es zwei Jordan-Kästchen der Länge 1 und 2 zum Eigenwert -1 gibt (siehe Beispiel 11.6(1)). Der Eigenraum E_{-1} hat also die Dimension 2, der Hauptraum $E_{-1}^{(2)}$ muss also Dimension 3 haben. (Diese Dimensionen ergeben sich auch aus der Formel in Satz 11.5(a).) Wir können als „Keim" einer Jordanbasis also mit einem beliebigen Vektor außerhalb E_{-1} beginnen. Wir wählen den ersten Standardbasisvektor $v_2 := e_1$. Weiter setzen wir

$$v_1 := Av_2 + v_2 = \begin{pmatrix} -2 \\ 4 \\ 0 \end{pmatrix}.$$

Diese beiden Vektoren gehören zum Jordan-Kästchen der Länge 2. Um einen Basisvektor zum Jordan-Kästchen der Länge 1 zu bekommen, ergänzen wir v_1 durch

$$v_3 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

zu einer Basis von E_{-1}. In der Reihenfolge v_3, v_1, v_2 bilden unsere Vektoren eine Jordan-Basis zu der Jordanschen Normalform mit der Reihenfolge der Kästchen wie in Beispiel 11.6(1). Eine transformierende Matrix ist

$$S = \begin{pmatrix} 0 & -2 & 1 \\ 2 & 4 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(2) Nun betrachten wir unser zweites Standardbeispiel, nämlich

$$A = \begin{pmatrix} -3 & -1 & 4 & -3 & -1 \\ 1 & 1 & -1 & 1 & 0 \\ -1 & 0 & 2 & 0 & 0 \\ 4 & 1 & -4 & 5 & 1 \\ -2 & 0 & 2 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{5\times 5}$$

(siehe Beispiel 11.6(2)). Für den Eigenwert $\lambda = 2$ finden wir durch Lösen des entsprechenden homogenen LGS den Eigenvektor

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ -2 \end{pmatrix},$$

den wir als ersten Vektor in die Jordan-Basis aufnehmen. Nun behandeln wir den Eigenwert $\lambda = 1$ und suchen als erstes einen Vektor für das Jordan-Kästchen der Länge 3. Hierzu müssen wir $E_1^{(3)}$, also den Kern von $(A - I_5)^3$, berechnen. Wir kennen aus Beispiel 11.6(2) bereits die Ränge von $A - I_5$ und $(A - I_5)^2$ (nämlich 3 und 2), und erhalten $\operatorname{rg}\left((A - I_5)^3\right) = 1$ durch Auflösen der Formel aus Satz 11.5(a). Es genügt also, eine Zeile von $(A - I_5)^3$ zu berechnen, wobei wir $(A - I_5)^2$ schon aus Beispiel 11.6(2) kennen. Am einfachsten ist die dritte Zeile von $(A - I_5)^3$, die sich zu $(2, 0, -2, 2, 0)$ ergibt. Wir wählen

$$v_5 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \in E_1^{(3)} \setminus E_1^{(2)}.$$

und weiter

$$v_4 := (A - I_5) \cdot v_5 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad v_3 := (A - I_5) \cdot v_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Die Vektoren v_3, v_4, v_5 gehören zum Jordan-Kästchen der Länge 3, was wir durch Nachrechnen von

$$A \cdot v_3 = v_3, \quad A \cdot v_4 = v_4 + v_3 \quad \text{und} \quad A \cdot v_5 = v_5 + v_4$$

bestätigen können. Für das Jordan-Kästchen der Länge 1 brauchen wir einen Vektor aus $E_1^{(1)}$ (also einen Eigenvektor), der zusammen mit v_3 linear unabhängig ist. Wir haben $A - I_5$ in Beispiel 11.6(2) bereits mit Spaltenoperationen behandelt und sind auf die Matrix

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

gekommen, an der man die Basis

$$\begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

des Eigenraums $E_1^{(1)}$ abliest. Wir können also als letzten Basisvektor

$$v_2 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

wählen. Die Nummerierung der v_i haben wir so gemacht, dass sie mit der gewählten Reihenfolge der Jordan-Kästchen in Beispiel 11.6(2) kompatibel ist. Als transformierende Matrix erhält man

$$S = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

◁

Als wichtige und interessante Anwendung werden wir die Jordansche Normalform verwenden, um homogene lineare Differentialgleichungen zu lösen. Hierzu müssen wir uns zunächst mit der Exponentialfunktion einer Matrix beschäftigen. Dies wird motiviert durch folgenden „Analogieschluss“: Für die Differentialgleichung $\dot{y}(t) = ay(t)$ mit $a \in \mathbb{C}$ ist $y(t) = \exp(at) \cdot y(0) = e^{at} \cdot y(0)$ die allgemeine Lösung. Also sollte für das Differentialgleichungssystem

$$\begin{pmatrix} \dot{y}_1(t) \\ \vdots \\ \dot{y}_n(t) \end{pmatrix} = A \cdot \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

mit $A \in \mathbb{C}^{n \times n}$ die Lösung durch

$$\begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \exp(At) \cdot \begin{pmatrix} y_1(0) \\ \vdots \\ y_n(0) \end{pmatrix}$$

gegeben sein. Im Moment erscheint dieser Analogieschluss ausgesprochen gewagt, denn wir wissen nicht einmal, ob und wie $\exp(At)$ definiert ist. Bevor wir uns hiermit beschäftigen, bemerken wir (zur Verstärkung der Motivation), dass sich homogene Differentialgleichungen höherer Ordnung in homogene lineare Differentialgleichungssysteme übersetzen lassen: Statt beispielsweise $\ddot{y}(t) = a \cdot y(t)$ zu lösen, kann man ebenso das System

$$\begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix} \cdot \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$$

lösen.

Wir definieren nun die Exponentialfunktion einer Matrix durch den folgenden Satz.

Satz 11.8. *Ist $A \in \mathbb{C}^{n \times n}$ eine quadratische Matrix, so konvergiert die Reihe*

$$\sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

Hierbei setzen wir $A^0 := I_n$. Mit Konvergenz ist gemeint, dass für alle $1 \leq i, j \leq n$ die Folge der (i,j)-ten Einträge der Teilsummen $\sum_{k=0}^{N} \frac{1}{k!} A^k$ für $N \to \infty$ gegen einen Grenzwert $c_{i,j} \in \mathbb{C}$ konvergiert. Wir schreiben $\exp(A) = e^A := (c_{i,j}) \in \mathbb{C}^{n \times n}$ *oder auch* $\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$.

Beweis. Um das Majorantenkriterium zu benutzen, müssen wir die Absolutbeträge aller Einträge von $\frac{1}{k!} A^k$ nach oben durch Glieder einer konvergenten Reihe abschätzen. Wir wählen $C \in \mathbb{R}$ so, dass $|a_{i.j}| \leq C$ für alle Einträge $a_{i,j}$ von A gilt. Wir zeigen nun per Induktion nach k, dass dann für den (i,j)-ten Eintrag von A^k gilt:

$$|(A^k)_{i,j}| \leq n^{k-1} C^k.$$

Dies ist korrekt für $k = 1$. Der Induktionsschluss lautet

$$|(A^{k+1})_{i,j}| = \left| \sum_{l=1}^{n} (A^k)_{i,l} a_{l,j} \right| \leq \sum_{l=1}^{n} |(A^k)_{i,l}||a_{l,j}| \leq n \cdot n^{k-1} C^k \cdot C = n^k C^{k+1}.$$

Damit wird jeder Eintag unserer Reihe (außer dem 0-ten Summanden) majorisiert durch die Reihe

$$\sum_{k=0}^{\infty} \frac{1}{k!} n^{k-1} C^k = \frac{1}{n} \exp(nC),$$

welche konvergiert. Damit konvergiert $\exp(A)$ sogar absolut. □

Nachdem der Begriff (und die Existenz) der Exponentialfunktion einer Matrix geklärt sind, stellt sich die Frage der Berechnung. Für einige Matrizen ist dies direkt möglich:

Beispiel 11.9. (1) Für $A = \left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix} \right) \in \mathbb{C}^{2 \times 2}$ gilt

$$A^k = \begin{pmatrix} a^k & 0 \\ 0 & b^k \end{pmatrix},$$

also

$$\exp(A) = \begin{pmatrix} \exp(a) & 0 \\ 0 & \exp(b) \end{pmatrix}.$$

Es ist klar, dass sich dies auf beliebige Diagonalmatrizen verallgemeinert.

(2) Für

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

gilt

$$A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{und} \quad A^k = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{für} \quad k \geq 3,$$

also

$$\exp(A) = I_3 + A + \frac{1}{2}A^2 = \begin{pmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Allgemein ist die Exponentialfunktion einer nilpotenten Matrix A (also einer Matrix mit $A^k = 0$ für ein k) ein Polynom in A. ◁

Für die Berechnung der Exponentialfunktion einer allgemeinen Matrix kommt nun die Jordansche Normalform ins Spiel. Da jede Matrix in $\mathbb{C}^{n \times n}$ ähnlich ist zu einer Matrix in Jordanscher Normalform, reduziert die folgende Proposition das Problem auf die Berechnung der Exponentialfunktion einer Matrix in Jordanscher Normalform.

