Gregor Kemper

Lineare Algebra für Informatik*

Vorlesungsmanuskript†
Technische Universität München

21. Oktober 2024

*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

In diesem Kapitel ist K immer ein Körper (z.B. $K = \mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{F}_2 \dots$).

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 definieren wir eine $m \times n$-Matrix 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, wobei das Bild von (i, j) mit $a_{i,j}$ bezeichnet wird. Somit wird die informelle Definition einer Matrix als rechteckige Anordnung „degradiert" zu einer bloßen Schreibweise.

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) $P_1,\dots,P_n$ seien (alle oder einige) Seiten des Internets. Für $i,j \in \{1,\dots,n\}$ definieren wir

$$w_{i,j} = \begin{cases} 1 & \text{falls } P_i \text{ einen Link auf } P_j \text{ enthält} \\ 0 & \text{sonst} \end{cases}.$$

Die Matrix $W := (w_{i,j}) \in \mathbb{R}^{n\times n}$ heißt *Weblink-Matrix*. Sie codiert die „referenzielle Struktur“ des Internets. Mit der Kenntnis von W kann man Fragen beantworten wie: Kann man sich von P_i nach P_j durchklicken? Wieviele Klicks braucht man? W ist quadratisch aber nicht symmetrisch. Wir betrachten ein Beispiel von drei Internet-Seiten:

$$\begin{tikzcd} \boxed{P_1} \arrow[r, shift left] \arrow[dr] & \boxed{P_2} \arrow[l, shift left] \\ & \boxed{P_3} \end{tikzcd} \tag{1.1}$$

Die Pfeile kennzeichnen Links. Als Weblink-Matrix ergibt sich

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

Die Erfolgsgeschichte von Google basiert (unter anderem) auf folgender Idee: Die Wichtigkeit einer Seite P_i ergibt sich daraus, wie viele andere wichtige Seiten einen Link auf P_i enthalten. Hierbei ist das Gewicht eines Links durch die Gesamtzahl der von derselben Seite ausgehenden Links zu dividieren. Wie lässt sich dieser selbst-refenzierende Wichtigkeitsbegriff entwirren? Wir werden hierauf im Kapitel 12 zurückkommen. Es ist jetzt schon klar, dass die Wichtigkeit der Seiten durch die Weblink-Matrix bestimmt wird.

(4) $Z_1, \dots, Z_n$ 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 $P_{i,1} + \cdots + P_{i,n} = 1$, 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(1) nach der Definition des Matrixprodukts klären.

(5) Als konkretes Beispiel für eine Transitionsmatrix gehen wir zurück zum Internet und stellen uns einen *Zufallssurfer* vor, der von jeder Seite aus rein zufällig irgendeinem Link folgt. Falls die Seite keinen Link enthält, wählt der Zufallssurfer eine beliebige Seite zufällig aus. Das System ist also der Surfer, und der Zustand ist die Seite, auf der er sich gerade befindet. Im oben betrachteten Miniatur-Beispiel (1.1) des Internets ergibt sich für den Zufallssurfer die Transitionsmatrix

$$T = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Wie hoch sind die Wahrscheinlichkeiten, nach 100 Klicks von der Seite P_i nach P_j zu gelangen? Im Vorgriff auf die Methode, wie man dies systematisch ausrechnet, sei die Transitionsmatrix T_{100} hier (näherungsweise) angegeben:

$$T_{100} \approx \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}.$$

Es fällt zunächst auf, dass alle Zeilen (näherungsweise) gleich sind. Die Interpretation ist naheliegend: Nach vielen Klicks ist die Wahrscheinlichkeit, auf einer bestimmten Seite zu landen, nahezu unabhängig von der Ausgangsseite. In Kapitel 12 werden wir die asymptotische Gleichheit der Zeilen nachweisen. Ein Blick auf T_{100} legt außerdem einen alternativen Wichtigkeitsbegriff für Intenet-Seiten nahe: die Wichtigkeit w_i einer Seite P_i ergibt sich als die Wahrscheinlichkeit, dass der Zufallssurfer nach einer hohen Zahl von Klicks auf P_i landet. Im Beispiel erhalten wir also $w_1 = 0.4$ und $w_2 = w_3 = 0.3$. Um beide Wichtigkeitsbegriffe zu vergleichen, wollen wir prüfen, ob die w_i dem Google-Wichtigkeitsbegriff aus (3) genügen. Gemäß des Verhaltens unseres Zufallssurfers ergänzen wir (1.1) um Links von P_3 auf alle Seiten, einschließlich sich selbst. Auf P_1 zeigt ein Link mit Gewicht $\frac{1}{3}$ von P_3 (da 3 Links von P_3 ausgehen) und ein Link mit Gewicht 1 von P_2. Also müsste P_1 die Wichtigkeit

$$w_2 + \frac{1}{3} \cdot w_3 = 0.3 + \frac{1}{3} \cdot 0.3 = 0.4$$

haben, und genau so ist es gemäß unseres alternativen Wichtigkeitsbegriffs! Weiterhin müsste P_2 die Wichtigkeit

$$\frac{1}{2} \cdot w_1 + \frac{1}{3} \cdot w_3 = \frac{1}{2} \cdot 0.4 + \frac{1}{3} \cdot 0.3 = 0.3$$

und P_3 die Wichtigkeit

$$\frac{1}{2} \cdot w_1 + \frac{1}{3} \cdot w_3 = \frac{1}{2} \cdot 0.4 + \frac{1}{3} \cdot 0.3 = 0.3$$

haben, was ebenfalls mit dem alternativen Begriff zusammenfällt. Diese Beobachtungen sind überraschend, da unsere beiden Begriffe von Wichtigkeit scheinbar nichts miteinander zu tun haben. Ist die Übereinstimung ein Zufall, der nur in diesem Beispiel eintritt? In Kapitel 12 werden wir beweisen, dass dies nicht der Fall ist: Beide Begriffe von Wichtigkeit decken sich immer! Für den Nachweis werden wir fast sämtliche bis dahin entwickelte Theorie brauchen. ◁

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. Es ist klar, dass für $A, B, C \in K^{m\times n}$ die Regeln*

$$(A + B) + C = A + (B + C) \quad \textit{und} \quad A + B = B + A$$

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

(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 $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$:*

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

(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

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

(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 13 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 8 kennenlernen. Ein paar Hinweise, weshalb die Definition sinnvoll ist, finden sich jedoch schon in folgendem Beispiel.

Beispiel 1.5. (1) Wie in Beispiel 1.2(4) 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. Die Matrix T_{100} in Beispiel 1.2(5) wurde als $T_{100} := T^{100}$ berechnet. Hier findet das Produkt bzw. die Potenzen von Transitionsmatrizen also eine natürliche Interpretation.

(2) Ein Beispiel aus der Wirtschaft soll hier nur angedeutet werden. Wir nehmen an, dass das Unternehmen U_i einen gewissen Anteil $a_{i,j}$ des Unternehmens U_j besitzt. Die $a_{i,j}$ kann man in einer quadratischen Matrix A zusammenfassen, die die wirtschaftlichen Verflechtungen beschreibt.

Ähnlich wie im Beispiel (1) sieht man, dass A^2 dann die Anteile „zweiter Generation“ beschreibt, und so weiter. ◁

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

(a) $(K^{m\times n}, +)$ *ist eine abelsche Gruppe.*
(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) 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. Die Richtigkeit von (a) wurde schon in Definition 1.3(a) angemerkt. 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

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \text{aber} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \tag{1.2}$$

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

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

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. Dieser Eintrag wird als **Pivotelement** *bezeichnet.*

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

(c) Über jedem Pivotelement, also 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. $\triangleleft$

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.

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

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

$$
\left(\begin{array}{cccc|c} 1 & 0 & 2 & 1 & -3 \\ 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)\begin{array}{l} {\scriptstyle -1} \\ {\scriptstyle 1} \\ \\ \\ \end{array} \xrightarrow[\text{Typ III}]{} \left(\begin{array}{cccc|c} 1 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)
$$

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.) Dieser Eintrag wird in den folgenden beiden Schritten als Pivotelement verwendet.
(5) Bringe das Pivotelement in die $(r+1)$-te Zeile (Operation Typ I).
(6) Erzeuge unterhalb des Pivotelements 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. Dies ist auch die Anzahl der Pivotelemente. Für $i = 1, \ldots, r$ sei $j_i \in \{1, \ldots, n+1\}$ die Position (= Spalte), in der das Pivotelement 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 $k_1, \dots, k_{n-r}$ diejenigen Zahlen in $\{1, \dots, n\}$, die nicht eines der j_i sind. Also $\{1, \dots, n\} \setminus \{j_1, \dots, j_r\} = \{k_1, \dots, k_{n-r}\}$.

(5) Die Lösungsmenge ist

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

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

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

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
Komplexe Zahlen

Dieses kurze Kapitel ist ein Einschub über komplexe Zahlen. Die Motivation zur Einführung der komplexen Zahlen ist, dass negative Zahlen, insbesondere -1, in $\mathbb{R}$ keine Quadratwurzel haben. Um Abhilfe zu schaffen, kann man sich ein Wurzel aus -1 als „imaginäre Zahl" denken. Man muss aber gar nicht so philosophisch werden. Wenn man sich die reellen Zahlen als die skalaren Matrizen $\operatorname{diag}(a, a)$ in $\mathbb{R}^{2\times 2}$ eingebettet denkt, dann gibt es bereits eine Wurzel aus -1, denn $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. Dies gibt die Idee, die komplexen Zahlen durch

$$\mathbb{C} := \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a, b \in \mathbb{R} \right\} = \{a \cdot I_2 + b \cdot i \mid a, b \in \mathbb{R}\} \quad \text{mit} \quad i := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

zu definieren. Weil die so definierte Menge unter Addition abgeschlossen ist, ist $(\mathbb{C}, +)$ eine Untergruppe der abelschen Gruppe $(\mathbb{R}^{2\times 2}, +)$ (siehe Satz 1.6). Weiter gilt für $z_1 = a_1 I_2 + b_1 i$ und $z_2 = a_2 I_2 + b_2 i \in \mathbb{C}$:

$$z_1 \cdot z_2 = a_1 a_2 I_2 + (a_1 b_2 + a_2 b_1) i + b_1 b_2 i^2 = (a_1 a_2 - b_1 b_2) I_2 + (a_1 b_2 + a_2 b_1) i \in \mathbb{C}.$$

An dieser Formel sieht man auch, dass das Kommutativgesetz

$$z_1 z_2 = z_2 z_1$$

gilt. $\mathbb{C}$ hat I_2 als Einselement, und das Assoziativ- und Distributivgesetz vererbt sich von $\mathbb{R}^{2\times 2}$ auf $\mathbb{C}$. Damit ist $\mathbb{C}$ ein kommutativer Ring. Für $z = aI_2 + bi \neq 0$ gilt $a^2 + b^2 > 0$. Wir können direkt nachprüfen, dass für

$$z^{-1} := \frac{1}{a^2 + b^2} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \frac{a}{a^2 + b^2} I_2 - \frac{b}{a^2 + b^2} i \in \mathbb{C}$$

die Gleichung $z^{-1} z = I_2$ gilt. Dies zeigt, dass $\mathbb{C}$ sogar ein Körper ist. Durch

$$\varepsilon \colon \mathbb{R} \to \mathbb{C},\ a \mapsto aI_2$$

wird eine injektive Abbildung gegeben, für die die Regeln $\varepsilon(a+b) = \varepsilon(a) + \varepsilon(b)$, $\varepsilon(a \cdot b) = \varepsilon(a) \cdot \varepsilon(b)$ $(a, b \in \mathbb{R})$ und $\varepsilon(1) = I_2$ gelten. Wir werden $a \in \mathbb{R}$ mit $\varepsilon(a) \in \mathbb{C}$ identifizieren und so $\mathbb{R}$ als Unterkörper von $\mathbb{C}$ auffassen. Mit dieser Identifikation gilt

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$$

und

$$i^2 = -1.$$

Die Multiplikations- und Inversionsregel liest sich als

$$(a_1 + b_1 i) \cdot (a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i$$

und

$$(a + bi)^{-1} = \frac{a - bi}{a^2 + b^2} \quad (\text{für} \quad a + bi \neq 0).$$

Damit ist eine „philosophiefreie“ Konstruktion von $\mathbb{C}$ und die Herleitung der wichtigsten (algebraischen) Eigenschaften geschafft! Ein Element von $\mathbb{C}$ heißt eine **komplexe Zahl**. Zu $z = a + bi$ heißen a der **Realteil** und b der **Imaginärteil**, geschrieben als

$$a =: \mathrm{Re}(z), \quad b =: \mathrm{Im}(z).$$

Da jede komplexe Zahl zwei reelle Komponenten hat, lässt sich $\mathbb{C}$ veranschaulichen als „komplexe Zahlenebene“.

