JOCHEN HEINLOTH

LINEARE ALGEBRA 2

Inhaltsverzeichnis

Einleitung

Diese Notizen sind nicht als Skript gedacht, sondern eher als Hilfestellung zur Erinnerung an die Vorlesung. Während des Semesters ist es schwierig, zusätzlich zur Vorlesung ein Buch zu schreiben – Sie können gerne einmal versuchen selbst eine Vorlesung im Computer aufzuschreiben, dann sehen Sie vielleicht, was ich meine – daher werden Sie hier auch mehr Tippfehler finden als mir lieb ist. Hinweise zu Fehlern und Tippfehlern nehme ich gerne entgegen, zum Beispiel per email oder das Moodle-Forum.

Mein Dank an alle Studierenden und Tutor:innen, die mir Kommentare, Fehler und Verbesserungsvorschläge geschickt haben: Fereshteh Fattahi, Kevin Kristen, Florian Leptien, Natascha Scheibke, Christian Willeke, Carla Wussow und ebenfalls vielen Dank an die anonymen Hinweisgeber:innen auf Moodle.

Als Literaturquellen können Sie die gleichen Referenzen wie für die Lineare Algebra 1 verwenden: Viele Bücher zur linearen Algebra sind über die Universitätsbibliothek auch online verfügbar, ein Beispiel ist das Buch von Gerd Fischer[1]*, es gibt auch viele Skripte wie zum Beispiel das von Ulrich Görtz*[2] *oder das von Wolfgang Soergel*[3]*, englische Quellen wie das Buch von David Lay, Steven Lay und Judi McDonald*[4] *sind häufig sehr viel anwendungsorientierter und darum eine gute Ergänzung. Sie werden sehen, dass die Quellen fast den gleichen Inhalte umfassen, aber unterschiedlich erklären, welche Darstellung Ihnen persönlich am leichtesten zugänglich ist, wird von Ihren Vorlieben und Vorkenntnissen abhängen.*

[1] Gerd Fischer. *Lineare Algebra*, volume 17 of *Grundkurs Mathematik*. Friedr. Vieweg & Sohn, Braunschweig, fifth edition, 1979. ISBN 3-528-17217-7. URL `https://link.springer.com/book/10.1007/978-3-8348-9365-9`. In collaboration with Richard Schimpl

[2] Ulrich Görtz. Lineare algebra. Vorlesungsskript, 2020. URL `https://math.ug/lecture-notes.html`

[3] Wolfgang Soergel. Lineare algebra. Vorlesungsskript, 2022. URL `http://home.mathematik.uni-freiburg.de/soergel/Skripten/XXLA1.pdf`

[4] David Lay, Steven Lay, and Judi McDonald. *Linear Algebra and Its Applications*. Harlow: Pearson Education, Limited, 2015. ISBN 9781292092232. URL `https://elibrary.pearson.de/book/99.150005/9781292092249`

Diese Notizen schließen an die Vorlesung Lineare Algebra 1 an, die wir mit der Hauptraumzerlegung für Endomorphismen und einem Ausblick wie diese ermöglicht zu zeigen, dass wir für große Matrizen oftmals durch Berechnung von $A^n\mathbf{v}$ Eigenvektoren näherungsweise bestimmen können, abgeschlossen hatten. Wir hatten dort auch gesehen, dass dies zum Beispiel für die Bewertung von Knoten in großen Netzwerken verwendet wird.

In diesem Semester werden wir dieses Thema zunächst aufgreifen und sehen, wie wir auch für Matrizen (oder lineare Abbildungen) die nicht diagonalisierbar sind, Basen finden können, mit denen wir die Abbildung geometrisch verstehen können. Das wird uns nur noch wenige Vorlesungen beschäftigen.

Danach wechseln wir das Thema. Wir werden dann versuchen

Vektorräume geometrischer zu verstehen, indem wir einen Längen- und Abstandsbegriff für Vektoren finden und damit auch Vorstellungen wie „senkrecht“ mathematisch genau fassen. Das ist nicht nur für die Anschauung praktisch, sondern ermöglicht auch alltägliche Anwendungen wie GPS oder das mp3 Musikformat, aber dazu später mehr.