Proposition 11.10. *Es seien $A, B \in \mathbb{C}^{n \times n}$ ähnlich, also $B = S^{-1}AS$ mit $S \in \mathrm{GL}_n(\mathbb{C})$. Dann gilt*

$$\exp(B) = S^{-1} \exp(A) S.$$

Beweis. Für die N-te Teilsumme gilt:

$$\sum_{k=0}^{N} \frac{1}{k!} B^k = \sum_{k=0}^{N} \frac{1}{k!} (S^{-1}AS)^k = \sum_{k=0}^{N} \frac{1}{k!} S^{-1} A^k S = S^{-1} \left(\sum_{k=0}^{N} \frac{1}{k!} A^k \right) S.$$

Die Behauptung folgt also, wenn wir allgemein für eine Folge A_N von Matrizen in $\mathbb{C}^{n \times n}$ zeigen können, dass aus der Konvergenz $A_N \to M \in \mathbb{C}^{n \times n}$ folgt, dass $S^{-1}A_N S$ gegen $S^{-1}MS$ konvergiert. Indem wir A_N durch $A_N - M$ ersetzen, können wir annehmen, dass A_N eine Nullfolge ist, und müssen dasselbe für $S^{-1}A_N S$ zeigen. Wir schreiben $S = (s_{i,j})$, $S^{-1} = (\tilde{s}_{i,j})$ und wählen $C \in \mathbb{R}$ als obere Schranke für alle $|s_{i,j}|$ und $|\tilde{s}_{i,j}|$.

Für $\varepsilon > 0$ gibt es $N_0 \in \mathbb{N}$, so dass für $N \geq N_0$ und für $1 \leq i, j \leq n$ gilt:

$$|(A_N)_{i,j}| \leq \frac{\varepsilon}{n^2C^2}.$$

Für den $(i.j)$-ten Eintag von $S^{-1}A_NS$ folgt

$$|(S^{-1}A_NS)_{i,j}| = \left|\sum_{k,l=1}^{n} \tilde{s}_{i,k}(A_N)_{k,l}s_{l,j}\right| \leq$$
$$\sum_{k,l=1}^{n} |\tilde{s}_{i,k}||(A_N)_{k,l}||s_{l,j}| \leq n^2C^2\frac{\varepsilon}{n^2C^2} = \varepsilon.$$

Also ist $S^{-1}A_NS$ eine Nullfolge, wie behauptet. □

Aus Beispiel 11.9 wissen wir, wie man die Exponentialfunktion einer Diagonalmatrix berechnet, und auch, wie es bei einem Jordan-Kästchen zum Eigenwert 0 und allgemeiner bei nilpotenten Matrizen geht. Nun kann man jede Matrix $B \in \mathbb{C}^{n\times n}$, die in Jordanscher Normalform ist, als Summe der entsprechenden Diagonalmatrix D und der Matrix mit Eintägen auf der oberen Nebendiagonale schreiben:

$$B = D + N \tag{11.2}$$

mit N eine nilpotente Matrix. Ist e die maximale Länge eines Jordan-Blocks, so folgt $N^e = 0$. Außerdem gilt $DN = ND$, d.h. D und N kommutieren, was man leicht anhand der einzelnen Jordan-Blöcke überprüft. Nebenbei sei gesagt, dass eine Darstellung einer Matrix als Summe zweier kommutierender Matrizen, von denen eine diagonalisierbar und die andere nilpotent ist, die *additive Jordan-Zerlegung* genannt wird. Man kann zeigen, dass sie eindeutig ist. Folgendes Lemma zeigt, wie wichtig die Bedingung ist, dass beide Matrizen kommutieren.

Lemma 11.11. *Sind seien $A, B \in \mathbb{C}^{n\times n}$ zwei kommutierende Matrizen, so gelten:*

(a)
$$(A+B)^n = \sum_{i=0}^{n} \binom{n}{i} A^i B^{n-i}$$
für $n \in \mathbb{N}_0$, und

(b)
$$\exp(A+B) = \exp(A)\cdot\exp(B).$$

Beweis. Beide Nachweise folgen exakt den Beweisen der entsprechenden Aussagen in der Analysis (Binomialformel und Additionstheorem der Exponentialfunktion), wobei (b) aus (b) hergeleitet wird. Für (a) braucht man die Voraussetzung $AB = BA$, denn nur mit dieser kann man z.B.

$$(A+B)^2 = A^2 + AB + BA + B^2 = A^2 + 2AB + B^2$$

rechnen. □

Nun ist klar, wie man die Expontialfunktion einer Matrix $A \in \mathbb{C}^{n\times n}$ berechnen kann: Zunächst bestimmt man die Jordansche Normalform B und eine Matrix $S \in \mathrm{GL}_n(\mathbb{C})$ mit $A = S^{-1}BS$. Dann zerlegt man B gemäß (11.2) in eine Diagonalmatrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ und eine nilpotente Matrix N. Ist e die maximale Länge eines Jordan-Blocks, so gilt

$$\exp(A) = S^{-1} \operatorname{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n}) \cdot \left(\sum_{k=0}^{e-1} \frac{1}{k!} N^k\right) S.$$

Wie wir in der Motivation für die Exponentialfunktion einer Matrix gesehen haben, benötigen wir auch $\exp(At)$ für Skalare t. (Wir schreiben hier ausnahmsweise At statt tA.) Da $At = S^{-1}BtS$ und da Bt die additive Jordan-Zerlegung $Bt = Dt + Nt$ hat, ergibt sich

$$\exp(At) = S^{-1} \operatorname{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}) \cdot \left(\sum_{k=0}^{e-1} \frac{1}{k!} N^k t^k\right) S. \tag{11.3}$$

Nun kommen wir zurück auf unsere ursprüngliche Motivation, das Lösen von homogenen Systemen von linearen Differentialgleichungen erster Ordnung mit konstanten Koeffizienten. Zunächst erinnern wir uns, dass man die Ableitung einer Potenzreihe innerhalb des Konvergenzradius durch Ableitung der Reihenglieder erfolgt. Nach Satz 11.8 ist jeder Eintrag von $\exp(At)$ eine überall konvergente Potenzreihe in t, also ergibt sich

$$\frac{d}{dt}\exp(At) = \sum_{k=0}^{\infty} \frac{d}{dt}\frac{1}{k!}A^k t^k = \sum_{k=1}^{\infty} \frac{1}{(k-1)!}A^k t^{k-1} = A \cdot \exp(At).$$

Also erfüllt für jedes $v \in \mathbb{C}^n$

$$y(t) := \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix} := \exp(At) \cdot v$$

die Differenzialgleichung $\dot{y}(t) = Ay(t)$ und $y(0) = v$, genau wie in unserem Analogieschluss vermutet. Insgesamt erfüllt y das durch $\dot{y}(t) = A \cdot y(t)$ und $y(0) = v$ gegebene *Anfangswertproblem*. Umgekehrt ist die Lösung auch eindeutig: Ist nämlich $y(t)$ eine Lösung von $\dot{y}(t) = Ay$ und $y(0) = v$, so folgt mit der Produktregel

$$\frac{d}{dt}\left(\exp(-At) \cdot y(t)\right) = -A\exp(-At)y(t) + \exp(-At)Ay(t) = 0,$$

da A mit jedem Glied der Reihe $\exp(-At)$ kommutiert, also auch mit $\exp(-At)$. Es folgt die Konstanz von $\exp(-At) \cdot y(t)$, also $\exp(-At) \cdot y(t) = \exp(0) \cdot y(0) = v$ und wegen Lemma 11.11 $y(t) = \exp(At)v$.

Wegen (11.3) ist jeder Eintrag $y_i(t)$ eine Linearkombination von Funktionen der Form $t^k e^{\lambda_j t}$, wobei $k > 0$ nur vorkommt, wenn A nicht diagonalisierbar ist. Wenn wir $\lambda_j = \rho_j + \omega_j \sqrt{-1}$ mit $\rho_j, \omega_j \in \mathbb{R}$ schreiben, so gilt

$$e^{\lambda_j t} = e^{\rho_j t} \left(\cos(\omega_j t) + \sqrt{-1} \sin(\omega_j t) \right),$$

was sich im Falle $\rho_j < 0$ physikalisch als eine gedämpfte Schwingung interpretieren lässt.

Die Anwendung auf Differentialgleichungen unterstreicht die Bedeutung der Jordanschen Normalform. Allerdings wird dies durch folgende Bemerkung etwas relativiert.