Wir haben auf $\mathbb{C}$ die sogenannte **komplexe Konjugation**: Einer komplexen Zahl $z = a + bi$ wird die **komplex konjugierte**

$$\overline{z} := a - bi$$

zugeordnet. Also ist $\overline{z}$ nichts anders als die Transponierte der 2×2-Matrix z. Es folgt, dass für $z_1, z_2 \in \mathbb{C}$ die Regeln

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \quad \text{und} \quad \overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$$

gelten. Auf $\mathbb{C}$ gibt es keine (sinnvolle) Anordnung „$\leq$“. Allerdings können wir einen **Betrag** definieren, indem wir beobachten, dass für $z = a + bi$

$$z \cdot \overline{z} = a^2 + b^2 \in \mathbb{R}_{\geq 0}$$

gilt und dann

$$\mid z \mid := \sqrt{z \cdot \overline{z}} = \sqrt{a^2 + b^2} \in \mathbb{R}_{\geq 0}.$$

setzen. Mit dieser Definition kann man die Regel für die Inversion auch als

$$z^{-1} = \frac{\overline{z}}{|z|^2} \quad \text{für} \quad z \in \mathbb{C} \setminus \{0\}$$

schreiben. Für den Betrag gelten folgende Regeln:

(1) Für $z \in \mathbb{C}$ gilt
$$|z| = 0 \iff z = 0.$$

(2) Für $z_1, z_2 \in \mathbb{C}$ gilt
$$|z_1 z_2| = |z_1||z_2|.$$

(3) Für $z_1, z_2 \in \mathbb{C}$ gilt
$$|z_1 + z_2| \leq |z_1| + |z_2| \quad (\text{„Dreiecksungleichung“}).$$

Hierbei ist (1) klar und (2) folgt aus der Multiplikativität der komplexen Konjugation. Die Dreiecksungleichung beweisen wir durch folgende Rechnung:

$$\begin{aligned}|z_1 + z_2|^2 = (z_1 + z_2)(\overline{z_1} + \overline{z_2}) &= \underbrace{z_1\overline{z_1}}_{=|z_1|^2} + \underbrace{z_1\overline{z_2} + \overline{z_1}z_2}_{=2\operatorname{Re}(z_1\overline{z_2})} + \underbrace{z_2\overline{z_2}}_{=|z_2|^2} \\ &\leq |z_1|^2 + 2\underbrace{|z_1\overline{z_2}|}_{\underset{(2)}{=}|z_1||\overline{z_2}|=|z_1||z_2|} + |z_2|^2 = \left(|z_1| + |z_2|\right)^2.\end{aligned}$$

Aus der Rechnung geht auch hervor, wann Gleichheit gilt:

$$\begin{aligned}|z_1 + z_2| = |z_1| + |z_2| \quad &\Leftrightarrow \quad \operatorname{Re}(z_1\overline{z_2}) = |z_1\overline{z_2}| \quad \Leftrightarrow \\ z_1\overline{z_2} = |z_1||z_2| \quad &\Leftrightarrow \quad |z_1| \cdot z_2 = |z_2| \cdot z_1. \qquad (3.1)\end{aligned}$$

Geometrisch kann man dies folgendermaßen ausdrücken: Die Dreiecksungleichung ist genau dann eine Gleichheit, wenn die beiden Zahlen in der komplexen Zahlenebene in die gleiche Richtung zeigen (oder mindestens eine der beiden Null ist). Dies entspricht auch der geometrischen Intuition.

Durch die Einführung des Betrags gewinnen wir für komplexe Zahlen Begriffe wie Nähe, Konvergenz und Stetigkeit. Hierdurch wird es möglich, analytische Begriffe auf komplexe Funktionen anzuwenden, was der Startpunkt der komplexen Analysis (= Funktionentheorie) ist.

Beispiel 3.1. Sei $z \in \mathbb{C}$ mit $|z| < 1$. Dann konvergiert die Folge z^n gegen 0: $\lim_{n\to\infty} z^n = 0$. ◁

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 abelsche (= 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.*

Wir hätten einen Vektorraum auch formaler als ein Tripel $(V, \boxplus, \boxdot)$ definieren können. Wir verwenden jedoch den etwas laxeren Sprachgebrauch „eine Menge ... zusammen mit Abbildungen ... “.

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) $V = \{0\}$ (abelsche 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) $\mathbb{C}$ 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 abelsche 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. ◁

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) $V = \mathbb{R}^2$. 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 $U_1 = U_2$). $\triangleleft$

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. $\square$

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 $v_1, \dots, v_n \in V$ Vektoren. Ein Vektor $v \in V$ heißt* **Linearkombination** *von $v_1, \dots, v_n$, falls es Skalare $a_1, \dots, a_n \in K$ 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 $v_1, \dots, v_n \in S$ gibt, so dass v eine Linearkombination von $v_1, \dots, v_n$ 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 $S = \{v_1, \dots, v_n\}$ 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 $v_1, \dots, v_n \in V$:

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

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 $v_1, \dots, v_n \in S$ und $a_1, \dots, a_n \in K$ 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. □

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 $S = \{x^i \mid i \in \mathbb{N}_0\} = \{1, x, x^2, \ldots\}$. ◁

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 $v_1, \ldots, v_m$ 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

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

also auch hier keine Änderung. □

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 $v_1, \ldots, v_n \in V$ heißen* **linear unabhängig**, *falls für alle $a_1, \ldots, a_n$ folgende Implikation gilt:*

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

*Gleichbedeutend damit ist: Für jede Linearkombination $v \in \langle v_1, \ldots, v_n\rangle$ gibt es eindeutig bestimmte $a_1, \ldots, a_n \in K$ mit $v = \sum_{i=1}^n a_i v_i$ („*eindeutige Darstellungseigenschaft*“). Der Beweis, dass lineare Unabhängigkeit und die eindeutige Darstellungseigenschaft gleichbedeutend sind, sei den Leserinnen und Lesern überlassen. Die Vektoren $v_1, \ldots, v_n$ 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 $v_1, \ldots, v_n \in S$ gilt, dass $v_1, \ldots, v_n$ 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 Gleicheitsbegriff 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 17 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

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

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. ◁

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) ⇒ (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

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

Hierbei können wir die v_i als paarweise verschieden annehmen. Aus der obigen Gleichung folgt

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

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

Nun zeigen wir „(b) ⇒ (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) ⇒ (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 $a_1, \dots, a_n \in K$ mit

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

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 $E' := \{v, v_2, \dots, v_n\}$ ergibt sich aus (6.1):

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

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*:*

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

Wie kann man eine Basis eines Unterraums $U \subseteq K^n$ finden? Wir nehmen an, U sei durch erzeugende Vektoren $v_1, \dots, v_m$ 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: $\dim(U) = \mathrm{rg}(A)$. 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* $v_1, \dots, v_n \in V$ *paarweise verschieden und* $S = \{v_1, \dots, v_n\}$*. Dann gelten:*

(a) *S ist eine Basis von V* $\iff \dim(V) = n$ *und S ist linear unabhängig* $\iff \dim(V) = n$ *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. □

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. □

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. ◁

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

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 $U = \langle S_{\max} \rangle = \langle B \rangle = V$. □

Kapitel 7
Lineare Codes

In diesem Kapitel werden die bisher erarbeiteten Konzepte auf die Datenübertragung über einen nicht perfekten Kanal angewandt. Wir stellen uns vor, dass nacheinander Bits $x_1, x_2, x_3, \ldots$ über einen Kanal gesendet (oder auf einem Datenträger gespeichert) werden. Hierbei sind Fehler möglich: Mit einer gewissen Wahrscheinlichkeit (etwa $p = 10^{-6}$) wird ein Bit fehlerhaft übertragen bzw. gespeichert. Um trotzdem die korrekten Daten rekonstruieren zu können, oder um zumindest mit großer Wahrscheinlichkeit auf einen Fehler aufmerksam zu werden, schickt man die Daten mit einer gewissen Redundanz.

Die naivste Idee ist hierbei das Wiederholen: Alle Daten werden zweimal gesendet (oder 3,4, $\ldots$ mal). Bei Einteilung in Viererblocks wird also statt (x_1, x_2, x_3, x_4) das „Wort" $(x_1, x_2, x_3, x_4, x_1, x_2, x_3, x_4)$ gesendet.

Als allgemeinen Rahmen wollen wir die folgende Situation betrachten: Ein Bit wird als ein Element des Körpers $K = \mathbb{F}_2 \,(= \mathbb{Z}/2\mathbb{Z})$ modelliert. Wir können jedoch auch Elemente eines anderen (endlichen) Körpers K betrachten. Der zu sendende Bit-Strom wird in Blocks der Länge k zerlegt, z.B. $k = 4$. Statt $(x_1, \ldots, x_k) \in K^k$ wird $(c_1, \ldots, c_n) \in K^n$ gesendet (bzw. gespeichert). Hierbei gibt es eine Zuordnung $(x_1, \ldots, x_k) \mapsto (c_1, \ldots, c_n)$. Diese ist häufig linear, d.h. gegeben durch eine Matrix $G \in K^{n \times k}$, also:

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = G \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

(Man beachte, dass wir hier je nach Bequemlichkeit Zeilen- und Spaltenvektoren schreiben.) Der gesendete Vektor $(c_1, \ldots, c_n)$ heißt **Codewort**, und $(x_1, \ldots, x_k)$ heißt **Informationswort**. G heißt **Generatormatrix**. Die Menge

$$C := \left\{ G \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \;\middle|\; \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \in K^k \right\}$$

aller Codewörter bildet einen Unterraum des K^n. Eine solche Datenübertragung ist nur sinnvoll, wenn die Zuordnung des Codeworts zu einem Datenwort injektiv ist. Das inhomogene LGS $G \cdot x = c$ muss also für alle $c \in C$ eindeutig lösbar sein, also $\mathrm{rg}(G) = k$. Aus unserem Test auf lineare Unabhängigkeit auf Seite 32 folgt, dass die Spalten von G linear unabhängig sind. Diese Spalten erzeugen C, also folgt

$$\dim(C) = k.$$

Ausgehend von dieser Situation machen wir folgende Definition:

Definition 7.1. *Ein* **linearer Code** *ist ein Unterraum $C \subseteq K^n$. Mit $k := \dim(C)$ bezeichnen wir C auch als einen (n, k)-Code. Die* **Länge** *von C ist n. Die* **Informationsrate** *ist k/n, die* **Redundanz** *ist $n - k$.*

Bei der Definition fällt auf, dass die Abbildung $K^k \to K^n$ nicht in die Definition des Codes aufgenommen wird. Für die meisten Fragestellungen der Codierungstheorie ist diese nämlich unerheblich. Als Generatormatrix eines Codes C kann man jede Matrix nehmen, deren Spalten eine Basis von C bilden. Wir bemerken noch, dass bisweilen auch nicht-lineare Codes betrachtet werden.

Beispiel 7.2. (1) Die Generatormatrix

$$G := \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix}$$

liefert den *Wiederholungscode*, bei dem alles einmal wiederholt wird. Dies ist ein (8,4)-Code, die Informationsrate ist also 1/2. Falls bei der Übertragung höchstens ein Fehler auftritt, wird dies beim Empfang festgestellt. Der Fehler kann jedoch nicht korrigiert werden. Man spricht von einem *1-fehlererkennenden Code*.

(2) Der sogenannte *Parity-Check-Code* ist gegeben durch die Generatormatrix

$$G := \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&1&1&1 \end{pmatrix}.$$

Als Abbildung kann man ihn als $(x_1, \ldots, x_4) \mapsto (x_1, \ldots, x_4, x_1+x_2+x_3+x_4)$ definieren. Dies ist ein (5,4)-Code. Falls einer oder 3 Fehler auftreten, wird dies erkannt. Also ist auch dieser Code *1-fehlererkennend*. Aber seine Informationsrate ist mit 4/5 höher als die des Wiederholungscodes. Der Parity-Check-Code ist wohl eine der ältesten Ideen der Informatik.

(3) Es ist auch möglich, jedes Informationswort dreimal zu senden. Der entsprechende Code hat die Generatormatrix

$$G = \begin{pmatrix} I_4 \\ I_4 \\ I_4 \end{pmatrix} \in K^{12\times 4}.$$

Dies ist ein (12,4)-Code. Falls höchstens ein Fehler auftritt, kann man diesen nach Empfang korrigieren. Man spricht von einem *1-fehlerkorrigierenden Code.* ◁

Das *Dekodieren* läuft folgendermaßen ab: Das empfangene Wort $c' = (c'_1, \ldots, c'_n)$ kann sich von dem gesendeten Wort c durch Übertragungsfehler unterscheiden. Falls c' ein Codewort ist, also $c' \in C$, so wird $c = c'$ angenommen, denn dann ist der wahrscheinlichste Fall, dass kein Fehler auftrat. In diesem Fall wird durch das Auflösen des LGS $G \cdot x = c'$ das (wahrscheinliche) Informationswort $x \in K^k$ ermittelt. Interessanter ist der Fall $c' \notin C$. Es wird (wieder) mit der Annahme gearbeitet, dass die Anzahl der Fehlerbits mit großer Wahrscheinlichkeit klein ist. Also sucht man ein Codewort $c'' \in C$, das sich von c' an möglichst wenig Koordinaten unterscheidet. Falls es genau ein solches c'' gibt, wird $c = c''$ angenommen und $x \in K^k$ mit $G \cdot x = c''$ ausgegeben. Andernfalls wird eine Fehlermeldung ausgegeben: dann ist sinnvolles Dekodieren nicht möglich. Die Güte eines Codes entscheidet sich darin, dass dieser Fall möglichst vermieden wird, und dass korrektes Dekodieren ($c'' = c$) mit möglichst hoher Wahrscheinlichkeit passiert.

Definition 7.3. *Für $c = (c_1, \ldots, c_n) \in K^n$ ist*

$$w(c) := \left| \left\{ i \in \{1, \ldots, n\} \,\middle|\, c_i \neq 0 \right\} \right|$$

das **Hamming-Gewicht** *von c. Für $c, c' \in K^n$ ist*

$$d(c, c') := w(c - c') = \left| \left\{ i \in \{1, \ldots, n\} \,\middle|\, c_i \neq c'_i \right\} \right|$$

der **Hamming-Abstand** *von c und c'. (Nebenbei: Dies ist eine* Metrik *auf K^n.) Für eine Teilmenge $C \subseteq K^n$ ist*

$$d(C) := \min \left\{ d(c, c') \,\middle|\, c, c' \in C,\ c \neq c' \right\}$$

der **Hamming-Abstand** *von C. (Falls $|C| \leq 1$, so setzen wir $d(C) := n+1$.) Falls C ein Unterraum ist, ergibt sich*

$$d(C) = \min \left\{ w(c) \middle| c \in C \setminus \{0\} \right\}.$$

Beispiel 7.4. (1) Der (8,4)-Wiederholungscode (Beispiel 7.2(1)) hat $d(C) = 2$.