ZUM ENDE DER VORLESUNG kommen wir dann noch einmal auf die Grundlagen zurück, die wir, nachdem wir uns besser an die mathematische Sprache gewöhnt haben, dann noch einmal von einem abstraken Standpunkt neu verstehen können, zum Beispiel können wir die Aussage „Zeilenrang=Spaltenrang“ damit ohne Rechnung neu einsehen. Als Bonus können wir mit diesen Methoden dann auch verstehen, was die reellen Zahlen wirklich sind und wieso wir damit überhaupt rechnen können, obwohl Sie gar keine Rechenregeln für die Addition von unendlichen Dezimalzahlen kennen. Ebenso können wir vom abstrakteren Standpunkt dem Rechnen mit Buchstaben und Symbolen einen ganz konkreten, formalen Rahmen geben. Das führt und dann auch zu einem erstaunlicherweise offenen Problem: Wir wissen noch immer nicht genau, was die schnellste Methode ist, um zwei $n \times n$-Matrizen miteinander zu multiplizieren, schlimmer noch, ist diese Frage schon für 3×3-Matrizen nicht endgültig geklärt. Für sehr viele moderne Algorithmen wäre es sehr nützlich, dafür schnellere, oder weniger Energieintensive Verfahren zu finden.

Die Jordannormalform – gute Basen für lineare Abbildungen

Erinnerung an die Hauptraumzerlegung aus dem vorigen Semester

Im letzten Semester hatten wir angefangen uns zu überlegen, wie wir für eine lineare Abbildung $A\colon V \to V$ eine angepasste Basis von V finden können, so dass die Matrix von A eine besonders einfache Gestalt bekommt. Das ist einerseits nützlich, um eine geometrische Vorstellung der Abbildung zu bekommen und andererseits erlaubt es uns, die Potenzen von A abzulesen und damit das Langzeitverhalten von $A^n\mathbf{v}$ für große n zu verstehen. Als Anwendung der geometrischen Vorstellung hatten wir gesehen, wie das zur Bewertung von Knotenpunkten in Netzwerken verwendet werden kann.

Unser erster Ansatz hierzu war wie folgt:

1. Bestimme die Eigenwerte von A, das sind die Nullstellen des charakteristischen Polynoms $\det(A - t\,\mathrm{id})$.

2. Bestimme für jeden Eigenwert c eine Basis des Eigenraums
$$\mathrm{Eig}(A, c) = \mathrm{Ker}(A - c \cdot \mathrm{id}) = \{\mathbf{v} \in V \mid A\mathbf{v} = c \cdot \mathbf{v}\}.$$

Wenn die Dimension der Eigenräume jeweils die Ordnung der Nullstelle des entsprechenden Eigenwertes im charakteristischen Polynom ist, bekommen wir so eine Basis von V bezüglich der A eine Diagonalmatrix ist.

Wenn das nicht funktioniert, d.h. falls die Dimension der Eigenräume zu klein ist, hatten wir stattdessen die Hauptraumzerlegung gefunden:

Satz (Hauptraumzerlegung aus Lineare Algebra 1)**.** *Sei $A\colon V \to V$ eine lineare Abbildung mit charakteristischem Polynom*
$$\det(A - t\,\mathrm{id}_V) = \prod_{i=1}^{r} (c_i - t)^{n_i}$$
wobei $c_1, \ldots, c_r \in K$ paarweise verschieden sind. Dann gilt
$$\bigoplus_{i=1}^{r} \mathrm{Ker}((A - c_i\,\mathrm{id}_V)^n) = V$$

Das hatten wir uns induktiv überlegt, indem wir zunächst für den Fall $c_1 = 0$ die Kette der Kerne

$$\operatorname{Ker}(A) \subseteq \operatorname{Ker}(A^2) \subseteq \ldots \operatorname{Ker}(A^n)$$

und der Bilder

$$\operatorname{Bild}(A) \supseteq \operatorname{Bild}(A^2) \supseteq \cdots \supseteq \operatorname{Bild}(A^n)$$

angeschaut hatten und gesehen hatten, dass in beiden Ketten die Teilmengen ab einem Punkt $=$ gelten muss.

An diesem Punkt bildet A das $\operatorname{Bild}(A^m)$ isomorph auf sich selbst ab, hat auf diesem Unterraum also nicht mehr den Eigenwert 0 hingegen ist auf $\operatorname{Ker}(A^m)$ der einzige Eigenwert 0.

Das konnten wir dann induktiv für die Abbildung $A - c_i \operatorname{id}_V$ anwenden um die Hauptraumzerlegung zu erhalten.