Anmerkung 11.12. Ein System von Differentialgleichungen der Form $\dot{y}(t) = Ay(t)$ (mit $A \in \mathbb{C}^{n \times n}$) lässt sich auch mit Hilfe einer Trigonalisierung von A gemäß Satz 10.19 lösen. Es sei nämlich

$$S^{-1}AS = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

mit $S \in \mathrm{GL}_n(K)$, $\lambda_i \in \mathbb{C}$. Mit $z(t) := S^{-1}y(t)$ ergibt sich

$$\dot{z}(t) = S^{-1}\dot{(y)}(t) = S^{-1}Ay(t) = S^{-1}ASz(t) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} z(t),$$

also

$$\dot{z_i}(t) = \lambda_i z_i(t) + \sum_{j=i+1}^{n} a_{i,j} z_j(t)$$

für $i = 1, \ldots, n$, wobei $a_{i,j} \in \mathbb{C}$. Dies ergibt

$$\frac{d}{dt}\left(e^{-\lambda_i t} z_i(t)\right) = -\lambda_i e^{-\lambda_i t} z_i(t) + e^{-\lambda_i t} \left(\lambda_i z_i(t) + \sum_{j=i+1}^{n} a_{i,j} z_j(t) \right) = \sum_{j=i+1}^{n} a_{i,j} e^{-\lambda_i t} z_j(t). \quad (11.4)$$

Hiermit lassen sich die z_i iterativ berechnen. Wir starten mit $z_n(t) = e^{\lambda_n t} z_n(0)$. Sind nun $z_{i+1}, \ldots, z_n$ bestimmt, so lässt sich $e^{-\lambda_i t} z_i(t)$ durch Integration aus (11.4) berechnen. Da alle z_j gemäß der obigen Beobachtung Linearkombinationen von Funktionen der Form $t^k e^{\lambda_l t}$ sind, lassen sich deren

Stammfunktionen berechnen. Die Lösung $y(t)$ des ursprünglichen Systems ergibt sich dann aus $y(t) = S \cdot z(t)$. Für die Lösung des Eigenwertproblems müssen die Konstanten bei der Wahl der Stammfunktion jeweils angepasst werden. ◁

Kapitel 12
Skalarprodukte

Bis jetzt haben wir die gesamte Theorie über beliebigen Körpern entwickelt. Dabei hat jeglicher Begriff von „Abstand“ gefehlt. Die Einführung eines Abstandsbegriffs ist über allgemeinen Körpern auch nicht (in geometrisch sinnvoller Weise) möglich. Nun spezialisieren wir den Grundkörper zu $\mathbb{R}$ oder $\mathbb{C}$ und führen das Skalarprodukt ein. Mit diesem werden dann Längen, Abstände und auch Winkel definiert.

Auf $\mathbb{R}^n$ ist das **Standard-Skalarprodukt** zweier Vektoren $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ und $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ durch

$$\langle v, w \rangle := \sum_{i=1}^{n} x_i y_i \quad (= v^T \cdot w) \in \mathbb{R}$$

definiert. Achtung: Die Notation ist anfällig für Verwechselungen mit dem Erzeugnis!

Es gelten die folgenden Regeln:

(a) Für alle $u, v, w \in \mathbb{R}^n$ und $a \in \mathbb{R}$ gelten:

$$\langle u, v + a \cdot w \rangle = \langle u, v \rangle + a \cdot \langle u, w \rangle$$

und

$$\langle u + a \cdot v, w \rangle = \langle u, w \rangle + a \cdot \langle v, w \rangle.$$

(Man sagt auch, dass das Skalarprodukt **bilinear** ist.)

(b) Für $v, w \in \mathbb{R}^n$ gilt

$$\langle v, w \rangle = \langle w, v \rangle.$$

(Man sagt auch, dass das Skalarprodukt **symmetrisch** ist.)

(c) Für $v \in \mathbb{R}^n$ mit $v \neq 0$ gilt

$$\langle v, v \rangle > 0.$$

(Man sagt auch, dass das Skalarprodukt **positiv definit** ist.)

Wir nehmen dies zum Anlass für folgende Definition:

Definition 12.1. *Es sei V ein* reeller Vektorraum *(d.h. ein Vektorraum über $\mathbb{R}$). Eine Abbildung*

$$V \times V \to \mathbb{R}, \ (v, w) \mapsto \langle v, w \rangle$$

heißt eine **symmetrische Bilinearform***, falls sie symmetrisch und bilinear ist. Eine symmetrische Bilinearform heißt ein* **Skalarprodukt***, wenn sie zusätzlich positiv definit ist.*

Ein reeller Vektorraum zusammen mit einem Skalarprodukt heißt ein **euklidischer Raum**.

Beispiel 12.2. (a) Für reelle Zahlen $a < b$ sei $V := C([a, b], \mathbb{R})$ der Vektorraum aller stetiger Funktionen $[a, b] \to \mathbb{R}$ auf dem abgeschlossenen Intervall $[a, b]$. Durch

$$\langle f, g \rangle := \int_a^b f(x) g(x) dx$$

wird ein Skalarprodukt auf V definiert.

(b) Auf $\mathbb{R}^2$ wird für $v = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ und $w = \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)$ ein Skalarprodukt erklärt durch

$$\langle v, w \rangle = 5x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2.$$

Die Bilinearität und Symmetrie sind klar, und die positive Definitheit geht aus

$$\langle v, v \rangle = 5x_1^2 + 6x_1x_2 + 2x_2^2 = (2x_1 + x_2)^2 + (x_1 + x_2)^2$$

hervor.

(c) Ebenso wie oben kann man

$$\langle v, v \rangle = x_1y_1 - x_2y_2$$

definieren und erhält ein Beispiel für eine nicht positiv definite, symmetrische Bilinearform. ◁

Zu einer symmetrischen Bilinearform auf $\mathbb{R}^n$ erhält man durch Einsetzen der Standardbasisvektoren Zahlen $a_{i,j} := \langle e_i, e_j \rangle$, die man zu einer Matrix $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ zusammenfassen kann. A ist symmetrisch und wird die **Darstellungsmatrix** der symmetrischen Bilinearform genannt. Die Bilinearform wird durch A „codiert,“ denn für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ und $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ gilt

$$\langle v, w \rangle = \langle \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j \rangle = \sum_{i,j=1}^{n} x_i y_j a_{i,j} = v^T \cdot A \cdot w. \tag{12.1}$$

Die Darstellungsmatrix des Standard-Skalarprodukts ist die Einheitsmatrix.

Allgemeiner kann man auch Darstellungsmatrizen von symmetrischen Bilinearformen auf endlich-dimensionalen Vektorräumen betrachten, indem man eine Basis wählt und die Basisvektoren in die Form einsetzt. Nun kann man auch überlegen, wie sich ein Basiswechsel auf die Darstellungsmatrix auswirkt. Wir werden dieses Thema nicht weiter verfolgen, sondern uns nun mit komplexen Vektorräumen beschäftigen.

In einem *komplexen Vektorraum* V (d.h. einem Vektorraum über $\mathbb{C}$) kann es kein Skalarprodukt im Sinne von Definition 12.1 geben (es sei denn, $V = \{0\}$). Denn für $0 \neq v \in V$ müsste $\langle v, v \rangle > 0$ gelten, also

$$\langle iv, iv \rangle = i^2 \langle v, v \rangle = -\langle v, v \rangle < 0.$$

(Darüber hinaus wäre beispielsweise $\langle (i+1) \cdot v, (i+1) \cdot v \rangle = 2i \langle v, v \rangle$ nicht einmal reell.) Man behilft sich, indem man die *komplexe Konjugation* benutzt, die wir nun in Erinnerung rufen: Für $z = a + bi \in \mathbb{C}$ ist das **komplex konjugierte**

$$\overline{z} := a - bi \in \mathbb{C}.$$

Man rechnet nach, dass für $z, w \in \mathbb{C}$ die Regeln

$$\overline{z + w} = \overline{z} + \overline{w} \quad \text{und} \quad \overline{z \cdot w} = \overline{z} \cdot \overline{w}$$

gelten. Außerdem gilt

$$\overline{z} \cdot z = a^2 + b^2 \in \mathbb{R}_{\geq 0},$$

was die Definition des Betrags $\mid z \mid := \sqrt{\overline{z} \cdot z}$ möglich macht. Nur die Null hat den Betrag Null.

Das Standard-Skalarprodukt auf $\mathbb{R}^n$ wird nun ersetzt durch das Produkt

$$\langle v, w \rangle := \sum_{i=1}^{n} \overline{x}_i y_i \quad (= \overline{v}^T \cdot w) \in \mathbb{C}$$

für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ und $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}^n$ ersetzt. Dies ist ein komplexes Skalarprodukt gemäß der folgenden Definition.

Definition 12.3. *Es sei V ein komplexer Vektorraum. Eine Abbildung*

$$V \times V \to \mathbb{C}, \ (v, w) \mapsto \langle v, w \rangle$$

heißt

(a) **sesquilinear***, falls für $u, v, w \in V$ und $a \in \mathbb{C}$ die Regeln*

$$\langle u, v + a \cdot w \rangle = \langle u, v \rangle + a \cdot \langle u, w \rangle$$

und

$$\langle u + a \cdot v, w \rangle = \langle u, w \rangle + \overline{a} \cdot \langle v, w \rangle$$

gelten;

(b) **hermitesch**, *falls für $v, w \in V$ die Regel*

$$\langle v, w \rangle = \overline{\langle w, v \rangle}$$

gilt;

(c) **positiv definit**, *falls für $v \in V \setminus \{0\}$*

$$\langle v, v \rangle \in \mathbb{R} \quad \textit{und} \quad \langle v, v \rangle > 0$$

gilt.