(2) Der (5,4)-Parity-Check-Code (Beispiel 7.2(2)) hat ebenfalls $d(C) = 2$.

(3) Der (12,4)-Wiederholungscode (Beispiel 7.2(3)) hat $d(C) = 3$. ◁

Folgende Überlegung zeigt, dass der Hamming-Abstand entscheidend ist für die Güte eines Codes.

Es sei zunächst $d(C) = 2e + 1$ ungerade. Das (durch Übertragungsfehler bedingte) Ändern von höchstens e Bits in einem Codewort ergibt ein $c' \in K^n$ mit $d(c, c') \leq e$. Dann ist c das eindeutig bestimmte Codewort $c'' \in C$ mit $d(c'', c') \leq e$. Aus $d(c'', c') \leq e$ und $c'' \in C$ folgt nämlich $d(c'', c) \leq 2e$, also $c'' = c$ wegen der Annahme. Dies bedeutet, dass korrekt dekodiert wird, falls höchstens e Übertragungsfehler auftreten. Der Code ist also e-fehlerkorrigierend. (Bei mehr als e Fehlern ist allerdings eine misslungene oder gar falsche Dekodierung möglich.)

Nun sei $d(C) = 2e + 2$ gerade. Nach obigem Argument ist C auch e-fehlerkorrigierend. Zusätzlich gilt: Bei $e + 1$ Fehlern gibt es kein Codewort $c'' \in C$ mit $d(c'', c') \leq e$ (denn dann wäre $c'' \neq c$ und $d(c, c'') \leq d(c, c') + d(c', c'') \leq e + 1 + e < d(C)$, ein Widerspruch). Dies bedeutet, dass $e + 1$ Fehler erkannt werden können. Bei $e + 1$ Fehlern kann aber (in der Regel) nicht mehr dekodiert werden. Ein Code mit Hamming-Abstand $2e+2$ ist also e-fehlerkorrigierend und $(e + 1)$-fehlererkennend.

Wir fassen zusammen:

Satz 7.5. *Sei $C \subseteq K^n$ ein Code.*

(a) Falls $d(C) = 2e + 1$, so ist C e-fehlerkorrigierend.
(b) Falls $d(C) = 2e+2$, so ist C e-fehlerkorrigierend und $(e+1)$-fehlererkennend.

Alles, was wir über das Dekodieren und den Hamming-Abstand gesagt haben, gilt auch für nicht-lineare Codes. Nun erinnern wir uns, dass wir lineare Codes betrachten wollen. Außerdem beobachten wir, dass in allen bisherigen Beispielen die Generatormatrix die Form

$$G = \begin{pmatrix} I_k \\ A \end{pmatrix} \tag{7.1}$$

mit $A \in K^{(n-k)\times k}$ hat. Wir machen dies zu einer weiteren Voraussetzung und bilden die Matrix

$$P := \begin{pmatrix} -A & I_{n-k} \end{pmatrix} \in K^{(n-k)\times n}.$$

P hat den Rang $n - k$, und es gilt

$$P \cdot G = \begin{pmatrix} -A & I_{n-k} \end{pmatrix} \cdot \begin{pmatrix} I_k \\ A \end{pmatrix} = 0.$$

Hieraus folgt $P \cdot c = 0$ für alle $c \in C$. Andererseits hat die Lösungsmenge L des homogenen LGS $P \cdot x = 0$ nach Proposition 6.11 die Dimension $n-(n-k) = k = \dim(C)$. Wegen Korollar 6.16(b) folgt $L = C$. Wir halten fest, dass für $c \in K^n$ gilt:

$$c \in C \quad \Longleftrightarrow \quad P \cdot c = 0.$$

P heißt die **Parity-Check-Matrix**. Nebenbei sei erwähnt, dass für lineare Codes auch ohne die Voraussetzung (7.1) eine Parity-Check-Matrix existiert.

Beispiel 7.6. (1) Der (8,4)-Wiederholungscode (Beispiel 7.2(1)) hat die Parity-Check-Matrix

$$P = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{pmatrix} \in K^{4 \times 8}.$$

(2) Der (5,4)-Parity-Check-Code (Beispiel 7.2(2)) hat die Parity-Check-Matrix

$$P = \begin{pmatrix} -1 & -1 & -1 & -1 & 1 \end{pmatrix} \in K^{1 \times 5}.$$

Mit Hilfe der Parity-Check-Matrix kann man das Dekodierungsverfahren verbessern. Es sei $c' \in K^n$ das empfangene Wort. Den Unterschied von c und c' quantifizieren wir durch den (dem Empfänger nicht bekannten) *Fehlervektor* $f := c' - c \in K^n$. Es ergibt sich

$$P \cdot c' = P \cdot (c + f) = 0 + P \cdot f = P \cdot f.$$

Der Vektor $P \cdot c' \in K^{n-k}$ heißt das **Syndrom** von c'. Es misst, wie weit c' von einem Codewort abweicht. Nach obiger Gleichung haben empfangenes Wort und Fehlervektor das gleiche Syndrom. Das Dekodieren kann nun so geschehen: Man berechnet das Syndrom $P \cdot c'$. Nun sucht man ein $f \in K^n$, welches unter allen $f' \in K^n$ mit $P \cdot f' = P \cdot c'$ minimales Hamming-Gewicht hat. Falls $c' \in C$, so ergibt sich automatisch $f = 0$. Falls es ein eindeutig bestimmtes solches f gibt, setzt man $c'' := c' - f \in C$ und gibt $x \in K^k$ mit $G \cdot x = c''$ aus. Falls es kein eindeutiges f gibt, gibt man eine Fehlermeldung aus. Dies entspricht genau dem oben beschriebenen Dekodierungsverfahren. Da es nur $|K|^{n-k}$ mögliche Syndrome gibt, kann man das f (oder Fehlermeldung) zu jedem Syndrom in einer Tabelle speichern. Oft gibt es noch bessere Methoden zur Ermittlung von f. Dies ist in folgendem Beispiel der Fall.

Der (7,4)-Hamming-Code

Wir definieren nun den sogenannten (7,4)-Hamming-Code. Dieser zeigt, dass Codierungstheorie zu mehr in der Lage ist, als die bisherigen, relativ offensichtlichen Beispiele von Codes zu analysieren. Der Hamming-Code $C \subset \mathbb{F}_2^7$

wird durch die Generatormatrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \in \mathbb{F}_2^{7\times 4}$$

definiert, als Abbildung $\mathbb{F}_2^4 \to \mathbb{F}_2^7$ also $(x_1, \dots, x_4) \mapsto (x_1, x_2, x_3, x_4, x_2 + x_3 + x_4, x_1 + x_3 + x_4, x_1 + x_2 + x_4)$. C ist ein (7,4)-Code, hat also höhere Informationsrate als der (8,4)-Wiederholungscode aus Beispiel 7.2(1). Die Parity-Check-Matrix ist

$$P = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Welchen Hamming-Abstand hat C? Dazu müssen wir $w(c)$ für $c \in C \setminus \{0\}$ ermitteln. Die Bedingung $c \in C$ ist gleichbedeutend mit $P \cdot c = 0$. Gibt es ein solches c mit $w(c) = 1$? Dies würde bedeuten, dass (mindestens) eine der Spalten von P eine Nullspalte ist, was nicht der Fall ist. Gibt es ein $c \in \mathbb{F}_2^7$ mit $P \cdot c = 0$ und $w(c) = 2$? Dies würde bedeuten, dass es in P zwei Spalten gibt, die linear abhängig sind. Auch dies ist nicht der Fall! Es folgt also $d(C) > 2$. In diesem Argument zeigt sich die eigentliche Idee des Hamming-Codes: Man beginnt mit der Parity-Check-Matrix und stellt sie so auf, dass sie keine zwei linear abhängigen Spalten enthält. Hieraus folgt dann $d(C) > 2$. Die Generatormatrix G leitet man dann aus der Parity-Check-Matrix her. Da C selbst (sogar mehr als) einen Vektor von Gewicht 3 enthält, folgt

$$d(C) = 3.$$

Der (7,4)-Hamming Code ist also 1-fehlerkorrigierend. Damit hat er einerseits eine höhere Informationsrate, andererseits bessere Fehlerkorrektureigenschaften als der (8,4)-Wiederholungscode!

Das Dekodieren ist hier ganz besonders einfach: Es gibt nur acht mögliche Syndrome, nämlich alle Vektoren von $\mathbb{F}_2^3$. Wir können diese schreiben als $v_0 = 0, v_1, \dots, v_7$, wobei v_i die i-te Spalte von P ist $(i > 0)$. Für v_0 ist der Nullvektor das Codewort kleinsten Gewichtes mit Syndrom v_0. Für v_i $(i > 0)$ ist dies der i-te Standardbasisvektor e_i, denn $P \cdot e_i = v_i$. Der vollständige Dekodieralgorithmus läuft also so ab: Man ermittelt das Syndrom $s := P \cdot c'$ des empfangenen Wortes $c' = (c'_1, \dots, c'_7)$. Falls $s = v_i$ mit $1 \le i \le 4$, so gibt man $(x_1, \dots, x_4) = (c'_1, \dots, c'_4) + e_i$ aus (d.h. das i-te Bit wird geändert). Andernfalls gibt man $(x_1, \dots, x_4) = (c'_1, \dots, c'_4)$ aus. (Falls das Syndrom einer der Vektoren v_5, v_6, v_7 ist, so wird e_i mit $i > 4$ zu c' hinzuaddiert, aber dies ändert $(x_1, \dots, x_4)$ nicht.) In dem wahrscheinlichen Fall, dass bei der Übertragung höchstens ein Fehler auftritt, wird so das korrekte Informationswort ausgegeben.

Der Bauer-Code

Einen weiteren interessanten Code erhalten wir durch folgende Erweiterung des (7,4)-Hamming Codes: Wir hängen einfach zusätzlich noch ein Parity-Bit $c_8 = c_1 + \dots + c_7$ an, d.h. wir benutzen die Abbildung

$$(x_1, \dots, x_4) \mapsto (x_1, x_2, x_3, x_4, x_2{+}x_3{+}x_4, x_1{+}x_3{+}x_4, x_1{+}x_2{+}x_4, x_1{+}x_2{+}x_3).$$

Der hierdurch definierte Code C wird *Bauer-Code* (nach F. L. Bauer, Informatiker an der TU München) genannt. Es ist ein (8,4)-Code. Was ist der Hamming-Abstand $d(C)$? Auf jeden Fall mindestens 3, denn die ersten 7 Bits sind ja identisch mit dem Hamming-Code. Aber falls ein Wort $(c_1, \dots, c_7)$ des Hamming-Codes das Gewicht 3 hat, so ist $c_1 + \dots + c_7 = 1$, also hat das entsprechende Wort in C Gewicht 4. Wir erhalten $d(C) = 4$. Der Bauer-Code ist also 1-fehlerkorrigierend und 2-fehlererkennend. Er hat damit wesentlich bessere Eigenschaften als der (8,4)-Wiederholungscode.

Kapitel 8
Lineare Abbildungen

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

Definition 8.1. *Eine Abbildung $\varphi\colon V \to W$ 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 8.2. (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 8.3. *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 8.4. *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 8.5. (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 8.6. *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

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

definieren. Die lineare Unabhängigkeit von B liefert $\mathrm{Kern}(\varphi) = \{0\}$, also ist φ nach Satz 8.4(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 8.7. *Es sei $n := \dim(V) < \infty$. Dann gilt*

$$V \cong K^n.$$

Beispiel 8.8. $V = \{f \in K[x] \mid \deg(f) < 3\} \cong K^3$. Ein Isomorphismus wird gegeben durch

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

◁

Der Isomorphismus aus Satz 8.7 kann immer erst nach Wahl einer Basis angegeben werden. Man spricht auch von einem *nicht kanonischen* Isomorphismus. Satz 8.7 besagt, dass man sich beim Studium von endlich-dimensionalen Vektorräumen immer auf den Fall $V = K^n$ zurückziehen kann.

Satz 8.9 (Dimensionssatz für lineare Abbildungen)**.** *Sei $\varphi\colon V \to W$ linear. Dann gilt:*

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

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

$$a_1 v_1 + \cdots + a_m v_m + b_1 v_1' + \cdots + b_n v_n' = 0 \tag{8.1}$$

mit $a_i, b_i \in K$. Anwendung von φ auf (8.1) liefert:

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

Wegen der linearen Unabhängigkeit der w_i liefert dies $b_1 = \cdots = b_n = 0$. Nun folgt aus (8.1)

$$a_1 v_1 + \cdots + a_m v_m = 0,$$

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 $\varphi(v) = \sum_{i=1}^{n} b_i w_i$ mit $b_i \in K$. Mit $\widetilde{v} := v - \sum_{i=1}^{n} b_i v_i'$ folgt

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

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

$$\widetilde{v} = a_1 v_1 + \cdots + a_m v_m.$$

Insgesamt erhalten wir

$$v = \widetilde{v} + \sum_{i=1}^{n} b_i v_i' = \sum_{i=1}^{m} a_i v_i + \sum_{i=1}^{n} b_i v_i',$$

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

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

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 8.9 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 8.10. *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 8.10 gegebenen Interpretationen des Rangs laufen unter der Merkregel

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

Korollar 8.11. *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 8.4(c) ist Injektivität gleichbedeutend mit $\mathrm{Kern}(\varphi) = \{0\}$, also mit $\dim(\mathrm{Kern}(\varphi)) = 0$. Wegen Satz 8.9 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 8.11 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 alle Pivotelemente 1 sind.
(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 8.12. *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

$$A \text{ invertierbar} \iff A \text{ regulär}.$$

Beispiel 8.13. 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. $\triangleleft$

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

Satz 8.14 (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 9
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 9.1. *Die oben definierte Matrix A heißt die* **Darstellungsmatrix** *von φ (bezüglich der Basen B und C). Schreibweise:*

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

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

Anmerkung 9.2. (a) Die Notation $D_{C,B}(\varphi)$ dieser Vorlesung ist nicht allgemein gebräuchlich. Viele Lehrbücher verwenden für die Darstellungsmatrix andere oder gar keine Notation.