Wir sollten uns nun noch überlegen, wie wir für die Haupträume gute Basen finden können, wenn es nicht genügend Eigenvektoren gibt, d.h. falls

$$\operatorname{Eig}(A, c) \subsetneq \operatorname{Ker}((A - c \operatorname{id}_V)^n).$$

Um das zu verstehen, würde ich mit zunächst den Fall anschauen, dass es überhaupt nur einen Eigenwert gibt, d.h.

$$\det(A - t \cdot \operatorname{id}_V) = (c - t)^n.$$

Wie bei der Hauptraumzerlegung ist es dann gut, zunächst den Fall 0 zu verstehen, also den Fall, dass $\operatorname{Ker}(A^n) = V$, also $A^n = 0$ gilt. So eine Abbildung ist uns schon begegnet.

Die Strategie, zunächst Fälle zu betrachten, in denen möglichst wenige zusätzliche Komplikationen auftauchen kennen Sie mittlerweile.

Beispiel 1. Die Ableitung definiert eine lineare Abbildung

$$\begin{aligned} D\colon \mathbb{Q}[t]_{\leq 3} &\to \mathbb{Q}[t]_{\leq 3} \\ p(t) &\mapsto D(p) := p'(t) \end{aligned}$$

für die $D^4 = 0$ gilt, aber $D^3 \neq 0$. Da $\operatorname{Ker}(D) = \{p(t) = c \text{ konstant}\}$, ist der Eigenraum zum Eigenwert 0 hier eindimensional.

Die Matrix von D bezügliche der Basis $\mathbf{v}_1 = 1, \mathbf{v}_2 = t, \mathbf{v}_3 = t^2, \mathbf{v}_4 = t^3$ ist:

$$_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Die Einträge $3, 2, 1$ machen es noch etwas unhandlich, die Potenzen von D auszurechnen, probieren Sie das einmal aus.

Übersichtlicher wird die Matrix wenn wir stattdessen die Basis

$$\begin{aligned} \mathbf{w}_4 &= t^3, \\ \mathbf{w}_3 &= D\mathbf{w}_4 = 3t^2, \\ \mathbf{w}_2 &= D^2\mathbf{w}_4 = D\mathbf{w}_3 = 6t, \\ \mathbf{w}_1 &= D^3\mathbf{w}_4 = D\mathbf{w}_2 = 6 \end{aligned}$$

wählen, dann bildet D die Basisvektoren auf sich ab:

$$\mathbf{w}_4 \xrightarrow{D} \mathbf{w}_3 \xrightarrow{D} \mathbf{w}_2 \xrightarrow{D} \mathbf{w}_1 \xrightarrow{D} 0.$$

Das bedeutet

$$_{\underline{\mathbf{w}}}\mathrm{Mat}_{\underline{\mathbf{w}}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

und also

$$_{\underline{\mathbf{w}}}\mathrm{Mat}_{\underline{\mathbf{w}}}(D^2) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, {}_{\underline{\mathbf{w}}}\mathrm{Mat}_{\underline{\mathbf{w}}}(D^3) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, D^4 = 0,$$

die 1sen rutschen also bei den Potenzen jeweils eine Nebendiagonale weiter nach oben, denn

$$\mathbf{w}_4 \xrightarrow{D^2} \mathbf{w}_2 \xrightarrow{D^2} 0,$$

$$\mathbf{w}_3 \xrightarrow{D^2} \mathbf{w}_1 \xrightarrow{D} 0$$

und genauso für D^3.

Definition 2. Eine lineare Abbildung $A\colon V \to V$ für die es ein n gibt so dass $A^n = \underbrace{A \circ A \circ \cdots \circ A}_{n \text{ Faktoren}} = 0$ gilt heißt *nilpotent*.

Frage. *Was passiert im Beispiel der Ableitung, wenn wir statt $\mathbb{Q}$ den Körper $K = \mathbb{Z}/3\mathbb{Z}$ oder $K = \mathbb{Z}/2\mathbb{Z}$ verwenden? Was passiert wenn wir Polynome höheren Grades $K[t]_{\leq d}$ betrachten?*