Man spricht dann auch von einer **Sesquilinearform** *bzw. einer* **hermiteschen Form**. *Eine positiv definite, hermitesche Sesquilinearform heißt ein* **komplexes Skalarprodukt**.

Ein komplexer Vektorraum zusammen mit einem komplexen Skalarprodukt heißt ein **unitärer Raum**.

Anmerkung. Man drückt die Bedingung der Sesquilinearität auch aus, indem man sagt, dass die Form linear im zweiten und *semilinear* im ersten Argument ist. Einige Autoren treffen die umgekehrte Konvention, indem sie Linearität im ersten und Semilinearität im zweiten Argument fordern. ◁

Beispiel 12.4. Für reelle Zahlen $a < b$ sei $V := C([a, b], \mathbb{C})$ der Vektorraum aller stetiger Funktionen $[a, b] \to \mathbb{C}$ auf dem abgeschlossenen Intervall $[a, b] \subseteq \mathbb{R}$. Durch

$$\langle f, g \rangle := \int_a^b \overline{f(x)} g(x) dx$$

wird ein komplexes Skalarprodukt auf V definiert. ◁

Zu einer hermiteschen Sesquilinearform auf einem endlich-dimensionalen Vektorraum mit einer Basis $\{v_1, \ldots, v_n\}$ erhält man eine Matrix $A = (a_{i,j}) \in \mathbb{C}^{n \times n}$ durch $a_{i,j} := \langle v_i, v_j \rangle$. Es folgt $a_{i,j} = \overline{a_{j,i}}$ für alle $i, j \in \{1, \ldots, n\}$, also

$$A^T = \overline{A}.$$

Matrizen mit dieser Eigenschaft nennt man **hermitesch**. Die Darstellungsmatrizen von hermiteschen Sesquilinearformen sind also hermitesche Matrizen.

Von nun an sei V ein euklidischer oder unitärer Raum. Wir kommen nun zum Abstands- und Längenbegriff.

Definition 12.5. *Für $v \in V$ heißt*

$$|| \, v \, || := \sqrt{\langle v, v \rangle} \in \mathbb{R}_{\geq 0}$$

die **Länge** *(auch:* **Norm***) von v.*

Für $v, w \in V$ heißt

$$d(v,w) := ||v - w|| \in \mathbb{R}_{\geq 0}$$

der **Abstand** *von v und w.*

Proposition 12.6 (Schwarzsche Ungleichung)**.** *Für $v, w \in V$ gilt*

$$|\langle v, w\rangle| \leq ||v|| \cdot ||w||.$$

Hierbei gilt Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis. Wir können $w \neq 0$ annehmen, da für $w = 0$ die Ungleichung und die Zusatzbehauptung erfüllt sind.

Für $a \in \mathbb{R}$ oder (im Falle eines komplexen Vektorraums) $a \in \mathbb{C}$ gilt

$$0 \leq ||v - aw||^2 = \langle v - aw, v - aw\rangle = ||v||^2 - a\langle v, w\rangle - \overline{a}\langle w, v\rangle + \overline{a}a||w||^2.$$

Speziell für $a = \frac{\langle w,v\rangle}{||w||^2}$ ergibt dies

$$\begin{aligned} 0 &\leq ||v||^2 - \frac{\langle w,v\rangle\langle v,w\rangle}{||w||^2} - \frac{\overline{\langle w,v\rangle}\langle w,v\rangle}{||w||^2} + \frac{\overline{\langle w,v\rangle}\langle w,v\rangle}{||w||^2} \\ &= \frac{1}{||w||^2}\left(||v||^2||w||^2 - |\langle v,w\rangle|^2\right). \end{aligned}$$

Dies liefert die Ungleichung und zeigt, dass genau dann Gleichheit gilt, wenn $v = \frac{\langle w,v\rangle}{||w||^2} \cdot w$. Die lineare Abhängigkeit ist also notwendig für die Gleichheit. Ist umgekehrt $v = aw$ mit $a \in \mathbb{R}$ bzw. $a \in \mathbb{C}$, so folgt

$$\frac{\langle w,v\rangle}{||w||^2} = \frac{a||w||^2}{||w||^2} = a,$$

also Gleichheit. □

Nun können wir die wichtigsten Eigenschaften der Länge und des Abstands beweisen.

Satz 12.7. *Für alle $u, v, w \in V$ und $a \in \mathbb{R}$ bzw. $a \in \mathbb{C}$ gelten:*

(a) Falls $v \neq 0$, so folgt $||v|| > 0$.
(b) $||a \cdot v|| = |a| \cdot ||v||$.
*(c) $||v + w|| \leq ||v|| + ||w||$ (*Dreiecksungleichung*).*
(d) Falls $v \neq w$, so folgt $d(v,w) > 0$.
(e) $d(v,w) = d(w,v)$.
*(f) $d(u,w) \leq d(u,v) + d(v,w)$ (*Dreiecksungleichung*).*

Beweis. Die Teile (a), (b), (d) und (e) sind unmittelbar klar. Für den Nachweis von (c) rechnen wir:

$$\begin{aligned} ||v+w||^2 &= ||v||^2 + \langle v,w\rangle + \langle w,v\rangle + ||w||^2 = ||v||^2 + 2\,\mathrm{Re}\left(\langle v,w\rangle\right) + ||w||^2 \\ &\le ||v||^2 + 2\,|\langle v,w\rangle| + ||w||^2 \underset{\text{Proposition 12.6}}{\le} ||v||^2 + 2||v||\cdot||w|| + ||w||^2 \\ &= \left(||v|| + ||w||\right)^2, \end{aligned}$$

wobei $\mathrm{Re}(z) := a$ für $z = a + bi \in \mathbb{C}$ den Realteil bezeichnet. Der Nachweis von (f) wird durch

$$d(u,w) = ||u-w|| = ||u-v+v-w|| \underset{(c)}{\le} ||u-v|| + ||v-w|| = d(u,v) + d(v,w)$$

erbracht. □

Wir nehmen diesen Satz zum Anlass, ein paar Begriffe zu erwähnen, die in dieser Vorlesung nicht weiter vorkommen werden, aber insbesondere für die Quantenmechanik wichtig sind.

Anmerkung 12.8. (a) Ein **normierter Vektorraum** ist ein reeller oder komplexer Vektorraum V mit einer Abbildung

$$V \to \mathbb{R}_{\ge 0},\ v \mapsto ||v||,$$

die (a)–(c) aus Satz 12.7 erfüllt.

(b) Ein **metrischer Raum** ist eine Menge V mit einer Abbildung

$$d\colon V \times V \to \mathbb{R}_{\ge 0},$$

die (d)–(f) aus Satz 12.7 erfüllt. Die Abbildung d heißt dann eine **Metrik** auf V.

(c) Sobald man einen Abstandsbegriff hat, kann man von konvergenten Folgen und von Cauchy-Folgen sprechen. Vollständigkeit bedeutet, dass jede Cauchy-Folge konvergent ist. Ein **Banachraum** ist ein vollständiger normierter Raum. Ein **Hilbertraum** ist ein vollständiger euklidischer oder unitärer Raum. ◁

Wir erhalten eine hierarchische Anordnung unserer Begriffe: Jeder euklidische oder unitäre Raum ist normiert, und jeder normierte Raum ist metrisch. Jeder Hilbertraum ist ein Banachraum.

Beispiel 12.9. (1) Beispiele für Normen, die nicht von einem Skalarprodukt kommen, sind die *Manhattan-Norm* auf $\mathbb{R}^n$, definiert durch

$$||v|| = \sum_{i=1}^{n} |v_i|$$

(wobei v_i die Komponenten von $v \in \mathbb{R}^n$ sind) und die *Maximum-Norm* auf $C([a,b],\mathbb{C})$, definiert durch

$$||f|| := \max\left\{|f(x)| \mid x \in \mathbb{R},\ a \le x \le b\right\}.$$

(2) Ein Beispiel für eine Metrik, die nicht von einer Norm kommt, ist die *Hamming-Metrik* auf $\mathbb{R}^n$ (oder K^n mit einem Körper K), definiert durch

$$d(v,w) := |\{i \in \{1,\dots,n\} \mid v_i \neq w_i\}|,$$

wobei v_i und w_i die Komponenten von $v, w \in \mathbb{R}^n$ sind.

(3) Es ist nicht schwer zu zeigen, dass jeder endlich-dimensionale euklidische oder unitäre Raum ein Hilbertraum ist. Ebenso ist jeder endlich-dimensionale normierte Raum ein Banachraum.