(b) Es erscheint zunächst unnatürlich, dass bei $D_{C,B}(\varphi)$ die Basis des Zielraums W als erstes und die des Definitionsraums V als zweites geschrieben wird. Der Grund hierfür ist, dass sich durch diese Konvention wesentlich schönere und leichter zu merkende Formeln ergeben, etwa in Satz 9.5. ◁

Als Merkregel halten wir fest:

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

Wegen Satz 8.14 ist φ durch seine Darstellungsmatrix eindeutig bestimmt, und jede Matrix taucht als Darstellungsmatrix einer linearen Abbildung auf.

Beispiel 9.3. (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 8.2(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 9.1 ein Zusammenhang?

Satz 9.4. *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 $A := D_{C,B}(\varphi)$ 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 $\varphi = \varphi_A$. $\square$

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 9.5. *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*

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

Als Merkregel halten wir fest:

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

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

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

Für $j \in \{1, \dots, n\}$ 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 $D_{C,B}(\psi) \cdot D_{B,A}(\varphi)$ ist, folgt die Behauptung. $\square$

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 9.5 gilt.

Kombiniert man den Satz mit Satz 9.4, 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 8.12 entstanden waren.

Definition 9.6. *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 $B = \{v_1, \dots, v_n\}$ eine Basis von V, und $B' = \{v_1', \dots, v_n'\}$ sei eine weitere Basis. Wir können die „neuen“ Basisvektoren v_j' mit Hilfe der alten ausdrücken:

$$v_j' = \sum_{i=1}^{n} a_{i,j} v_i \tag{9.1}$$

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 $j \in \{1, \dots, n\}$ 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 (9.1) 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 9.7. *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 9.8. *Es seien B, B' endliche Basen von V und C, C' endliche Basen von W. Dann gilt für eine lineare Abbildung $\varphi\colon V \to W$:*

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

Anmerkung. Man kann Satz 9.8 ziemlich bequem aus Satz 9.5 herleiten, indem man die Basiswechselmatrix $S_{B,B'}$ als Darstellungsmatrix der Identität bezüglich der Basen B und B' interpretiert, also $S_{B,B'} = D_{B,B'}(\mathrm{id}_V)$. Satz 9.7 ergibt sich dann als Forgerung von Satz 9.8. ◁

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

Definition 9.9. *(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 9.10. 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 in verschiedenen Weisen zu interpretieren. Dies führt immer wieder zu großer Verwirrung. Man sollte den Formalismus des Basiswechsels als Rezept verstehen, dessen Korrektheit bewiesen wurde, und wortwörtlich anwenden.

Kapitel 10
Determinanten

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

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

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

Definition 10.1. *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 10.2. (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 $i > 2$. (Für Leserinnen und Leser, denen die Zykelschreibweise geläufig ist: $\sigma = (1,2)$.) 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 $\{1,\ldots,n\}$ fest. (In Zykelschreibweise: $\sigma = (i,j)$.) 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 10.3. *Für $\sigma, \tau \in S_n$ gilt*

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

Die Abbildung $\operatorname{sgn}\colon S_n \to \{1,-1\}$ ist also ein Gruppen-Homomorphismus.

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

$$\operatorname{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{x_{\sigma(i)} - x_{\sigma(j)}}{x_i - x_j}. \tag{10.1}$$

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 (10.1).

Nun setzen wir $y_i := x_{\sigma(i)}$. Ebenso wie die x_i sind auch die y_i paarweise verschieden, also gilt wegen (10.1) 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

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

$\square$

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

Definition 10.4. *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 10.5. Für $n \le 3$ machen wir Definition 10.4 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 10.2(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*:

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

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

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

Ab jetzt behandeln wir nur noch die Determinante.

Lemma 10.6. *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 Spezialfall 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

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

wobei im letzten Schritt die Regel $\operatorname{sgn}(\tau) = \operatorname{sgn}(\tau^{-1})$ verwendet wurde, die aus Satz 10.3 folgt.

(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 10.3 für die letzte Gleichheit benutzt wurde. Satz 10.3 liefert auch $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$, 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 $1 \le j < k \le n$ 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 10.2(3)). Für alle $i, l \in \{1, \ldots, n\}$ gilt dann

$$a_{i,l} = a_{i,\tau(l)}. \tag{10.2}$$

Wir definieren

$$A_n := \{\sigma \in S_n \mid \operatorname{sgn}(\sigma) = 1\}.$$

(Nebenbei gesagt folgt aus Satz 10.3, dass A_n eine Untergruppe der S_n ist; sie heißt die *alternierende Gruppe.*) Wegen $\operatorname{sgn}(\tau) = -1$ folgt aus Satz 10.3, dass S_n die disjunkte Vereinigung von A_n und $\tau \cdot A_n := \{\tau\sigma \mid \sigma \in A_n\}$ ist:

$$S_n = A_n \,\dot{\cup}\, \tau \cdot A_n.$$

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 (10.2) für die letzte Gleichheit verwendet wurde. □

Der wohl wichtigste Satz über die Determinante ist der folgende.

Satz 10.7 (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 10.6(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 10.6(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 10.8. *Für $A \in K^{n\times n}$ gilt die Äquivalenz*

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

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 10.7 und Beispiel 10.5(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 $\{v = v_1, v_2, \ldots, v_n\}$ von K^n. $B := (v_1, \ldots, v_n) \in K^{n \times n}$ sei die Matrix mit den v_i als Spalten. Da die Spalten von B linear unabhängig sind, folgt nach Korollar 8.10, 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 10.7 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: $A \in \mathrm{GL}_n(K)$);
- 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 10.7.

Korollar 10.9. *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 10.7 und 10.8 folgt

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

□

Korollar 10.9 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 9.7 über in eine ähnliche Matrix.

Definition 10.10. *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 10.7 folgt, dass* $\mathrm{SL}_n(K)$ *eine Untergruppe der* $\mathrm{GL}_n(K)$ *ist, womit* $\mathrm{SL}_n(K)$ *selbst eine Gruppe ist.*

Für den Rest des Kapitels beschäftigen wir uns mit dem effizienten Berechnen der Determinante. Die Definition 10.4 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 10.11. *Es sei* $A = (a_{i,j}) \in K^{n\times n}$ *mit* $n \geq 2$*. Für* $i, j \in \{1, \ldots, n\}$ *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* $i \in \{1, \ldots, n\}$ *gilt*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det(A_{i,j}), \tag{10.3}$$

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

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \cdot \det(A_{i,j}). \tag{10.4}$$

Wir lassen den Beweis weg. Er ist nicht besonders schwer, aber sehr notationslastig. Man beachte, dass die Formeln (10.3) und (10.4) das rekursive Berechnen der Determinante ermöglichen. Die Berechnung gemäß Formel (10.3) wird als *Entwicklung nach der* i*-ten Zeile* bezeichnet, und gemäß (10.4) als *Entwicklung nach der* j*-ten Spalte*. Man kann eine dieser Formeln anwenden und dabei i bzw. j nach Opportunitätsgesichtspunkten auswählen.

Beispiel 10.12. 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 10.11 und Lemma 10.6(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 8 angegebene Verfahren. Die Formel kann jedoch nützlich sein, wenn in A Parameter vorkommen, oder um die auftretenden Nenner zu kontrollieren.

Beispiel 10.13. 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 10.4 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*

$$A = \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix} \tag{10.5}$$

gilt

$$\det(A) = a_1 \cdots a_n. \tag{10.6}$$

Zur Erklärung: (10.5) 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 (10.6) gilt auch für untere Dreiecksmatrizen.

(3) Für eine Matrix

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

mit $B \in K^{l\times l}$, $D \in K^{(n-l)\times(n-l)}$ und $C \in K^{(n-l)\times l}$ 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 10.6(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 10.7 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 10.7 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 10.14. 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).

◁

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 11
Eigenwerte

Auch in diesem Kapitel sei K ein Körper.

Definition 11.1. *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. E_λ ist auch definiert, wenn $\lambda \in K$* kein *Eigenwert ist.*

Beispiel 11.2. 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. $\triangleleft$

Definition 11.1 lässt sich auf lineare Abbildungen $\varphi\colon V \to V$ eines K-Vektorraums V übertragen: Man fordert $\varphi(v) = \lambda v$ für ein $v \in V \setminus \{0\}$. Auch für $\lambda \in K$, die nicht Eigenwerte sind, kann man E_λ wie in Definition 11.1 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 11.3. *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 10 äquivalent zu $\det(A - \lambda I_n) = 0$ (siehe Seite 68). Diese Überlegungen nehmen wir zum Anlass für eine Definition.

Definition 11.4. *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 11.5. *Die Eigenwerte einer quadratischen Matrix A sind die Nullstellen des charakteristischen Polynoms χ_A.*

Beispiel 11.6. (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 11.7. 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 verbleibende 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:

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

◁

Über $\mathbb{R}$ zerfallen also nicht alle Polynome. Wie das obige Beispiel suggeriert, haben wir über $\mathbb{C}$ bessere Chancen. In der Tat gilt der folgende

Satz 11.8 (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 11.8 haben, bezeichnet man als **algebraisch abgeschlossen**. Wir werden Satz 11.8 nicht beweisen, da dies mit unseren derzeitigen Mitteln nicht möglich ist. Der Beweis benötigt Methoden aus der Funktionentheorie (= komplexe Analysis) oder der Analysis.

Korollar 11.9. *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 11.10. *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 11.11. (1) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ hat die Eigenwerte 1 und -1 (siehe Beispiel 11.2). 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 11.12. *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 $\dim(E_\lambda) \geq 1$.

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 $1 \leq i \leq m$ 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 9.7), 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 10.9 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 11.13. *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 11.14. (1) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist diagonalisierbar (siehe Beispiel 11.2).

(2) $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ ist nicht diagonalisierbar. Es fehlen Eigenwerte (siehe Beispiel 11.6(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 11.11(2)). ◁

Wir werden folgendes Kriterium für Diagonalisierbarkeit beweisen. Es besagt, dass die in Beispiel 11.14(2) und (3) aufgetretenen Hindernisse für die Diagonalisierbarkeit tatsächlich die einzig möglichen Hindernisse sind.

Satz 11.15. *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 11.16. *Es seien $\lambda_1, \ldots, \lambda_r \in K$ paarweise verschiedene Eigenwerte einer Matrix $A \in K^{n\times n}$. Weiter seien $v_i \in E_{\lambda_i}$ mit*

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

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 $i \in \{2, \ldots, r\}$. Wegen $\lambda_i \neq \lambda_1$ folgt $v_i = 0$ für $i \in \{2, \ldots, r\}$. Nun folgt auch $v_1 = -(v_2 + \cdots + v_r) = 0$. □

Beweis von Satz 11.15. 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 $B_i := B \cap E_{\lambda_i}$:

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

wobei die mittlere Ungleichung aus Satz 11.12 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 11.16 folgt, dass B linear unabhängig ist. Außerdem gilt

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

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

Damit ist klar: Auch über einem algebraisch abgeschlossenen Körper sind nicht alle quadratischen Matrizen diagonalisierbar. Eine über keinem Körper diagonalisierbare Matrix findet sich in Beispiel 11.14(3). Es gilt aber die folgende Abschwächung:

Satz 11.17. *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 $\chi_A = \prod_{i=1}^{n}(x - \lambda_i)$.

Beweis. 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 $\{v_1, \dots, v_n\}$ von K^n. Wir bilden die Matrix $S = (v_1, \dots, v_n) \in \mathrm{GL}_n(K)$ mit den v_i als Spalten. Da v_1 ein Eigenvektor zum Eigenwert λ_1 ist, folgt

$$S^{-1}AS = \left(\begin{array}{c|c} \lambda_1 & * \cdots * \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right) \tag{11.1}$$

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 ähnliche Matrizen identische charakteristische Polynome haben. □

Anmerkung 11.18. (a) In Wirklichkeit gilt Satz 11.17 unter der allgemeineren Voraussetzung, dass das charakteristische Polynom χ_A in Linearfaktoren zerfällt. Unser Beweis funktioniert auch unter dieser Voraussetzung, denn aus (11.1) 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.