Das hatten wir in der Vorlesung ausprobiert. In $\mathbb{Z}/2\mathbb{Z}[t]_{\leq 3}$ erhalten wir für die Basis $1, t, t^2, t^3$ die Bilder

$$D(t^3) = 3t^2 = t^2, D(t^2) = 2t = 0, D(t) = 1, D(1) = 0$$

d.h. $t^3 \mapsto t^2 \mapsto 0, t \mapsto 1 \mapsto 0$ und damit

$$_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In $\mathbb{Z}/3\mathbb{Z}[t]_{\leq 3}$ erhalten wir für die Basis $1, t, t^2, t^3$ die Bilder

$$D(t^3) = 3t^2 = 0 D(t^2) = 2t, D(2t) = 2, D(2) = 0$$

d.h. $t^3 \mapsto 0, t^2 \mapsto 2t \mapsto 2 \mapsto 0$ und damit für die Basis $\mathbf{w}_4 = t^3, \mathbf{w}_3 = t^2, \mathbf{w}_2 = 2t, \mathbf{w}_1 = 2$

$$_{\underline{\mathbf{w}}}\mathrm{Mat}_{\underline{\mathbf{w}}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

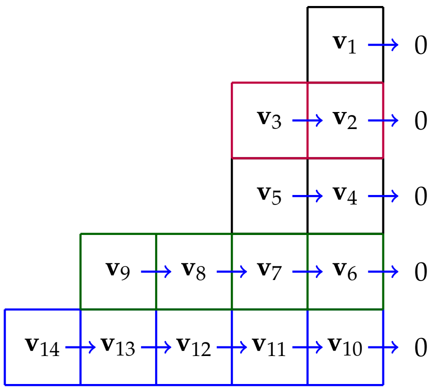
Beispiel 3. Können Sie für die Matrizen

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

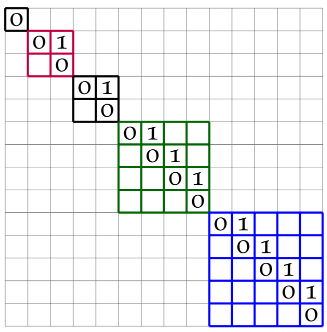
auch Basen finden, so dass wir die Potenzen der entsprechenden Matrizen sofort ablesen können?

Die Normalform für nilpotente Abbildungen

Aus den Beispielen könnten wir die Hoffnung bekommen, dass wir vielleicht für jede nilpotente Abbildung $N \colon V \to V$ Basen finden können, so dass die Abbildungsvorschrift durch Ketten wie zuvor gegeben ist:



Die zugehörige Matrix ist dann eine Blockmatrix



.

Notation 4. Eine $r \times r$ Matrix der Form

$$J_r(c) = \begin{pmatrix} c & 1 & 0 & \cdots & 0 \\ & c & 1 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & c \end{pmatrix}$$

Für den Moment sind uns nur Beispiele mit $c = 0$ begegnet, aber Sie können vielleicht raten, warum wir die allgemeinere Notation einführen.

heißt *Jordanblock der Länge r zum Eigenwert c.*

Bemerkung. Sowohl an der Blockmatrix, als auch am Diagramm der Abbildungsvorschrift können wir sofort die Dimensionen der Kerne $\dim(\mathrm{Ker}(N), \dim(\mathrm{Ker}(N^2)), \ldots$ sofort ablesen:

1. Im Diagramm bilden die Vektoren in der letzten Spalte eine Basis von $\mathrm{Ker}(N)$, die Vektoren in den letzten beiden Spalten eine Basis von $\mathrm{Ker}(N^2)$, usw.

2. In der Blockmatrix ist die Anzahl der Böcke, die Dimension von $\mathrm{Ker}(N)$, $\dim(\mathrm{Ker}(N^2) = \#\{\text{Blöcke}\} + \#\{\text{Blöcke der Länge } \geq 2\}$, usw.

Eine ähnliche Beschreibung können Sie für die Dimensionen der Bilder $\dim(\mathrm{Bild}(N), \dim(\mathrm{Bild}(N^2)), \ldots$ selbst überlegen.

Für eine solche Beschreibung von N suchen wir also eine Basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ von V, so dass

1. Für jedes $\mathbf{v}_i \in B$ entweder $N\mathbf{v}_i = 0$ oder $N\mathbf{v}_i = \mathbf{v}_j$ für ein $\mathbf{v}_j \in B$ gilt.

2. (Kettenbedingung) Für jedes $\mathbf{v}_j$ existiert höchstens ein $\mathbf{v}_i \in B$ so dass $N\mathbf{v}_i = \mathbf{v}_j$.

Solche Basen können wir tatsächlich induktiv immer finden.