(4) Der euklidische Raum $C([a,b],\mathbb{R})$ (siehe Beispiel 12.2(a)) ist nicht vollständig, also kein Hilbertraum.

(5) Man kann zeigen, dass $C([a,b],\mathbb{R})$ und $C([a,b],\mathbb{C})$ zusammen mit der Maximum-Norm (siehe (1)) Banachräume sind. Der durch die Maximum-Norm gegebene Konvergenzbegriff ist die gleichmäßige Konvergenz.

(6) Das wohl einfachste Beispiel für einen unendlich-dimensionalen Hilbertraum ist der Raum ℓ^2 aller komplexer Folgen $\mathbf{a} = (a_n)$ mit der Eigenschaft, dass $\sum_{n=1}^{\infty} |a_n|^2$ konvergiert. Das Skalarprodukt wird durch

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{n=1}^{\infty} \overline{a}_n b_n$$

definiert. Der Nachweis der Vollständigkeit von ℓ^2 ist nicht ganz einfach.

(7) Schwieriger (aber wichtiger) ist der L^2-Raum: Man betrachtet die Menge $L^2(\mathbb{R})$ aller Funktionen $f\colon \mathbb{R} \to \mathbb{C}$, für die das Integral $\int_{-\infty}^{\infty} |f(x)|^2 dx$ existiert (und endlich ist). Dann wird durch

$$\langle f, g \rangle := \int_{-\infty}^{\infty} \overline{f(x)} g(x) dx$$

eine hermitesche Form definiert. Sie ist allerdings nicht positiv definit, da auch Funktionen, die nicht überall 0 sind, verschwindendes Integral haben können. Um aus der obigen Definition ein Skalarprodukt zu machen, muss man solche Funktionen als Null auffasen (Bildung eines „Faktorraums“). Nun erhält man einen unitären Raum, der auch mit $L^2(\mathbb{R})$ bezeichnet wird. Man kann zeigen, dass dieser ein Hilbertraum ist. ◁

Die Schwarzsche Ungleichung (Proposition 12.6) ermöglicht es, für Vektoren $v, w \in V$ positiver Länge in einem *euklidischen* Raum den **Winkel** zwischen v und w als die eindeutig bestimmte Zahl α in dem abgeschlossenen Intervall $[0,\pi]$ mit

$$\cos(\alpha) = \frac{\langle v, w \rangle}{||v|| \cdot ||w||}$$

zu definieren. Diese Definition erscheint zunächst willkürlich, sie liefert aber genau das Erwartete.

Beispiel 12.10. Für $v = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ und $w = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) \in \mathbb{R}^2$ ist

$$\frac{\langle v, w\rangle}{||v|| \cdot ||w||} = \frac{1}{\sqrt{2}},$$

also beträgt der Winkel $\pi/4$. ◁

In unitären Räumen lässt sich kein sinnvoller Winkelbegriff definieren, man kann aber (ebenso wie in euklidischen Räumen) davon sprechen, dass zwei Vektoren senkrecht aufeinander stehen. Dies ist Inhalt der folgenden Definition.

Definition 12.11. *Es sei V ein euklidischer oder unitärer Raum.*

(a) Zwei Vektoren $v, w \in V$ heißen **orthogonal** *(gleichbedeutend:* **senkrecht***), falls*

$$\langle v, w\rangle = 0.$$

(b) Eine Menge $S \subseteq V$ heißt ein **Orthogonalsystem***, falls je zwei Vektoren $v, w \in S$ mit $v \neq w$ orthogonal sind.*
(c) Ein Orthogonalsystem $S \subseteq V$ heißt ein **Orthonormalsystem***, falls zusätzlich alle Vektoren $v \in S$ die Länge $||v|| = 1$ haben.*
(d) Ein Orthonormalsystem $S \subseteq V$ heißt **Orthonormalbasis***, falls es zusätzlich eine Basis ist.*
(e) Zu einem Unterraum $U \subseteq V$ heißt

$$U^{\perp} := \{v \in V \mid \langle v, u\rangle = 0 \text{ für alle } u \in U\}$$

das **orthogonale Komplement** *von U. Es ist klar, dass $U^{\perp}$ ein Unterraum von V ist.*

Beispiel 12.12. (1) Die Standardbasis ist eine Orthonormalbasis von $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ mit dem Standard-Skalarprodukt.
(2) Die Vektoren

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix} \quad \text{und} \quad v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}$$

bilden ein Orthonormalsystem im $\mathbb{R}^3$.
(3) Im Raum $C([0, 2\pi], \mathbb{C})$ der stetigen komplexen Funktionen auf den Intervall $[0, 2\pi]$ mit dem Skalarprodukt aus Beispiel 12.4 bilden die Funktionen

$$f_n(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{int} \quad (n \in \mathbb{Z})$$

ein Orthonormalsystem. Die Theorie der Fourierreihen basiert hierauf. ◁

Satz 12.13. *Jedes Orthogonalsystem $S \subseteq V$ in einem euklidischen oder unitären Vektorraum, das nicht den Nullvektor enthält, ist linear unabhängig. Falls $|S| = \dim(V) < \infty$, so ist S eine Basis.*

Beweis. Seien $v_1, \dots, v_n \in S$ paarweise verschieden. Weiter sei

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

mit $a_i \in \mathbb{R}$ bzw. $a_i \in \mathbb{C}$. Für alle $j \in \{1, \dots, n\}$ folgt durch Bildung des Skalaprodukts mit v_j:

$$0 = \langle v_j, 0 \rangle = \left\langle v_j, \sum_{i=1}^{n} a_i v_i \right\rangle = \sum_{i=1}^{n} a_i \langle v_j, v_i \rangle = a_j \langle v_j, v_j \rangle.$$

Wegen $v_j \neq 0$ sind also sind alle $a_j = 0$, und die lineare Unabhängigkeit ist bewiesen.

Die zweite Aussage folgt mit Korollar 6.13(a). □

Orthonormalbasen haben einige günstige Eigenschaften. Ist beispielsweise $S = \{v_1, \dots, v_n\}$ eine Orthonormalbasis eines endlich-dimensionalen euklidischen oder unitären Vektorraums und $v \in V$, so sind die Skalarprodukte $\langle v_i, v \rangle$ genau die Koordinaten von v bezüglich der Basis S. Gilt nämlich $v = a_1 v_1 + \dots + a_n v_n$, so folgt

$$\langle v_i, v \rangle = \left\langle v_i, \sum_{j=1}^{n} a_j v_j \right\rangle = \sum_{j=1}^{n} a_j \langle v_i, v_j \rangle = a_i \langle v_i, v_i \rangle = a_i.$$

Mit Orthonormalbasen lassen sich also Koeffizienten „isolieren“. Es stellt sich die Frage, ob jeder endlich-dimensionale euklidische oder unitäre Vektorraum eine Orthonormalbasis hat. Diese Frage werden wir konstruktiv durch das Schmidtsche Orthogonalisierungsverfahren beantworten.

Algorithmus 12.14 (Schmidtsches Orthogonalisierungsverfahren)**.**

Eingabe: Vektoren $v_1, \dots, v_k$ eines euklidischen oder unitären Vektorraums V.

Ausgabe: Eine Orthonormalbasis $\{u_1, \dots, u_m\}$ des von den v_i erzeugten Unterraums von V.

(1) Setze $m := 0$.
(2) Für $i = 1, \dots, k$ führe Schritte (3) und (4) aus.
(3) Setze

$$w_i := v_i - \sum_{j=1}^{m} \langle u_j, v_i \rangle \cdot u_j. \tag{12.2}$$

(Im Fall $m = 0$ bedeutet dies $w_i := v_i$.)

(4) Falls $w_i \neq 0$, setze $m := m + 1$ und

$$u_m := \frac{1}{||w_i||} \cdot w_i.$$

Satz 12.15. *Algorithmus 12.14 liefert eine Orthonormalbasis von* $\langle v_1, \ldots, v_k \rangle \subseteq V$*.*

Beweis. Wir benutzen Induktion nach der Anzahl k der Erzeuger von V und können $k > 0$ voraussetzen. Nach Induktion gelten nach Durchlaufen der Schleife für $i = 1, \ldots, k-1$:

$$\langle u_i, u_j \rangle = \delta_{i,j} \quad (1 \le i, j \le m) \tag{12.3}$$

und

$$\langle v_1, \ldots, v_{k-1} \rangle = \langle u_1, \ldots, u_m \rangle, \tag{12.4}$$

wobei m das „aktuelle“ m nach $k-1$ Schleifendurchläufen ist. Aus (12.2) folgt für $i \le m$

$$\langle u_i, w_k \rangle = \langle u_i, v_k \rangle - \sum_{j=1}^{m} \langle u_j, v_k \rangle \cdot \langle u_i, u_j \rangle \underset{(12.3)}{=} \langle u_i, v_k \rangle - \langle u_i, v_k \rangle = 0.$$