(c) In Wirklichkeit gilt eine wesentlich weiter reichende Aussage als die von Satz 11.17: Jede quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, ist ähnlich zu einer sogenannten *Jordan-Matrix*, d.h. einer Block-Diagonalmatrix

$$\begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}, \tag{11.2}$$

wobei jeder der Block die Gestalt

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & & 1 \\ 0 & & & \lambda_i \end{pmatrix} \in K^{k_i \times k_i}$$

(mit $\lambda_i \in K$, $k_i \in \mathbb{N}_{>0}$) hat. Die J_i nennt man *Jordan-Blöcke*. Im Fall $k_i = 1$ ist J_i ein einzelner Diagonal-Eintrag. Die Matrix (11.2) nennt man auch eine *Jordan-Normalform* von A. Der Beweis hierfür ist sehr viel komplizierter. $\triangleleft$

Kapitel 12
Die Google-Matrix und stochastische Matrizen

Im ersten Kapitel haben wir als Beispiel für eine Matrix die Weblink-Matrix betrachtet und einige interessante Beobachtungen gemacht. Auf dieses Thema kommen wir nun zurück. Als kurze Erinnerung: Wir betrachten die Seiten $P_1, \ldots, P_n$ des Internets (n geht also in die Milliarden!) und definieren die *Weblink-Matrix* $W = (w_{i,j}) \in \mathbb{R}^{n \times n}$ durch

$$w_{i,j} = \begin{cases} 1 & \text{falls } P_i \text{ einen Link auf } P_j \text{ enthält} \\ 0 & \text{sonst} \end{cases}.$$

In Beispiel 1.2(3) (auf Seite 6) findet sich ein konkretes (Miniatur-)Beispiel. Die Idee des *Google-Pageranks* ist es, dass sich die Wichtigkeit einer Seite P_i daraus ergibt, wie viele andere wichtige Seiten einen Link auf P_i enthalten. Hierbei ist das Gewicht eines Links durch die Gesamtzahl der von derselben Seite ausgehenden Links zu dividieren. In Beispiel 1.2(5) haben wir dann das Modell eines *Zufallssurfers* betrachtet und gefragt, mit welcher Wahrscheinlichkeit dieser nach einer großen Anzahl von Klicks auf einer bestimmten Seite landet. Am konkreten (Miniatur-)Beispiel haben wir die überraschende Beobachtung gemacht, dass diese Wahrscheinlichkeiten genau der Idee des Pageranks entsprechen. Wir haben angekündigt, dies allgemein nachzuweisen. Dieses Versprechen werden wir in diesem Kapitel einlösen.

Als ersten Schritt werden wir die Idee des Pageranks mathematisch ausdrücken. Wir ändern die Weblink-Matrix W ab, indem wir jede Zeile von W (die nicht lauter Nullen enthält) durch die Anzahl der Einsen in dieser Zeile dividieren. Wenn wir die so erhaltene Matrix mit H bezeichnen und die Wichtigkeit der Seite P_i mit α_i, dann übersetzt sich die Google-Idee in die Gleichung

$$p \cdot H = p \quad \text{mit} \quad p := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{1 \times n}. \tag{12.1}$$

Ebenso gut kann man $H^T \cdot p^T = p^T$ schreiben. Wir erhalten also die Bedingung, dass p^T ein Eigenvektor zum Eigenwert $\lambda = 1$ von H^T ist. Es sollte also 1 ein Eigenwert von H^T sein, und damit p zumindest bis auf skalare Vielfache eindeutig bestimmt ist, müsste die geometrische Vielfachheit 1 sein.

Beispiel 12.1. Wir knüpfen an Beispiel 1.2(3) (auf Seite 6) an. Die Matrizen W und H sind

$$W = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{und} \quad H = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ist $\lambda = 1$ ein Eigenwert von H^T? Es gilt

$$\chi_{H^T} = \chi_H = \det \begin{pmatrix} x & -\frac{1}{2} & -\frac{1}{2} \\ -1 & x & 0 \\ 0 & 0 & x \end{pmatrix} = x \cdot \det \begin{pmatrix} x & -\frac{1}{2} \\ -1 & x \end{pmatrix} = x(x^2 - \frac{1}{2}),$$

wobei wir nach der dritten Zeile entwickelt haben. Die Eigenwerte sind also 0 und $\pm\sqrt{2}/2$, aber nicht 1. Mit der so definierten Matrix H lässt sich also die Idee des Google-Pageranks nicht realisieren! ◁

Wie können wir garantieren, dass $\lambda = 1$ ein Eigenwert von H^T ist? Aufgrund der Charakterisierung von Eigenwerten als Nullstellen des charakteristischen Polynoms sehen wir, dass 1 genau dann ein Eigenwert von H^T ist, wenn es ein Eigenwert von H ist. Es fällt auf, dass bis auf die Nullzeilen alle Zeilen von H die Summe 1 haben. Anders ausgedrückt: Der Vektor $H \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ hat in allen Komponenten, wo H keine Nullzeile hat, den Eintrag 1. Wenn wir die Nullzeilen von H durch Zeilen mit der Summe 1 ersetzen, so wird $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ automatisch zu einem Eigenvektor zum Eigenwert 1. Dies motiviert die Bildung einer neuen Matrix $S \in \mathbb{R}^{n \times n}$, die aus H entsteht, indem jede Nullzeile in H durch die Zeile $(1/n, \dots, 1/n)$ ersetzt wird.

Beispiel 12.2. Im obigen Beispiel erhalten wir

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

Das charakteristische Polynom ist

$$\chi_S = \det \begin{pmatrix} x & -\frac{1}{2} & -\frac{1}{2} \\ -1 & x & 0 \\ -\frac{1}{3} & -\frac{1}{3} & x - \frac{1}{3} \end{pmatrix} \underset{(1)}{=} \det \begin{pmatrix} x & x^2 - \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3}(x+1) & x - \frac{1}{3} \end{pmatrix} \underset{(2)}{=}$$
$$\det \begin{pmatrix} x^2 - \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{3}(x+1) & x - \frac{1}{3} \end{pmatrix} = x^3 - \frac{1}{3}x^2 - \frac{2}{3}x = x(x-1)\left(x + \frac{2}{3}\right),$$

wobei wir im Schritt (1) das x-fache der ersten Spalte zu der zweiten addiert haben, und im Schritt (2) nach der zweiten Zeile entwickelt haben. Also hat S und damit auch S^T den Eigenwert 1 mit algebraischer Vielfachheit 1. Mit Satz 11.12 folgt, dass auch die geometrische Vielfachheit 1 ist, wodurch ein Vektor $p \in \mathbb{R}^{1\times 3}$ mit $p \cdot S = p$ bis auf skalare Vielfache eindeutig bestimmt ist. Eine Rechnung liefert

$$p = a \cdot (4, 3, 3) \quad \text{mit} \quad a \in \mathbb{R}.$$

Der Pagerank-Vektor ist also hinreichend eindeutig definiert. ◁

Wie können wir den Übergang von H zu S rechtfertigen? Einerseits durch die Auffassung, dass eine Seite ohne Outlinks ebensogut einen Link auf jede Seite haben könnte. Andererseits dadurch, dass die Definition von S konsistent mit unserem Modell des Zufallssurfers ist (siehe Beispiel 1.2(5) auf Seite 7).

Nach Konstruktion ist S eine (zeilen-)stochastische Matrix. Nach der vorausgegangenen Überlegung wird hierdurch garantiert, dass 1 ein Eigenwert von S und damit auch von S^T ist. Ist auch garantiert, dass die geometrische Vielfachheit 1 ist?

Beispiel 12.3. Wir betrachten ein weiteres Miniaturbeispiel, das aus vier Internet-Seiten $P_1, \ldots, P_4$ besteht, wobei P_1 und P_2 gegenseitig Links aufeinander haben, und ebenso P_3 und P_4. Wir erhalten

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = H = S.$$

Eine kurze Rechnung zeigt, dass S die Eigenwerte 1 und -1 hat, und dass der Eigenraum E_1 die Dimension 2 hat. Genauer: Jeder Vektor der Form (a, a, b, b) mit $a, b \in \mathbb{R}$ erfüllt (12.1), es fehlt also die Eindeutigkeit (bis auf skalare Vielfache). ◁

Im obigen Beispiel ist der durch die Links gegebene Graph nicht zusammenhängend. Um Hoffnung zu haben, dass die Matrix S den Eigenwert 1 mit geometrischer Vielfachheit 1 hat, müssen wir sie nochmals abändern, so dass ein Zufallssurfer mit positiver Wahrscheinlichkeit von jeder Seite zu jeder anderen gelangen kann. Wir tun dies, indem wir ein $\alpha \in \mathbb{R}$ mit $0 < \alpha < 1$ wählen und

$$G := (1 - \alpha) \cdot S + \frac{\alpha}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

setzen. Die Matrix G heißt die **Google-Matrix**. Diese Matrix wurde (oder wird) von der Google Suchmaschine verwendet, wobei Google (laut vorliegenden Informationen) $\alpha = 0.15$ wählte. Wie kann man die Definition von G

interpretieren oder rechtfertigen? Interpretiert als Transitionsmatrix, die das Verhalten unseres Zufallssurfers modelliert, deutet sich der Übergang von S zu G so: der Zufallssurfer entscheidet sich mit einer Wahrscheinlichkeit α, von der momentanen Seite aus keinem Link zu folgen, sondern die nächste Seite rein zufällig auszuwählen. Inwieweit dies ein Surferverhalten realistisch darstellt, sei dahingestellt. Letztendlich wird die Definition der Google-Matrix durch ihre mathematischen Eigenschaften, die wir nun herleiten werden, und durch den wirtschaftlichen Erfolg der Google Suchmaschine gerechtfertigt. Man sieht sofort, dass G ebenso wie S eine (zeilen-)stochastische Matrix ist. Außerdem sind alle Einträge von G positiv, was bei S nicht der Fall ist. Im folgenden werden wir Resultate für Matrizen mit diesen Eigenschaften herleiten.

Definition 12.4. *Eine Matrix $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ heißt* **stochastisch** *(oder auch* **zeilen-stochastisch***), falls $a_{i,j} \geq 0$ für alle i, j gilt und außerdem $\sum_{j=1}^n a_{i,j} = 1$ für alle i. A heißt* **positiv**, *falls $a_{i,j} > 0$ für alle i, j.*

Die Teile der obigen Definition sind unabhängig, d.h. eine positive Matrix muss nicht stochastisch sein. Außerdem ist der Begriff der Positivität zu unterscheiden von der positiven Definitheit, die wir in Kapitel 13 definieren werden.

Lemma 12.5. *Es seien $A, B \in \mathbb{R}^{n\times n}$ stochastisch. Dann ist auch $A \cdot B$ stochastisch.*

Beweis. Wir schreiben $A = (a_{i,j})$, $B = (b_{i,j})$ und $A \cdot B = (c_{i,j})$. Für alle i, j gilt

$$c_{i,j} = \sum_{k=1}^n a_{i,k} b_{k,j} \geq 0,$$

und für alle i gilt

$$\sum_{j=1}^n c_{i,j} = \sum_{j=1}^n \sum_{k=1}^n a_{i,k} b_{k,j} = \sum_{k=1}^n a_{i,k} \cdot \underbrace{\left(\sum_{j=1}^n b_{k,j} \right)}_{=1} = \sum_{k=1}^n a_{i,k} = 1.$$

□

Satz 12.6. *Sei $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ stochastisch.*

(a) A hat den Eigenwert 1, und für alle Eigenwerte $\lambda \in \mathbb{C}$ von A gilt: $|\lambda| \leq 1$. (Um von komplexen Eigenwerten sprechen zu können, fassen wir A als komplexe Matrix auf.)
(b) Falls A zusätzlich positiv ist, so gilt $m_a(1) = 1$ (die algebraische und damit auch die geometrische Vielfachheit ist 1), und für alle Eigenwerte $\lambda \in \mathbb{C} \setminus \{1\}$ gilt $|\lambda| < 1$.

Beweis. (a) Wegen

$$A \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{12.2}$$

ist 1 ein Eigenwert von A. Es sei $\lambda \in \mathbb{C}$ irgendein Eigenwert. Hierzu sei $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}^n \setminus \{0\}$ ein Eigenvektor. Wir wählen i so, dass $|x_i|$ maximal wird. Es gilt $\sum_{j=1}^n a_{i,j} x_j = \lambda \cdot x_i$, also

$$|\lambda| \cdot |x_i| = \left| \sum_{j=1}^n a_{i,j} x_j \right| \le \sum_{j=1}^n |a_{i,j} \cdot x_j| = \sum_{j=1}^n a_{i,j} \cdot |x_j| \le \sum_{j=1}^n a_{i,j} \cdot |x_i| = |x_i|. \tag{12.3}$$

Hieraus ergibt sich $|\lambda| \le 1$.