Satz 5 (Jordanbasen für nilpotente Abbildungen)**.** *Für jede nilpotente Abbildung*

$$N \colon V \to V$$

eines n-dimensionalen K-Vektorraums V existiert eine Basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ von V, so dass

1. *für jedes $\mathbf{v}_i \in B$ entweder $N\mathbf{v}_i = 0$ oder $N\mathbf{v}_i = \mathbf{v}_j$ für ein $\mathbf{v}_j \in B$ gilt und*

2. *(Kettenbedingung) für jedes $\mathbf{v}_j \in B$ höchstens ein $\mathbf{v}_i \in B$ existiert, mit $N\mathbf{v}_i = \mathbf{v}_j$.*

Bei diesem Beweis passiert ein kleines Wunder. Wenn ich mich ersteinmal dafür entschieden habe, einen Induktionsbeweis zu führen, bleibt mir in jedem Schritt wenig Auswahl. Würden Sie einfach so versuchen, eine Basis zu basteln würden Sie wahrscheinlich schnell viele Fallunterscheidungen benötigen. Es ist eine gute Idee, das einmal zu probieren.

Beweis. Wir beweisen die Aussage mittels Induktion über die kleinste Zahl m für die $N^m = 0$ gilt. Für $m = 1$ gilt $N = 0$, d.h. N ist die 0-Abbildung. In diesem Fall gilt für jede Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V, dass $N\mathbf{v}_i = 0$ für alle i gilt.

Angenommen die Aussage für eine Zahl m, dann auch für $m+1$: Wenn $N^{m+1} = 0$ gilt, so erhalten wir wie im Beweis der Dimensionsformel Abbildungen

$$\mathrm{Ker}(N) \subseteq V \twoheadrightarrow \mathrm{Bild}(N).$$

Auf dem Vektorraum $U := \mathrm{Bild}(N)$ gilt dann $N^m(\mathbf{u}) = 0$ für alle $\mathbf{u} \in U$, denn jedes $\mathbf{u} \in \mathrm{Bild}(N) = U$ ist nach Definition von der Form $N(\mathbf{v})$ für ein $\mathbf{v} \in V$ und darum gilt $N^m(\mathbf{u}) = N^m(N(\mathbf{v})) = N^{m+1}(\mathbf{v}) = 0$.

Also finden wir nach Induktionsvoraussetzung eine Basis $B_{\text{Bild}} = \{\mathbf{u}_1, \ldots, \mathbf{u}_\ell\}$ von $U = \mathrm{Bild}(N) \subset V$, die unsere Bedingungen erfüllt.

Wir wählen nun

1. Für alle $\mathbf{u}_i \in B_{\text{Bild}}$ für die es kein $\mathbf{u}_j \in B_{\text{Bild}}$ gibt mit $N(\mathbf{u}_j) = \mathbf{u}_i$ ein Urbild $\mathbf{v}_i \in V$ mit $N(\mathbf{v}_i) = \mathbf{u}_i$. Wir bezeichnen diese Vektoren mit $\{\mathbf{v}_i \mid i \in I\}$.

2. Wir ergänzen die Menge $\{\mathbf{u}_j \in B_{\text{Bild}} \,|\, N\mathbf{u}_j = 0\}$ zu einer Basis $\{\mathbf{u}_j \in B_{\text{Bild}} \,|\, N\mathbf{u}_j = 0\} \cup \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ von $\operatorname{Ker}(N)$.

Dann ist $B_{\text{Bild}} \cup \{\mathbf{v}_i \,|\, i \in I\} \cup \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ eine Basis von V, denn

$$\{\mathbf{u}_j \in B_{\text{Bild}} \,|\, N\mathbf{u}_j = 0\} \cup \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$$

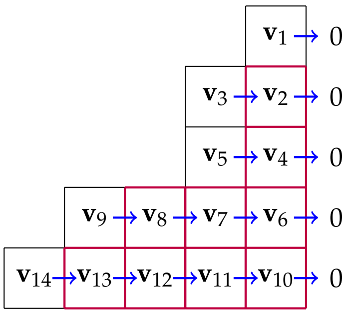
ist nach Konstruktion eine Basis von $\operatorname{Ker}(N)$ und das Komplement

$$\{\mathbf{u}_j \in B_{\text{Bild}} \,|\, N\mathbf{u}_j \neq 0\} \cup \{\mathbf{v}_i \,|\, i \in I\}$$

wird unter N auf eine Basis von $\operatorname{Bild}(N)$ abgebildet. Wie im Beweis der Dimensionsformel muss die Menge dann aber eine Basis von V sein.