Außerdem folgt aus (12.2)

$$\langle u_1, \ldots, u_m, w_k \rangle = \langle u_1, \ldots, u_m, v_k \rangle \underset{(12.4)}{=} \langle v_1, \ldots, v_k \rangle.$$

Falls $w_k = 0$, so folgt $\langle v_1, \ldots, v_k \rangle = \langle u_1, \ldots, u_m \rangle$. Falls $w_k \neq 0$, so wird $\{u_1, \ldots, u_{m+1}\}$ ein Orthonormalsystem und ein Erzeugendensystem von $\langle v_1, \ldots, v_k \rangle$, also nach Satz 12.13 eine Orthonormalbasis. □

Beispiel 12.16. Wir wollen Algorithmus 12.14 auf

$$V := \left\langle \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right\rangle \subseteq \mathbb{R}^3$$

anwenden. Wir erhalten

$$w_1 = v_1 = \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix} \quad \text{und} \quad u_1 = \frac{1}{||w_1||} \cdot w_1 = \begin{pmatrix} 3/5 \\ 0 \\ 4/5 \end{pmatrix}.$$

Im zweiten Schritt erhalten wir

$$w_2 = v_2 - \langle u_1, v_2 \rangle \cdot u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{3}{5} \cdot \begin{pmatrix} 3/5 \\ 0 \\ 4/5 \end{pmatrix} = \frac{1}{25} \cdot \begin{pmatrix} 16 \\ 0 \\ -12 \end{pmatrix}$$

und

$$u_2 = \frac{1}{||w_2||} \cdot w_2 = \begin{pmatrix} 4/5 \\ 0 \\ -3/5 \end{pmatrix}.$$

Der dritte Schritt liefert

$$w_3 = v_3 - \langle u_1, v_3\rangle \cdot u_1 - \langle u_2, v_3\rangle \cdot u_2 = \begin{pmatrix}1\\0\\2\end{pmatrix} - \frac{11}{5}\cdot\begin{pmatrix}3/5\\0\\4/5\end{pmatrix} + \frac{2}{5}\cdot\begin{pmatrix}4/5\\0\\-3/5\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Also ist $\{u_1, u_2\}$ eine Orthonormalbasis von V. ◁

Wenn man das Schmidtsche Orthogonalisierungsverfahren auf eine Basis $B = \{v_1, \ldots, v_k\}$ von V anwendet, bekommt man eine Orthonormalbasis $B' = \{u_1, \ldots, u_k\}$. Es ist interessant, dass die Basiswechselmatrix $S_{B,B'}$ automatisch eine obere Dreiecksmatrix wird. Dies folgt aus (12.4).

Aus der Korrektheit von Algorithmus 12.14 folgt:

Korollar 12.17. *Jeder endlich-dimensionale euklidische oder unitäre Vektorraum hat eine Orthonormalbasis.*

Zwischen euklidischen bzw. unitären Vektorräumen kann man „strukturerhaltende“ Abbildungen studieren.

Definition 12.18. *Es seien V und W zwei euklidische bzw. zwei unitäre Räume. Eine lineare Abbildung $\varphi\colon V \to W$ heißt* **orthogonal** *bzw.* **unitär***, falls für alle $v, w \in V$ gilt:*

$$\langle \varphi(v), \varphi(w)\rangle = \langle v, w\rangle.$$

Eine unitäre oder orthogonale Abbildung φ ist injektiv, denn aus $\varphi(v) = 0$ für $v \in V$ folgt $\langle v, v\rangle = \langle \varphi(v), \varphi(v)\rangle = 0$, also $v = 0$. Weiter gilt

$$||\varphi(v)|| = ||v||$$

für alle $v \in V$ und damit auch

$$d(\varphi(v), \varphi(w)) = d(v, w)$$

für $v, w \in V$, φ ist also „abstandserhaltend“. Abbildungen zwischen metrischen Räumen mit dieser Eigenschaft nennt man auch *Isometrien.* Es ist nicht schwer zu zeigen, dass jede lineare Isometrie zwischen euklidischen oder unitären Räumen eine orthogonale bzw. unitäre Abbildung ist. Orthogonale Abbildungen sind auch „winkelerhaltend“.

Beispiel 12.19. (1) Jede Drehung um den Nullpunkt definiert eine orthogonale Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$.

(2) Auf dem Raum $V = C([a, b], \mathbb{C})$ der stetigen Funktionen eines Intervalls $[a, b]$ in $\mathbb{C}$ wird durch $\varphi\colon V \to V,\ f \mapsto \hat{f}$ mit $\hat{f}(x) = f(a + b - x)$ eine unitäre Abbildung gegeben. ◁

Was sind die orthogonalen bzw. unitären Abbildungen $V \to V$ für $V = K^n$ mit $K = \mathbb{R}$ bzw. $K = \mathbb{C}$? Ist φ eine solche, so muss φ jede Orthonormalbasis wieder auf eine Orthonormalbasis abbilden. Ist $A \in K^{n \times n}$ die Darstellungsmatrix von φ bezüglich der Standardbasis (also $\varphi = \varphi_A$), so folgt, dass die Spalten von A eine Orthonormalbasis von V bilden. Dies kann man ausdrücken durch die Bedingungen

$$A^T \cdot A = I_n \quad (\text{für } K = \mathbb{R}) \tag{12.5}$$

bzw.

$$\overline{A}^T \cdot A = I_n \quad (\text{für } K = \mathbb{C}), \tag{12.6}$$

wobei $\overline{A}$ durch komplexe Konjugation aller Einträge aus A hervorgeht. (Die zweite Bedingung umfasst eigentlich die erste, da $\overline{A} = A$ für $K = \mathbb{R}$.) Ist umgekehrt $A \in K^{n \times n}$ eine Matrix, die (12.5) bzw. (12.6) erfüllt, so folgt für $v, w \in V$

$$\langle \varphi_A(v), \varphi_A(w) \rangle = (\overline{Av})^T \cdot (Aw) = \overline{v}^T \overline{A}^T Aw = \langle v, w \rangle.$$

Dies bedeutet, dass genau die Matrizen mit (12.5) bzw. (12.6) orthogonale bzw. unitäre Abbildungen $V \to V$ definieren. Wir nehmen dies zum Anlass für die folgende Definition.

Definition 12.20. *(a) Eine Matrix $A \in \mathbb{R}^{n \times n}$ heißt* **orthogonal**, *falls sie (12.5) erfüllt. Dies ist gleichbedeutend damit, dass die Spalten von A eine Orthonormalbasis von $\mathbb{R}^n$ bilden, und wegen $A \cdot A^T = I_n$ auch damit, dass die Zeilen von A eine Orthonormalbasis von $\mathbb{R}^n$ bilden.*

(b) Eine Matrix $A \in \mathbb{C}^{n \times n}$ heißt **unitär**, *falls sie (12.6) erfüllt. Dies ist gleichbedeutend damit, dass die Spalten von A eine Orthonormalbasis von $\mathbb{C}^n$ bilden, und wegen $A \cdot \overline{A}^T = I_n$ auch damit, dass die Zeilen von A eine Orthonormalbasis von $\mathbb{C}^n$ bilden.*

(c) Die Untergruppe

$$\mathrm{O}_n := \left\{ A \in \mathbb{R}^{n \times n} \mid A^T \cdot A = I_n \right\} \subseteq \mathrm{GL}_n(\mathbb{R})$$

heißt die **orthogonale Gruppe**, *und*

$$\mathrm{SO}_n := \mathrm{O}_n \cap \mathrm{SL}_n(\mathbb{R})$$

heißt die **spezielle orthogonale Gruppe**.

(d) Die Untergruppe

$$\mathrm{U}_n := \left\{ A \in \mathbb{C}^{n \times n} \mid \overline{A}^T \cdot A = I_n \right\} \subseteq \mathrm{GL}_n(\mathbb{C})$$

heißt die **unitäre Gruppe**, *und*

$$\mathrm{SU}_n := \mathrm{U}_n \cap \mathrm{SL}_n(\mathbb{C})$$

heißt die **spezielle unitäre Gruppe**.

Kapitel 13
Hauptachsentransformation

Ziel dieses Kapitels ist der Nachweis, dass jede symmetrische, reelle Matrix diagonalisierbar ist. Durchweg sei $A \in \mathbb{R}^{n \times n}$ symmetrisch, also

$$A^T = A.$$

Wir benötigen drei Lemmata.