(b) Es sei $\lambda \in \mathbb{C}$ ein Eigenwert von A mit $|\lambda| = 1$ und v wie oben ein Eigenvektor. Dann gilt Gleichheit in (12.3), also

$$\left| \sum_{j=1}^n a_{i,j} x_j \right| = \sum_{j=1}^n |a_{i,j} \cdot x_j|$$

und

$$\sum_{j=1}^n a_{i,j} \cdot |x_j| = \sum_{j=1}^n a_{i,j} \cdot |x_i|.$$

Da alle $a_{i,j}$ positiv sind, folgt aus der zweiten Gleichung, dass alle $|x_j|$ gleich sind. Die erste Gleichung liefert durch mehrfache Anwendung von (3.1), dass alle $\frac{x_j}{|x_j|}$ gleich sind. Es folgt $v = a \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ mit $a \in \mathbb{C}$. Wegen (12.2) folgt $A \cdot v = v$, also $\lambda = 1$. Damit ist gezeigt, dass jeder Eigenwert $\neq 1$ betragsmäßig kleiner als 1 ist, und außerdem, dass $\langle \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \rangle$ der Eigenraum zu $\lambda = 1$ ist, also

$$m_g(1) = 1. \tag{12.4}$$

Es bleibt zu zeigen, dass auch $m_a(1) = 1$. Wir nehmen an, dass $m_a(1) > 1$ gilt. Dann liefert Satz 11.17 (zusammen mit Anmerkung 11.18(a)), dass $S \in \mathrm{GL}_n(\mathbb{C})$ existiert, so dass

$$S^{-1}AS = \begin{pmatrix} 1 & a & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 0 & \lambda_3 & & * \\ \vdots & & & \ddots & * \\ 0 & & \cdots & 0 & \lambda_n \end{pmatrix},$$

wobei $\lambda_3, \dots, \lambda_n$ Eigenwerte sind und $a \in \mathbb{C}$. Wegen (12.4) gilt $a \neq 0$. Für $k \in \mathbb{N}$ gilt

$$S^{-1}A^kS = \left(S^{-1}AS\right)^k = \begin{pmatrix} 1 & k \cdot a & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 0 & \lambda_3^k & & * \\ \vdots & & & \ddots & * \\ 0 & & \cdots & 0 & \lambda_n^k \end{pmatrix}.$$

Ist $u \in \mathbb{C}^n$ die zweite Spalte von S und $w \in \mathbb{C}^{1 \times n}$ die erste Zeile von S^{-1}, so folgt

$$w \cdot A^k \cdot u = k \cdot a.$$

Wir wählen $c \in \mathbb{R}$ so, dass alle Komponenten w_i von w und u_i von u betragsmäßig durch c nach oben beschränkt sind. Der (i,j)-te Eintrag von A^k sei mit $(A^k)_{i,j}$ bezeichnet. Dann folgt

$$\begin{aligned} k \cdot |a| = \left|w \cdot A^k \cdot u\right| = \left|\sum_{i,j=1}^n w_i (A^k)_{i,j} u_j\right| &\leq \sum_{i,j=1}^n |w_i| \cdot |u_j| \cdot |(A^k)_{i,j}| \\ &\leq c^2 \sum_{i,j=1}^n (A^k)_{i,j} = n \cdot c^2, \end{aligned}$$

wobei wir für die letzte Gleichheit benutzt haben, dass A^k nach Lemma 12.5 stochastisch ist. Da diese Ungleichung für alle $k \in \mathbb{N}$ gilt, folgt der Widerspruch $a = 0$. Wir schließen $m_a(1) = 1$. $\square$

Nach Satz 12.6 gilt für alle Eigenwerte $\lambda \neq 1$ der Google-Matrix G oder ihrer Transponierten G^T die Ungleichung $|\lambda| < 1$. In Wirklichkeit gilt sogar $|\lambda| \leq 1 - \alpha$ (mit dem α, das für die Bildung der Google-Matrix gewählt wurde). Diese schärfere Ungleichung werden wir nicht beweisen oder benutzen.

Beispiel 12.7. Wir betrachten einige Beispiele von stochastischen Matrizen und deren Eigenwerten.

(1) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ hat die Eigenwerte 1 und -1.

(2) $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ hat den Eigenwert 1 mit algebraischer und geometrischer Vielfachheit 2.

(3) $A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ hat die Eigenwerte 1 und 0.

(4) Die Google-Matrix unseres Miniatur-Beispiels ist

$$G = 0.85 \cdot \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + \frac{0.15}{3} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Die Eigenwerte sind 1, 0 und $\frac{-17}{30} \approx -0.567$. ◁

Den Pagerank-Vektor kann man erhalten, indem man einen Eigenvektor zum Eigenwert 1 von G^T ausrechnet. Normalerweise tut man das mit dem Gauß-Algorithmus. Dessen Aufwand beträgt aber $O(n^3)$ Körperoperationen. Bei $n \approx 10^9$ bedeutet das einen Aufwand in der Größenordnung von 10^{27} *Floating Point* Operationen. Selbst ein Petaflop-Rechner würde größenordnungsmäßig 100000 Jahre benötigen! Wir müssen also eine bessere Methode zum Berechnen eines Eigenvektors finden. Hier kommt der Zufallssurfer wieder ins Spiel. In Beispiel 1.2(5) haben wir beobachtet, dass die Potenzen der Google-Matrix (im Beispiel hatten wir $\alpha = 0$) gegen eine Matrix konvergieren, die lauter identische Zeilen hat, und deren Zeilen zudem der Eigenschaft eines Pagerank-Vektors genügen. Als Formeln ausgedrückt:

$$\lim_{k\to\infty} G^k = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}, \quad \text{wobei} \quad G^T \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}. \tag{12.5}$$

Wir werden nun beweisen, dass dies allgemein für positive, stochastische Matrizen gilt.

Laut den Sätzen 11.17 und 12.6 gilt für jede positive, stochastische Matrix $A \in \mathbb{R}^{n\times n}$

$$A \sim \begin{pmatrix} 1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \quad \text{mit} \quad \lambda_2, \dots, \lambda_n \in \mathbb{C},\ |\lambda_i| < 1,$$

wobei wir die Schreibweise $A \sim B$ für die Ähnlichkeit von A und B steht. Aus dem folgenden Lemma folgt, dass sogar noch etwas mehr gilt.

Lemma 12.8. *Es sei*

$$A = \left(\begin{array}{c|ccc} 1 & b_2 & \cdots & b_n \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right) \in K^{n\times n}$$

mit $B \in K^{(n-1)\times(n-1)}$ *(K irgendein Körper), so dass 1* kein *Eigenwert von B ist. Dann gilt*

$$A \sim \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right).$$

Beweis. Nach Voraussetzung ist $B - I_{n-1}$ invertierbar. Mit $b := (b_2, \dots, b_n) \in K^{1\times(n-1)}$ können wir also $x := b \cdot (B - I_{n-1})^{-1}$ setzen. Es folgt

$$x + b - x \cdot B = x \cdot (I_{n-1} - B) + b = 0. \tag{12.6}$$

Mit

$$S := \left(\begin{array}{c|c} 1 & x \\ \hline 0 & \\ \vdots & I_{n-1} \\ 0 & \end{array}\right) \in \mathrm{GL}_n(K)$$

gilt

$$S^{-1}AS = \left(\begin{array}{c|c} 1 & -x \\ \hline 0 & \\ \vdots & I_{n-1} \\ 0 & \end{array}\right) \cdot \left(\begin{array}{c|c} 1 & b \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right) \cdot \left(\begin{array}{c|c} 1 & x \\ \hline 0 & \\ \vdots & I_{n-1} \\ 0 & \end{array}\right) =$$

$$\left(\begin{array}{c|c} 1 & -x \\ \hline 0 & \\ \vdots & I_{n-1} \\ 0 & \end{array}\right) \cdot \left(\begin{array}{c|c} 1 & x+b \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right) = \left(\begin{array}{c|c} 1 & x+b-x\cdot B \\ \hline 0 & \\ \vdots & B \\ 0 & \end{array}\right),$$

woraus mit (12.6) die Behauptung folgt. □

Anmerkung. Der Beweis von Lemma 12.8 wäre überflüssig gewesen, wenn wir die Existenz der Jordan-Normalform vorausgesetzt hätten. ◁

Um (12.5) zu beweisen, brauchen wir zwei Lemmata und die folgende *ad hoc* Schreibweise: Für $A = (a_{i,j}) \in \mathbb{C}^{n\times n}$ und $C = (c_{i,j}) \in \mathbb{R}^{n\times n}$ schreiben wir $A \le C$, falls für alle $i, j \in \{1, \dots, n\}$ gilt: $|a_{i,j}| \le c_{i,j}$.

Lemma 12.9. *Es seien* $A, B \in \mathbb{C}^{n\times n}$ *und* $C, D \in \mathbb{R}^{n\times n}$ *mit* $A \le C$ *und* $B \le D$. *Dann folgt*

$$AB \le CD.$$

Beweis. Mit den üblichen Bezeichnungen für die Einträge unserer Matrizen gilt für alle $i, j \in \{1, \dots, n\}$

$$\left|\sum_{k=1}^{n} a_{i,k} b_{k,j}\right| \le \sum_{k=1}^{n} |a_{i,k}| \cdot |b_{k,j}| \le \sum_{k=1}^{n} c_{i,k} d_{k,j}.$$

Dies ergibt die Behauptung. □

Lemma 12.10. *Es sei*

$$B = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \in \mathbb{C}^{n\times n}$$

mit $|\lambda_i| < 1$ für alle i. Dann gilt

$$\lim_{k\to\infty} B^k = 0.$$

Hiermit ist gemeint, dass es für jedes reelle, positive ε ein $m \in \mathbb{N}$ gibt, so dass

$$B^k \le \begin{pmatrix} \varepsilon & \cdots & \varepsilon \\ \vdots & & \vdots \\ \varepsilon & \cdots & \varepsilon \end{pmatrix}$$

für $k \ge m$ gilt, d.h. jeder Matrix-Eintrag konvergiert gegen 0.

Beweis. Wir setzen $\lambda := \max\{|\lambda_1|, \ldots, |\lambda_n|\} < 1$ und wählen $c \in \mathbb{R}$ als obere Schranke für die Beträge sämtlicher Einträge von B. Dann gilt

$$B \le \begin{pmatrix} \lambda & c & \cdots & c \\ 0 & \lambda & & c \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda \end{pmatrix} = \lambda \cdot I_n + c \cdot \underbrace{\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & & & 1 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix}}_{=:D}.$$

Für die Matrix D gilt $D^n = 0$. Mit Lemma 12.9 folgt für $k \ge n-1$ unter Benutzung der binomischen Formel

$$B^k \le (\lambda \cdot I_n + c \cdot D)^k = \sum_{i=0}^{k} \underbrace{\binom{k}{i}}_{:=\frac{k!}{i!(k-i)!}} c^i \lambda^{k-i} D^i = \sum_{i=0}^{n-1} \binom{k}{i} c^i \lambda^{k-i} D^i \xrightarrow{k\to\infty} 0.$$

Hieraus ergibt sich die Behauptung. □

Nun können wir den Hauptsatz dieses Kapitels beweisen, von dem (12.5) ein Spezialfall ist.

Satz 12.11. *Es sei $A \in \mathbb{R}^{n \times n}$ positiv und stochastisch. Dann gilt*

$$\lim_{k\to\infty} A^k = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}.$$

Dabei ist $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n_{\geq 0}$ *der eindeutig bestimmte Eigenvektor von A^T zum Eigenwert 1 mit $\alpha_1 + \cdots + \alpha_n = 1$.*

Beweis. Aus den Sätzen 12.6 und 11.17 und aus Lemma 12.8 erhalten wir die Existenz einer Matrix $S \in \mathrm{GL}_n(\mathbb{C})$, so dass

$$S^{-1}AS = \left(\begin{array}{c|cccc} 1 & 0 & \cdots & 0 \\ \hline 0 & \lambda_2 & & * \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_n \end{array}\right) \tag{12.7}$$

mit $\lambda_i \in \mathbb{C}$, $|\lambda_i| < 1$. Wir wählen eine obere Schranke $d \in \mathbb{R}$ für die Beträge aller in S und S^{-1} vorkommenden Einträge. Wenn wir mit E_n die $n \times n$-Matrix mit sämtlichen Einträgen 1 bezeichnen, erhalten wir also $S \leq d \cdot E_n$, und ebenso für S^{-1}. Es sei ε eine positive, reelle Zahl. Wegen Lemma 12.10 gibt es ein $m \in \mathbb{N}$, so dass

$$\begin{pmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}^k \leq \frac{\varepsilon}{n^2 d^2} E_{n-1}$$

für alle $k \geq m$ gilt. Also gilt für diese k:

$$\begin{aligned} A^k - S \cdot \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right) \cdot S^{-1} = \\ S \cdot \left(\left(\begin{array}{c|cccc} 1 & 0 & \cdots & 0 \\ \hline 0 & \lambda_2 & & * \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_n \end{array}\right)^k - \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right) \right) \cdot S^{-1} \leq \\ dE_n \cdot \frac{\varepsilon}{n^2 d^2} E_n \cdot dE_n = \frac{\varepsilon}{n^2} E_n^3 = \varepsilon E_n. \end{aligned}$$

Dies bedeutet

$$\lim_{k\to\infty} A^k = S \cdot \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right) \cdot S^{-1}.$$

Was ist das Produkt auf der rechten Seite der Gleichung? Wegen (12.7) steht in der ersten Spalte von S ein Eigenvektor zum Eigenwert 1 von A. Da A eine stochastische Matrix ist, ist der Vektor $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ ein Eigenvektor zum Eigenwert 1. Da wir außerdem wissen, dass der Eigenraum eindimensional ist, können wir annehmen, dass in der ersten Spalte von S dieser Vektor steht. Die erste Zeile von S^{-1} bezeichnen wir mit $(\alpha_1, \ldots, \alpha_n)$. Wir erhalten

$$S \cdot \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right) \cdot S^{-1} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{array}\right) \cdot S^{-1} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix},$$

und insgesamt folgt

$$\lim_{k\to\infty} A^k = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}.$$

Hieraus folgt mit Lemma 12.5, dass die α_i nicht-negativ reell mit Summe 1 sind, und auch, dass $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ ein Eigenvektor von A^T zum Eigenwert 1 ist, denn es gilt

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix} \cdot A = \left(\lim_{k\to\infty} A^k\right) \cdot A = \lim_{k\to\infty} A^{k+1} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}.$$

Schließlich folgt aus Satz 12.6(b), dass ein Eigenvektor zum Eigenwert 1 von A^T mit Koeffizientensumme 1 eindeutig bestimmt ist. □

Aus Satz 12.11 folgt, dass für jeden Zeilenvektor $v \in \mathbb{R}^{1\times n}$ mit der Koeffizientensumme 1 gilt:

$$\lim_{k\to\infty} \left(v \cdot A^k\right) = (\alpha_1, \ldots, \alpha_n).$$

Anders gesagt: indem man A immer wieder von links an v heranmultipliziert, nähert man sich einem Eigenvektor zum Eigenwert 1 von A^T an.

Wir kommen auf die Google-Matrix G und den Pagerank-Vektor zurück. Satz 12.11 und die obige Idee gibt uns einen Algorithmus, wie man den Pagerank-Vektor (also einen Eigenvektor zum Eigenwert 1 von G^T) näherungsweise berechnen kann. Der Algorithmus läuft wie folgt.