Lemma 13.1. *Es sei $\lambda \in \mathbb{C}$ ein Eigenwert von A (aufgefasst als Matrix in $\mathbb{C}^{n \times n}$). Dann gilt $\lambda \in \mathbb{R}$.*

Beweis. Wir haben einen Eigenvektor $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}^n$. Wir schreiben $\overline{v} = \begin{pmatrix} \overline{x_1} \\ \vdots \\ \overline{x_n} \end{pmatrix}$ für den komplex-konjugierten Vektor. Da A reelle Einträge hat, gilt

$$A \cdot \overline{v} = \overline{A \cdot v} = \overline{\lambda v} = \overline{\lambda}\overline{v},$$

also

$$\overline{\lambda} \cdot \overline{v}^T \cdot v = (A \cdot \overline{v})^T \cdot v = \overline{v}^T \cdot A^T \cdot v = \overline{v}^T \cdot A \cdot v = \overline{v}^T \cdot \lambda \cdot v = \lambda \cdot \overline{v}^T \cdot v.$$

Wegen

$$\overline{v}^T \cdot v = \langle v, v \rangle \neq 0$$

folgt $\overline{\lambda} = \lambda$, also $\lambda \in \mathbb{R}$. □

Lemma 13.2. *Es sei $U \subseteq \mathbb{R}^n$ ein Unterraum mit $U \neq \{0\}$, so dass für alle $u \in U$ gilt: $A \cdot u \in U$. Dann enthält U einen Eigenvektor von A.*

Beweis. Nach Voraussetzung haben wir eine lineare Abbildung

$$\varphi\colon U \to U,\ u \mapsto A \cdot u.$$

Bezüglich einer Basis $B = \{b_1, \dots, b_k\}$ von U sei

$$C = (c_{i,j}) = D_B(\varphi) \in \mathbb{R}^{k \times k} \tag{13.1}$$

die Darstellungsmatrix. Es sei $\lambda \in \mathbb{C}$ eine Nullstelle des charakteristischen Polynoms χ_C, also ein Eigenwert von C, aufgefasst als Matrix in $\mathbb{C}^{k \times k}$. Mit einem zugehörigen Eigenvektor $\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$ gilt also

$$C \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}. \tag{13.2}$$

Der Vektor $v := \sum_{i=1}^k x_i b_i \in \mathbb{C}^n$ ist $\neq 0$, da nicht alle x_i Null sind und die b_i auch als Vektoren in $\mathbb{C}^n$ linear unabhängig sind. Es gilt

$$A \cdot v = \sum_{i=1}^k x_i A b_i = \sum_{i=1}^k x_i \varphi(b_i) \underset{(13.1)}{=} \sum_{i=1}^k x_i \left(\sum_{j=1}^k c_{j,i} b_j \right) = \sum_{j=1}^k \left(\sum_{i=1}^k c_{j,i} x_i \right) b_j \underset{(13.2)}{=} \sum_{j=1}^k \lambda x_j b_j = \lambda v,$$

also ist λ ein Eigenwert von A mit Eigenvektor v. Wegen Lemma 13.1 folgt $\lambda \in \mathbb{R}$. Als Eigenwert von C hat λ also auch einen reellen Eigenvektor, d.h. wir können $x_i \in \mathbb{R}$ annehmen. Damit ist $v \in U$ der gesuchte Eigenvektor. $\square$

Ab jetzt sei $\langle \cdot, \cdot \rangle$ das Standard-Skalarprodukt auf $\mathbb{R}^n$.

Lemma 13.3. *Es seien λ und μ zwei verschiedene Eigenwerte von A. Dann gilt für alle $v \in E_\lambda$ und $w \in E_\mu$*

$$\langle v, w \rangle = 0.$$

Beweis. Wir haben

$$(\lambda - \mu) v^T w = (\lambda v^T) w - v^T (\mu w) = (Av)^T w - v^T (Aw) = v^T A^T w - v^T A w = 0.$$

Wegen $\lambda - \mu \neq 0$ folgt $\langle v, w \rangle = v^T w = 0$. $\square$

Nun können wir das Hauptresultat des Kapitels beweisen.

Satz 13.4 („Hauptachsentransformation"). *Für jede symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ gibt es eine Orthonormalbasis von $\mathbb{R}^n$, die aus Eigenvektoren von A besteht.*

Anders gesagt: Es gibt $S \in \mathrm{O}_n(\mathbb{R})$, so dass $S^{-1} A S (= S^T A S)$ eine Diagonalmatrix ist.

Insbesondere ist A diagonalisierbar.

Beweis. Es seien $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ die (verschiedenen) Eigenwerte von A. Für jedes i existiert wegen Korollar 12.17 eine Orthonormalbasis B_i von E_{λ_i}. Wir setzen

$$B := B_1 \cup \dots \cup B_r.$$

Wegen Lemma 13.3 ist B ein Orthonormalsystem, wegen Satz 12.13 also eine Orthonormalbasis von $U := \langle B \rangle$. Nach Konstruktion besteht B aus Eigenvektoren.

Es sei $w \in U^\perp$. Wir behaupten, dass dann auch $A \cdot w$ in $U^\perp$ liegt. Um dies nachzuweisen genügt es zu zeigen, dass $A \cdot w$ auf allen $v \in B$ senkrecht steht. Es sei $v \in B_i$ mit $i \in \{1, \dots, r\}$. Dann folgt

$$\langle v, Aw \rangle = v^T A w = v^T A^T w = (Av)^T w = (\lambda_i v)^T w = \lambda_i \langle v, w \rangle = 0,$$

wobei wir im letzten Schritt $w \in U^\perp$ verwendet haben. Falls $U^\perp \neq \{0\}$, so würde U also die Voraussetzungen von Lemma 13.2 erfüllen, und wir erhielten einen Eigenvektor $v \in U^\perp$ von A. Als Eigenvektor von A liegt v in U, also $\langle v, v \rangle = 0$. Dies ist ein Widerspruch zu $v \neq 0$. Wir schließen, dass $U^\perp = 0$ gilt.

Mit $B = \{b_1, \dots, b_k\}$ und $C := \begin{pmatrix} b_1^T \\ \vdots \\ b_k^T \end{pmatrix} \in \mathbb{R}^{k \times n}$ (die Matrix mit den Vektoren aus B als Zeilen) ist $U^\perp$ der Lösungsraum des homogenen LGS $C \cdot x = 0$. Aus $U^\perp = \{0\}$ folgt, dass das LGS mindestens n Gleichungen enthält. Also ist B eine linear unabhängige Teilmenge von $\mathbb{R}^n$ mit mindestens n (also genau n) Vektoren. Damit ist B eine Basis von $\mathbb{R}^n$. Dass B ein Orthonormalsystem aus Eigenvektoren von A ist, haben wir bereits gesehen. □

Beispiel 13.5. (1) Wir betrachten die symmetrische Matrix

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \in \mathbb{R}^{3 \times 3}.$$

Um A zu diagonalisieren, berechnen wir das charakteristische Polynom und erhalten

$$\chi_A = \det \begin{pmatrix} x-2 & -1 & -1 \\ -1 & x-2 & -1 \\ -1 & -1 & x-2 \end{pmatrix} = (x-2)^3 - 2 - 3(x-2) =$$
$$x^3 - 6x^2 + 9x - 4 = (x-1)(x^2 - 5x + 4) = (x-1)^2(x-4).$$

Damit wissen wir schon, dass A zu $\mathrm{diag}(1, 1, 4)$ ähnlich ist. Wir wollen eine orthogonale Transformationsmatrix ausrechnen. Hierfür müssen wir die

Eigenräume bestimmen. Der Eigenraum E_1 zum Eigenwert 1 ergibt sich als Lösungsraums des homogenen LGS mit Matrix $A - I_3$. Wir erhalten

$$E_1 = \left\langle \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\rangle .$$

Auf die Basis von E_1 wenden wir das Schmidtsche Orthogonalisierungsverfahren an. Der erste Schritt liefert

$$u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} .$$

Weiter erhalten wir

$$w_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{1}{\sqrt{2}} u_1 = \frac{1}{2} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} ,$$

also

$$u_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} .$$

Nun berechnen wir E_4 und erhalten durch Lösen des entsprechenden LGS (oder durch die Beobachtung, dass alle Zeilensummen von A gleich 4 sind)

$$E_4 = \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle .$$

Normieren liefert als letzten Vektor der Orthonormalbasis

$$u_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

Damit gilt

$$S = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \in O_3(\mathbb{R})$$

und

$$S^{-1} A S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} .$$

(2) Es stellt sich die Frage, ob Satz 13.4 auch über anderen Körpern außer $\mathbb{R}$ gilt, z.B. über $\mathbb{C}$. Um diese zu beantworten, betrachten wir die symmetrische Matrix

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix} \in \mathbb{C}^{2\times 2}.$$

Das charakteristische Polynom ist

$$\chi_A = \det \begin{pmatrix} x-1 & -i \\ -i & x+1 \end{pmatrix} = (x-1)(x+1)+1 = x^2,$$

also haben wir 0 als einzigen Eigenwert. Die algebraische Vielfachheit ist 2, die geometrische aber 1, also ist A nicht diagonalisierbar. Mit $\mathbb{C}$ statt $\mathbb{R}$ wäre Satz 13.4 also nicht korrekt. Ebenso verhält es sich mit $\mathbb{Q}$ statt $\mathbb{R}$. ◁

Als physikalische Anwendung von Satz 13.4 betrachten wir nun den *Trägheitstensor*. Dieser setzt die Winkelgeschwindigkeit und den Drehimpuls eines rotierenden starren Körpers in Beziehung. Wir leiten diese Beziehung her. Eine Rotation eines starren Körpers lässt sich durch die Winkelgeschwindigkeit $\omega \in \mathbb{R}^3$ folgendermaßen beschreiben: Ein zum Körper gehöriger Massepunkt, der sich am Ort $r \in \mathbb{R}^3$ befindet, bewegt sich mit dem Geschwindigkeitsvektor

$$v = \omega \times r.$$

Dabei ist das **Vektorprodukt** (auch genannt *Kreuzprodukt*) zweier Vektoren in $\mathbb{R}^3$ definiert durch

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.$$

(Dass diese Definition an die Sarrus-Regel erinnert, ist kein Zufall.) Der Drehimpuls kann dadurch definiert werden, dass ein zu dem starren Körper gehörender Massepunkt mit Masse m am Ort $r \in \mathbb{R}^3$ zum Drehimpuls den Beitrag $m(r \times v)$ leistet. Mit einer bekannten, leicht nachzurechnenden Formel für das Vektorprodukt ist dieser Beitrag gleich

$$m\,(r \times (\omega \times r)) = m(||r||^2\omega - rr^T\omega).$$

Indem wir Massepunkte durch eine kontinuierliche Masseverteilung mit Dichte $\rho(r)$ ersetzen, erhalten wir folgende Gleichung für den Drehimpuls L:

$$L = \int_{r=(x,y,z)\in\mathbb{R}^3} \rho(r)(||r||^2\omega - rr^T\omega)dxdydz = I \cdot \omega$$

mit

$$I = \int_{r=(x,y,z)\in\mathbb{R}^3} \rho(r)(||r||^2 I_3 - rr^T) dxdydz =$$
$$\int_{x,y,z} \rho(x,y,z) \begin{pmatrix} y^2+z^2 & -xy & -xz \\ -xy & x^2+z^2 & -yz \\ -xz & -yz & x^2+y^2 \end{pmatrix} dxdydz \in \mathbb{R}^{3\times 3},$$

wobei I_3 für die Einheitsmatrix steht. Somit ist eine Formel für den Trägheitstensor I gefunden, der nichts anderes als eine reelle 3×3-Matrix ist. Eine wichtige Beobachtung ist nun, dass I symmetrisch ist. Also liefert Satz 13.4, dass es für jeden starren Körper drei senkrecht zueinander stehende Achsen gibt, so dass bei einer Drehung um diese Achsen die Drehgeschwindigkeit und der Drehimpuls in dieselbe Richtung zeigen. Diese Achsen heißen *Hauptträgheitsachsen*. Der Titel dieses Abschnitts ist durch diese Benennung motiviert. Bei unserer Herleitung haben wir den Körper zu einem bestimmten Zeitpunkt betrachtet. Da er jedoch rotiert, rotieren seine Hauptträgheitsachsen mit ihm. Rotiert der Körper um einer der Hauptträgheitsachsen, so stört dies nicht: ω und L können konstant in dieselbe Richtung zeigen. Andernfalls wird die Bewegung aber kompliziert, da bei konstantem L die Winkelgeschwindigkeit ω selbst rotiert (*Nutation*).

Aber auch die Rotationen um Hauptträgheitsachsen sind nicht gleichberechtigt. Um das zu sehen, betrachten wir die kinetische Energie der Rotation. Die kinetische Energie eines Massepunktes ist

$$\frac{m}{2}||v||^2 = \frac{m}{2}||\omega \times r||^2 = \frac{m}{2}\left(||\omega||^2||r||^2 - \langle \omega, r\rangle^2\right),$$

wobei man die letzte Gleichung leicht nachrechnen kann. Andererseits gilt

$$\omega^T I \omega = \int_{r\in\mathbb{R}^3} \rho(r)\left(||\omega||^2||r||^2 - \langle \omega, r\rangle^2\right) dxdydz, \tag{13.3}$$

also ergibt sich die Energie zu

$$E = \frac{1}{2}\omega^T I \omega.$$

Vergleich mit (12.1) zeigt, dass sich die Energie durch Einsetzen von ω in beide Argumente einer symmetrischen Bilinearform ergibt. Wir fragen nun, bei welcher Rotation sich die niedrigste Energie ergibt, wenn man den Betrag des Drehimpulses $||L||$ fest lässt, was aufgrund des Drehimpulserhaltungssatzes physikalisch sinnvoll ist. Es seien $\lambda_1 \leq \lambda_2 \leq \lambda_3$ die Eigenwerte von I, und wir wählen unser Koordinatensystem entlang der Hauptträgheitsachsen. Wir werden sehen (siehe Beispiel 13.8(2)), dass alle λ_i positiv sind, falls nicht alle Masse des starren Körpers auf einer Geraden konzentriert ist, was wir nun annehmen. Mit $L = (L_1, L_2, L_3)^T$ und $||L|| = 1$ erhalten wir

$$2E = L^T I^{-1} L = \sum_{i=1}^{3} \lambda_i^{-1} L_i^2 = (\lambda_1^{-1} - \lambda_3^{-1}) L_1^2 + (\lambda_2^{-1} - \lambda_3^{-1}) L_2^2 + \lambda_3^{-1},$$

was durch $L_1 = L_2 = 0$ minimiert wird. Die minimale Energie ergibt sich also bei Rotation um die Hauptträgheitsachse, die zu dem *größten* Eigenwert des Trägheitstensors gehört. Dies erklärt, weshalb es für einen Kreisel energetisch günstig ist, aufrecht zu bleiben.

Anmerkung. Satz 13.4 ist ein Spezialfall des sogenannten *Spektralsatzes*: Für $B \in \mathbb{C}^{n \times n}$ mit

$$\overline{B}^T B = B \overline{B}^T$$

(solche Matrizen nennt man *normal*) gibt es eine *unitäre* Matrix $S \in \mathbb{C}^{n \times n}$ (d.h. $\overline{S}^T \cdot S = I_n$), so dass $S^{-1} B S$ eine Diagonalmatrix ist. ◁

Aufgrund von Satz 13.4 wird folgende Definition sinnvoll.

Definition 13.6. *Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ heißt*

- **positiv definit***, falls alle Eigenwerte von A positiv sind;*
- **positiv semidefinit***, falls alle Eigenwerte von A positiv oder Null sind;*
- **negativ definit***, falls alle Eigenwerte von A negativ sind;*
- **negativ semidefinit***, falls alle Eigenwerte von A negativ oder Null sind;*
- **indefinit***, falls es sowohl positive als auch negative Eigenwerte gibt.*

Satz 13.7. *Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ ist genau dann positiv definit, wenn für alle $v \in \mathbb{R}^n \setminus \{0\}$ gilt:*

$$\langle v, A \cdot v \rangle > 0.$$

A ist positiv semidefinit, wenn $\langle v, A \cdot v \rangle \geq 0$ gilt. Entsprechendes gilt für negativ (semi-)definit.

Beweis. Wegen Satz 13.4 haben wir $S \in \mathrm{O}_n(\mathbb{R})$, so dass

$$S^T A S = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D$$

gilt, wobei die λ_i die Eigenwerte von A sind. Wegen $S^T \in \mathrm{GL}_n(\mathbb{R})$ ist für jeden Vektor $v \in \mathbb{R}^n \setminus \{0\}$ auch $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := S^T \cdot v$ ungleich 0, und jeder Vektor aus $\mathbb{R}^n \setminus \{0\}$ tritt als ein $S^T \cdot v$ mit $v \in \mathbb{R}^n \setminus \{0\}$ auf. Es gilt

$$\langle v, A \cdot v \rangle = v^T S D S^T v = (x_1, \dots, x_n) D \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \lambda_i x_i^2.$$

Hieraus folgen alle Behauptungen. □

Beispiel 13.8. (1) Wir betrachten

$$A = \begin{pmatrix} a & 0 & -a & 0 \\ 0 & b & 0 & -b \\ -a & 0 & a & 0 \\ 0 & -b & 0 & b \end{pmatrix} \quad \text{mit} \quad a, b \in \mathbb{R}.$$

Wir wenden Satz 13.7 zur Feststellung der Definitheitseigenschaften von A an. Für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ gilt

$$\langle v, A \cdot v \rangle = (x_1, x_2, x_3, x_4) \cdot \begin{pmatrix} a(x_1 - x_3) \\ b(x_2 - x_4) \\ -a(x_1 - x_3) \\ -b(x_2 - x_4) \end{pmatrix} = a(x_1 - x_3)^2 + b(x_2 - x_4)^2.$$

Damit ist A positiv semidefinit, falls $a, b \geq 0$, negativ semidefinit, falls $a, b \leq 0$, und sonst indefinit.

(2) Ist $I \in \mathbb{R}^{3 \times 3}$ der Trägheitstensor eines starren Körpers, so ergibt sich aus (13.3) und der Schwarzschen Ungleichung (Proposition 12.6), dass I positiv semidefinit ist. Weiter ergibt sich, dass I positiv definit ist, falls nicht die gesamte Masse des Körpers auf einer Geraden konzentriert ist. ◁

Notation

Index