Algorithmus 12.12 (Google-Algorithmus)**.**

(1) Bilde die Weblink-Matrix, daraus die Google-Matrix (mit „geeigneter“ Wahl des Parameters α).
(2) Wähle einen Vektor $p \in \mathbb{R}^{1 \times n}$ mit Komponentensumme 1.
(3) Ersetze immer wieder p durch $p \cdot G$, bis sich hierdurch p kaum noch ändert. (Hierbei sollte ein Schwellenwert für die Änderung von p vorgegeben werden, dessen Unterschreitung ein Terminationskriterium liefert.)
(4) Verwende p als Pagerank-Vektor. Seine i-te Komponente beinhaltet also die Wichtigkeit der Seite P_i im Internet.

Wir bemerken noch, dass die komplette Matrix G nie abgespeichert werden muss, da sich aufgrund der einfachen Bauart von G die Multiplikation $p \cdot G$ sehr einfach gestaltet. Aber selbst wenn man die Multiplikation ohne Optimierung durchführt, kommt der Algorithmus mit $r \cdot O(n^2)$ *Floating Point* Operationen aus, wobei r die Anzahl der Schleifendurchläufe ist. Setzt man $r = 10$ und $n = 10^9$ an, so ergibt sich gegenüber der Berechnung eines Eigenvektors mit dem Gauß-Algorithmus eine Verbesserung um den Faktor von etwa 10^8. Dies ist äußerst grob geschätzt, aber es zeigt, dass die Idee von Google erst durch Satz 12.11 und den daraus resultierenden Algorithmus praktikabel wird.

Kapitel 13
Skalarprodukt

Wie immer steht K für einen Körper. Wir werden K bald als $K = \mathbb{R}$ festlegen.

Definition 13.1. *Für* $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ *und* $w = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in K^n$ *heißt*

$$\langle v, w \rangle := \sum_{i=1}^{n} x_i y_i (= v^T w) \in K$$

*das (Standard-)***Skalarprodukt** *von v und w. Achtung: Die Notation ist anfällig für Verwechselungen mit dem Erzeugnis!*

Die Vektoren v und w heißen **senkrecht** *(=* **orthogonal***) zueinander, falls $\langle v, w \rangle = 0$. Für einen Unterraum $U \subseteq K^n$ heißt*

$$U^{\perp} := \{ v \in K^n \mid \langle u, v \rangle = 0 \text{ für alle } u \in U \}$$

das **orthogonale Komplement** *(auch:* Senkrecht-Raum*) von U.*

Für $V = \mathbb{R}^n$ entspricht der Orthogonalitätsbegriff dem, was man von der euklidischen Geometrie gewohnt ist. Über anderen Körpern gibt es „kontraintuitive" Effekte. Zum Beispiel steht der Vektor $\left(\begin{smallmatrix} 1 \\ i \end{smallmatrix} \right) \in \mathbb{C}^2$ auf sich selbst senkrecht. Oft wird das in Definition 13.1 definierte Produkt nur für $K = \mathbb{R}$ als Skalarprodukt bezeichnet, während man über $K = \mathbb{C}$ ein anderes Produkt benutzt.

Wir fassen ein paar Eigenschaften des Skalarprodukts zusammen.

Proposition 13.2. *(a) Für alle $u, v, w \in K^n$ und $a \in K$ 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 $v, w \in K^n$ *gilt*

$$\langle v, w\rangle = \langle w, v\rangle.$$

(Man sagt auch, dass das Skalarprodukt symmetrisch *ist.)*

(c) Es gilt

$$(K^n)^\perp = \{0\},$$

d.h. nur der Nullvektor steht auf allen Vektoren senkrecht. (Man sagt auch, dass das Skalarprodukt nicht ausgeartet *ist.)*

Ab jetzt setzen wir $K = \mathbb{R}$ voraus. Für alle $v \in \mathbb{R}^n$ ist $\langle v, v\rangle$ nicht-negativ.

Definition 13.3. *Für* $v \in \mathbb{R}^n$ *heißt*

$$|\, v \,| := \sqrt{\langle v, v\rangle} \in \mathbb{R}_{\geq 0}$$

die **Länge** *von* v*. Man spricht auch von der* euklidischen Länge.

Für die Länge gelten folgende Regeln:

Proposition 13.4. *Für* $v, w \in \mathbb{R}^n$ *und* $a \in \mathbb{R}$ *gelten:*

(a) $|v + w| \leq |v| + |w|$ *(„Dreiecksungleichung“).*
(b) $|av| = |a| \cdot |v|$ *(wobei* $|a|$ *den üblichen Betrag der reellen Zahl* a *bezeichnet).*
(c) Falls $v \neq 0$*, so gilt* $|v| > 0$*.*
(d) $|\langle v, w\rangle| \leq |v| \cdot |w|$ *(„Schwarzsche Ungleichung“). Auch in diese Formel koexistieren der reelle Betrag und die Länge von Vektoren.*

Den Beweis lassen wir weg.

Definition 13.5. *Eine Menge* $S = \{v_1, \ldots, v_k\} \subset \mathbb{R}^n$ *heißt ein* **Orthonormalsystem***, falls* v_i *und* v_j *für* $i \neq j$ *orthogonal sind, und* $|v_i| = 1$ *für alle* i *gilt. In geraffter Schreibweise lautet die Bedingung:*

$$\langle v_i, v_j\rangle = \delta_{i,j} \quad \textit{mit} \quad \delta_{i,j} := \begin{cases} 1 & \textit{falls } i = j \\ 0 & \textit{sonst} \end{cases}$$

(das sogenannte Kronecker-Delta*). Wenn* $A \in \mathbb{R}^{n \times k}$ *die Matrix mit den* v_i *als Spalten ist, ergibt sich:*

$$S \textit{ Orthonormalsystem} \quad \Longleftrightarrow \quad A^T \cdot A = I_k.$$

Beispiel 13.6. (1) Die Standardbasis von $\mathbb{R}^n$ ist ein Orthonormalsystem.
(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 $\mathbb{R}^3$. ◁

Satz 13.7. *Jedes Orthonormalsystem in $\mathbb{R}^n$ ist linear unabhängig.*

Beweis. Sei $S = \{v_1, \ldots, v_k\} \subset \mathbb{R}^n$ ein Orthonormalsystem. Weiter sei

$$a_1 v_1 + \cdots + a_k v_k = 0$$

mit $a_i \in \mathbb{R}$. Für alle $j \in \{1, \ldots, k\}$ folgt durch Bildung des Skalarprodukts mit v_j:

$$0 = \langle v_j, 0 \rangle = \left\langle v_j, \sum_{i=1}^{k} a_i v_i \right\rangle = \sum_{i=1}^{k} a_i \langle v_j, v_i \rangle = a_j.$$

Also sind alle $a_j = 0$, und die lineare Unabhängigkeit ist bewiesen. □

Aus Satz 13.7 und Korollar 6.13(a) folgt:

Korollar 13.8. *Sei $U \subseteq \mathbb{R}^n$ ein Unterraum, $k := \dim(U)$, und $S = \{v_1, \ldots, v_k\} \subset U$ sei ein Orthonormalsystem. Dann ist S eine Basis von U. Man nennt S dann eine* **Orthonormalbasis** *von U.*

Beispiel 13.9. (1) Die Standardbasis ist eine Orthonormalbasis von $\mathbb{R}^n$.
(2) Die Vektoren

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

bilden eine Orthonormalbasis im $\mathbb{R}^2$. ◁

Besitzt jeder Unterraum des $\mathbb{R}^n$ eine Orthonormalbasis? Diese Frage werden wir konstruktiv durch das Gram-Schmidtsche Orthogonalisierungsverfahren beantworten.

Algorithmus 13.10 (Gram-Schmidtsches Orthogonalisierungsverfahren)**.**

Eingabe: Ein Unterraum $U = \langle v_1, \ldots, v_k \rangle \subseteq \mathbb{R}^n$ (erzeugt von Vektoren v_i).
Ausgabe: Eine Orthonormalbasis $\{u_1, \ldots, u_m\}$ von U.

(1) Setze $m := 0$.
(2) Für $i = 1, \ldots, k$ führe Schritte (3) und (4) aus.
(3) Setze

$$w_i := v_i - \sum_{j=1}^{m} \langle u_j, v_i \rangle \cdot u_j. \tag{13.1}$$

(Im Fall $m = 0$ bedeutet dies $w_i := v_i$.)
(4) Falls $w_i \neq 0$, setze $m := m + 1$ und

$$u_m := \frac{1}{|w_i|} \cdot w_i.$$

Wir werden die Korrektheit des Algorithmus nicht ganz streng nachweisen. Wir bemerken jedoch, dass aus (13.1) folgt, dass für $l \leq m$

$$\langle u_l, w_i \rangle = \langle u_l, v_i \rangle - \sum_{j=1}^{m} \langle u_j, v_i \rangle \cdot \langle u_l, u_j \rangle = \langle u_l, v_i \rangle - \langle u_l, v_i \rangle = 0$$

gilt (wobei wir als Induktionsvoraussetzung $\langle u_l, u_j \rangle = \delta_{j,l}$ benutzt haben). Also steht w_i auf allen bisherigen u_l senkrecht. Außerdem folgt aus (13.1), dass

$$\langle u_1, \ldots, u_m, w_i \rangle = \langle u_1, \ldots, u_m, v_i \rangle = \langle v_1, \ldots, v_i \rangle$$

(letztere Gleichheit wieder unter Verwendung einer Induktionsannahme). Falls $w_i = 0$, so folgt $\langle v_1, \ldots, v_i \rangle = \langle u_1, \ldots, u_m \rangle$. Falls $w_i \neq 0$, so wird $\{u_1, \ldots, u_{m+1}\}$ ein Orthonormalsystem. Zum Schluss haben wir $\langle u_1, \ldots, u_m \rangle = U$, also bilden die u_i nach Satz 13.7 eine Orthonormalbasis.

Beispiel 13.11. Wir wollen Algorithmus 13.10 auf

$$U := \left\langle \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right\rangle \subseteq \mathbb{R}^3$$

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 U. ◁

Aus der Korrektheit von Algorithmus 13.10 folgt:

Satz 13.12. *Jeder Unterraum von $\mathbb{R}^n$ hat eine Orthonormalbasis.*

Wir schließen das Kapitel mit einer Definition ab.

Definition 13.13. *Eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$ heißt* **orthogonal**, *falls*

$$A^T \cdot A = I_n.$$

(Äquivalent hierzu: $A^{-1} = A^T$, oder die Spalten von A bilden eine Orthonormalbasis von $\mathbb{R}^n$.)

Weiter heißen

$$\mathrm{O}_n(\mathbb{R}) := \left\{A \in \mathbb{R}^{n\times n} \mid A^T \cdot A = I_n\right\}$$

die **orthogonale Gruppe** *und*

$$\mathrm{SO}_n(\mathbb{R}) := \mathrm{O}_n(\mathbb{R}) \cap \mathrm{SL}_n(\mathbb{R})$$

die **spezielle orthogonale Gruppe**.

Beides sind Gruppen mit dem Matrizenprodukt, denn für $A, B \in \mathrm{O}_n(\mathbb{R})$ gilt

$$(AB)^T \cdot (AB) = B^T A^T AB = I_n$$

und

$$(A^{-1})^T A^{-1} = (A^T)^T A^{-1} = AA^{-1} = I_n.$$

Für $A \in \mathrm{O}_n(\mathbb{R})$ und $v, w \in \mathbb{R}^n$ gilt

$$\langle Av, Aw\rangle = (Av)^T Aw = v^T A^T Aw = v^T w = \langle v, w\rangle,$$

also erhält die lineare Abbildung φ_A Skalarprodukte. Insbesondere folgt $|A \cdot v| = |v|$, also ist φ_A auch längenerhaltend. Abbildungen mit dieser Eigenschaft heißen *Isometrien*.

Kapitel 14
Symmetrische Matrizen

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 14.1. *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

$$\overline{v}^T \cdot v = \sum_{i=1}^{n} \overline{x_i} x_i = \sum_{i=1}^{n} |x_i|^2 \neq 0$$

folgt $\overline{\lambda} = \lambda$, also $\lambda \in \mathbb{R}$. $\square$

Lemma 14.2. *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 $B = \{b_1, \ldots, b_k\}$ von U sei

$$C = (c_{i,j}) = D_B(\varphi) \in \mathbb{R}^{k \times k} \tag{14.1}$$

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

$$C \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}. \tag{14.2}$$

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

$$A \cdot v = \sum_{i=1}^k x_i A b_i = \sum_{i=1}^k x_i \varphi(b_i) \underset{(14.1)}{=} \sum_{i=1}^k x_i \left(\sum_{j=1}^k c_{j,i} b_j \right) =$$
$$\sum_{j=1}^k \left(\sum_{i=1}^k c_{j,i} x_i \right) b_j \underset{(14.2)}{=} \sum_{j=1}^k \lambda x_j b_j = \lambda v,$$

also ist λ ein Eigenwert von A mit Eigenvektor v. Wegen Lemma 14.1 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. $\square$

Lemma 14.3. *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 14.4 („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 Satz 13.12 eine Orthonormalbasis B_i von E_{λ_i}. Wir setzen

$$B := B_1 \cup \dots \cup B_r.$$

Wegen Lemma 14.3 ist B ein Orthonormalsystem, wegen Satz 13.7 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 14.2 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 14.5. (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 $\operatorname{diag}(1,1,4)$ ähnlich ist. Wir wollen eine orthogonale Transformationsmatrix ausrechnen. Hierfür müssen wir die Eigenräume bestimmen. Der Eigenraum E_1 zum Eigenwert 1 ergibt sich als Lösungsraum 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 Gram-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 14.4 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 14.4 also nicht korrekt. Ebenso verhält es sich mit $\mathbb{Q}$ statt $\mathbb{R}$. ◁

Satz 14.4 hat beispielsweise physikalische Anwendungen. Zu einem starren Körper betrachtet man den sogenannten *Trägheitstensor*. Dieser ist eine Matrix $I \in \mathbb{R}^{3\times 3}$, die die Winkelgeschwindigkeit (als Vektor) mit dem Drehimpuls verbindet, ähnlich wie die Masse die Geschwindigkeit mit dem Impuls verbindet. Es stellt sich heraus, dass I symmetrisch ist. Also liefert Satz 14.4, 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*. Wegen des Drehimpulserhaltungssatzes bedeutet dies, dass Drehungen um die Hauptträgheitsachsen „schlingerfrei“ möglich sind. Bei konstantem Drehimpuls ist eine Drehung um die Achse mit dem größten Eigenwert (= *Hauptträgheitsmoment*) die energetisch günstigste und daher stabilste.

Anmerkung. Satz 14.4 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 $S^{-1}BS$ eine Diagonalmatrix ist. ◁

Aufgrund von Satz 14.4 wird folgende Definition sinnvoll.

Definition 14.6. *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 14.7. *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 14.4 haben wir $S \in \mathrm{O}_n(\mathbb{R})$, so dass

$$S^T AS = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D$$

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 $\left(\begin{smallmatrix} x_1 \\ \vdots \\ x_n \end{smallmatrix}\right) := S^T \cdot v$ 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

$$\langle v, A \cdot v\rangle = v^T SDS^T v = (x_1, \ldots, x_n) D \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \lambda_i x_i^2.$$

Hieraus folgen alle Behauptungen. □

Beispiel 14.8. 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 14.7 zur Feststellung der Definitheitseigenschaften von A an. Für $v = \left(\begin{smallmatrix} x_1 \\ \vdots \\ x_4 \end{smallmatrix}\right) \in \mathbb{R}^4$ 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. ◁

Kapitel 15
Anwendungen in der Graphentheorie

Zur Erinnerung (und Festlegung der Notation): Ein **Graph** ist ein Paar $G = (V, E)$ mit V einer endlichen, nicht leeren Menge von „Knoten“ (auch *Ecken* genannt, englisch *vertices*) und E einer Menge

$$E \subseteq \{\{u, v\} \mid u, v \in V,\ u \neq v\}$$

von „Kanten“ (englisch *edges*). (Es gibt diverse Varianten des Graphbegriffs: gerichtete Graphen, Multigraphen, gewichtete Graphen, Hypergraphen … Der Einfachheit halber beschränken wir uns hier auf Graphen im oben genannten Sinne.) Zwei Graphen $G = (V, E)$ und $G' = (V', E')$ heißen **isomorph**, falls es eine Bijektion $f\colon V \to V'$ gibt, so dass

$$\{\{f(u), f(v)\} \mid \{u, v\} \in E\} = E'.$$

Definition 15.1. *Es sei $G = (V, E)$ ein Graph mit $V = \{u_1, \ldots, u_n\}$. Wir definieren*

$$g_{i,j} := \begin{cases} 1 & \textit{falls } \{i, j\} \in E \\ 0 & \textit{sonst} \end{cases} \qquad \textit{und} \quad A := (g_{i,j}) \in \mathbb{R}^{n \times n}.$$

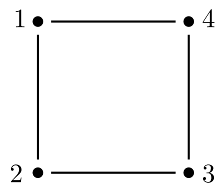
A heißt die **Adjazenzmatrix** *von G. Die Menge der Eigenwerte von A (gezählt mit Vielfachheiten) ist das* **Spektrum** *von G.*

Aus der Definition ist klar, dass die Adjazenzmatrix symmetrisch ist. Daher sind wegen Satz 14.4 alle Eigenwerte reell, und die algebraischen und geometrischen Vielfachheiten stimmen überein. Da das Spektrum eine Menge mit Vielfachheiten ist, ist es zweckmäßig, die Eigenwerte als der Größe nach geordnete Liste anzugeben.

Beispiel 15.2. Der Graph G mit

$$V = \{1, 2, 3, 4\} \quad \text{und} \quad E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 4\}\}$$

wird wie folgt gezeichnet:



Die Adjazenzmatrix ist

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

Deren charakteristisches Polynom ergibt sich nach kurzer Rechnung zu

$$\chi_A = \det \begin{pmatrix} x & -1 & 0 & -1 \\ -1 & x & -1 & 0 \\ 0 & -1 & x & -1 \\ -1 & 0 & -1 & x \end{pmatrix} = x^4 - 4x^2.$$

Als Spektrum bekommen wir $-2, 0, 0, 2$. ◁

Das Interesse am Spektrum eines Graphen ist durch folgenden Satz begründet.

Satz 15.3. *Die Spektren isomorpher Graphen stimmen überein.*

Beweis. Es seien $A = (g_{i,j})$ und $A' = (g'_{i,j}) \in \mathbb{R}^{n \times n}$ die Adjazenzmatrizen zweier isomorpher Graphen. Die Isomorphie bedeutet, dass es $\sigma \in S_n$ gibt mit

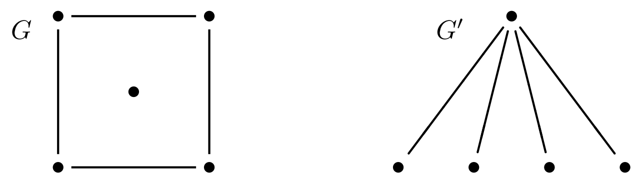
$$g'_{i,j} = g_{\sigma(i),\sigma(j)}.$$

Also geht A' aus A hervor, indem die Permuation σ auf die Zeilen und auf die Spalten angewandt wird. Ebenso geht die Matrix $(x \cdot I_n - A') \in \mathbb{R}[x]^{n \times n}$ aus $x \cdot I_n - A$ durch Permutation der Zeilen und Spalten mit σ hervor. Aus Lemma 10.6(b) folgt $\chi_{A'} = \chi_A$, also stimmen die Spektren überein. □

Man drückt Satz 15.3 auch aus, indem man sagt, dass das Spektrum eine Graph-Invariante ist. In analoger Sprechweise könnte man auch sagen, dass die Dimension eine Invariante eines Vektorraums ist, oder die Ordnung eine Invariante einer Gruppe. Die Adjezenzmatrix selbst ist aber keine Graph-Invariante.

Gilt auch die Umkehrung von Satz 15.3? Werden also Graphen bis auf Isomorphie durch ihr Spektrum bestimmt? Wie das folgende Beispiel zeigt, ist dies leider nicht der Fall.

Beispiel 15.4. Die Graphen G und G', gegeben durch



(bei G ist der in der Mitte gezeichnete Punkt mit keinem verbunden), haben beide das Spektrum $-2, 0, 0, 0, 2$. Sie sind aber nicht isomorph. (Dies kann man z.B. daran sehen, dass G' zusammenhängend ist, G aber nicht.) $\triangleleft$

Zwei Graphen mit demselben Spektrum nennt man *isospektral*. Wir führen nun eine Variante des Spektrums ein.

Definition 15.5. *Es sei G ein Graph mit Adjazenzmatrix $A = (g_{i,j}) \in \mathbb{R}^{n \times n}$. Für $i = 1, \ldots, n$ setzen wir*

$$d_i := \sum_{j=1}^{n} g_{i,j} \quad (= \textit{die Anzahl der vom } i\textit{-ten Knoten ausgehenden Kanten}),$$

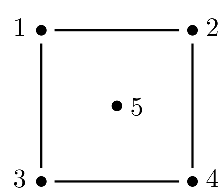
und nennen dies den **Grad** *des i-ten Knotens. Wir bilden die Matrix*

$$L = (l_{i,j}) \in \mathbb{R}^{n \times n} \quad \textit{mit} \quad l_{i,j} = \begin{cases} -g_{i,j} & \textit{falls } i \neq j \\ d_i & \textit{falls } i = j \end{cases}.$$

L heißt die **Laplace-Matrix** *von G. Die Menge der Eigenwerte von L (gezählt mit Vielfachheiten) ist das* **Laplace-Spektrum** *von G.*

Da auch L symmetrisch ist, sind die Eigenwerte reell. Außerdem haben isomorphe Graphen identische Laplace-Spektren. Dies beweist man genau so wie Satz 15.3.

Beispiel 15.6. (a) Wenn wir die Knoten des Graphen G aus Beispiel 15.4 wie folgt nummerieren,

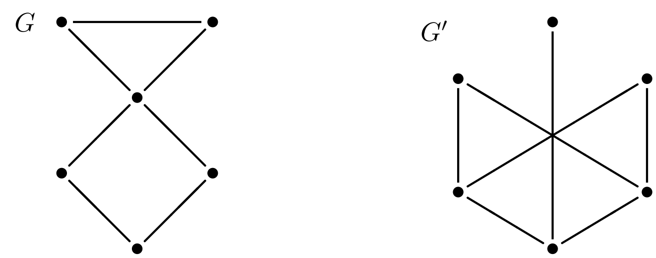


so ergibt sich die Laplace-Matrix

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

Rechnung liefert, dass das Laplace-Spektrum 0,0,2,2,4 ist. Der Graph G' aus Beispiel 15.4 hat im Gegensatz dazu das Laplace-Spektrum 0,1,1,1,5. Diese beiden Graphen lassen sich also durch ihre Laplace-Spektren trennen! Wir sehen also, dass das Laplace-Spektrum eine neue Invariante ist, die weitere Informationen liefert.

(b) Nun betrachten wir die folgenden Graphen G und G':



Aufstellen der Laplace-Matrizen und Berechnen der Eigenwerte ergibt, dass G und G' *beide* das Laplace-Spektrum

$$0, 3-\sqrt{5}, 2, 3, 3, 3+\sqrt{5}$$

haben. G und G' sind aber nicht isomorph. Dies kann man z.B. daran sehen, dass G' einen Knoten von Grad 1 enthält, G aber nicht. $\triangleleft$

Man kann auch Beispiele nicht isomorpher Graphen finden, bei denen das Spektrum und das Laplace-Spektrum übereinstimmen.

Satz 15.7. *Die Laplace-Matrix eines Graphen ist positiv semidefinit. Das Laplace-Spektrum besteht also aus lauter nicht-negativen Zahlen.*

Beweis. Es sei $A = (g_{i,j}) \in \mathbb{R}^{n\times n}$ die Adjazenzmatrix eines Graphen. Wir benutzen Satz 14.7. Für $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ gilt

$$\begin{aligned}\langle v, L\cdot v\rangle = \sum_{i,j=1}^{n} x_i l_{i,j} x_j = \sum_{i=1}^{n} d_i x_i^2 - \sum_{i\neq j} g_{i,j} x_i x_j = \\ \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} g_{i,j} x_i^2 - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} g_{i,j} x_i x_j = \sum_{1\le i<j\le n} g_{i,j}\left(x_i^2 + x_j^2 - 2x_i x_j\right) = \\ \sum_{1\le i<j\le n} g_{i,j}(x_i - x_j)^2 \ge 0. \quad (15.1)\end{aligned}$$

$\square$

Indem wir den obigen Beweis nochmal anschauen und analysieren, für welche Vektoren $v \in \mathbb{R}^n$ die Gleichung $\langle v, L \cdot v \rangle = 0$ gilt, erhalten wir einen interessanten Zusatz.

Satz 15.8. *Die Anzahl der Zusammenhangskomponenten eines Graphen G ist die Vielfachheit des Eigenwertes 0 im Laplace-Spektrum.*

Beweis. Für welche Vektoren $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ gilt $\langle v, L \cdot v \rangle = 0$? Wegen (15.1) muss $x_i = x_j$ für alle i, j mit $g_{i,j} = 1$ gelten. Wegen der Transitivität der Gleichheitsbeziehung gilt dann auch automatisch $x_i = x_j$, wenn i und j in derselben Zusammenhangskomponente von G liegen. Umgekehrt kann man für jede Zusammenhangskomponente Z_k eine Zahl $\alpha_k \in \mathbb{R}$ wählen und dann für alle Knoten $i \in Z_k$ $x_i := \alpha_k$ setzen. So erhält man einen Vektor v mit $\langle v, L \cdot v \rangle = 0$. Wir fassen zusammen: Mit

$$E_0 := \{v \in \mathbb{R}^n \mid \langle v, L \cdot v \rangle = 0\}$$

gilt

$$\dim(E_0) = \text{Anzahl der Zusammenhangskomponenten.} \tag{15.2}$$

Warum ist $\dim(E_0)$ die Vielfachheit des Eigenwertes 0 von L? Wegen Satz 14.4 gibt es eine Orthonormalbasis $\{v_1, \ldots, v_n\}$ aus Eigenvektoren. Also $L \cdot v_i = \lambda_i v_i$ mit $\lambda_i \geq 0$ wegen Satz 15.7. Durch Umordnen können wir $\lambda_1 = \cdots = \lambda_l = 0$ und $\lambda_i > 0$ für $i > l$ erreichen. Für $v = \sum_{i=1}^n y_i v_i \in \mathbb{R}^n$ folgt

$$\langle v, L \cdot v \rangle = \sum_{i,j=1}^n y_i \lambda_j y_j \langle v_i, v_j \rangle = \sum_{i=1}^n \lambda_i y_i^2,$$

also $v \in E_0$ genau dann, wenn $y_{l+1} = \cdots = y_n = 0$. Dies ergibt

$$\dim(E_0) = l = \text{Vielfachheit des Eigenwertes 0 von } L.$$

Mit (15.2) folgt die Behauptung. □

Notation

Index