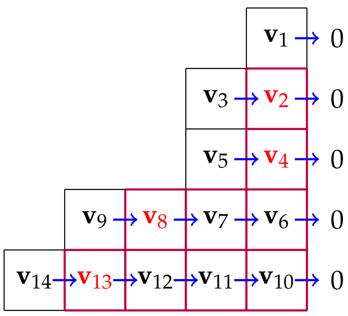
Außerdem erfüllt die Menge nach Konstruktion die Bedingungen aus dem Satz. □

Bemerkung. Um den Beweis zu verstehen und auch um sich das besser merken zu können, sollten Sie einmal versuchen, für jedem Beweisschritt anzuschauen, von welchen Vektoren in unserem Beispiel der nilpotenten Abbildung, die durch ein Diagramm gegeben war jeweils die Rede ist. Zum Beispiel können wir $\operatorname{Bild}(N)$ markieren



denn das entspricht genau den von 0 verschiedenen Spaltenvektoren der zugehörigen Matrix.

Die Vektoren in B_{Bild} die in $\operatorname{Bild}(N)$ noch ohne Urbild sind, wären im Beispiel



.

Markieren Sie die anderen Vektoren im Beweis einmal selbst im Diagramm. Sie werden sehen, dass der Induktionsbeweis die Vektoren nach und nach, von unten rechts (dem Bild der größten Potenz von N, die gerade noch $\neq 0$ ist) ausgehend, nach links und oben aufbaut.

Beispiel 6. Lassen Sie uns das Argument des Beweises am Beispiel

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

ausprobieren. Für die Induktion müssen wir induktiv die Bilder $V \twoheadrightarrow \mathrm{Bild}(B) \twoheadrightarrow \mathrm{Bild}(B^2) \twoheadrightarrow \dots$ kennen, also berechnen wir die Bilder, indem wir jeweils eine Basis angeben:

Das Bild ist der Spann der Spaltenvektoren, so sehen wir $\mathrm{Bild}(B)$.
Für $\mathrm{Bild}(B^2)$ wende ich B auf die gefundene Basis von $\mathrm{Bild}(B)$ an, was etwas weniger Schreibaufwand als B^2 zu berechnen ist, weil ich keine 4×4-Matrizen miteinander multiplizieren muss.

$$\begin{aligned} \mathrm{Bild}(B) &= \mathrm{Span}(\mathbf{e}_1, (\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)) \\ \mathrm{Bild}(B^2) &= \mathrm{Span}(\mathbf{e}_1) \\ \mathrm{Bild}(B^3) &= 0 \\ \mathrm{Ker}(B) \cap \mathrm{Bild}(B) &= \mathrm{Span}(\mathbf{e}_1) \\ \mathrm{Ker}(B) &= \mathrm{Span}(\mathbf{e}_1, \mathbf{e}_2 - \mathbf{e}_3) \end{aligned}$$

Im Beweis können wir also damit beginnen, dass $B\colon \mathrm{Bild}(B^2) \to 0$ die 0 Abbildung ist. Wir wählen also eine beliebige Basis von $\mathrm{Bild}(B^2)$, zum Beispiel $\mathbf{v}_1 = \mathbf{e}_1$.

Dann betrachten wir die Abbildung $B\colon \mathrm{Bild}(B) \twoheadrightarrow \mathrm{Bild}(B^2)$:

$$\mathrm{Ker}(B) \cap \mathrm{Bild}(B) \hookrightarrow \mathrm{Bild}(B) \twoheadrightarrow \mathrm{Bild}(B^2).$$

Der erste Schritt ist also für alle bereits gewählten Basisvektoren, die noch kein Urbild haben, ein Urbild in $\mathrm{Bild}(B)$ zu wählen. Ein Urbild von $\mathbf{v}_1 = \mathbf{e}_1$ ist $\mathbf{v}_2 := \frac{1}{2}(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)$.

Der nächste Schritt wäre, die bereits gewählten Elemente des Kerns von B zu einer Basis von $\mathrm{Ker}(B) \cap \mathrm{Bild}(B) = \mathrm{Span}(\mathbf{e}_1)$ zu ergänzen. Wir haben hierfür aber schon eine Basis.

Der nächste Induktionsschritt betrachtet nun:

$$\mathrm{Ker}(B) \hookrightarrow V \twoheadrightarrow \mathrm{Bild}(B).$$

Wir suchen nun für alle bisher gewählten Basisvektoren, die noch kein Urbild haben (das ist nur $\mathbf{v}_2 := \frac{1}{2}(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)$) ein Urbild in V. Ein Urbild ist $\mathbf{v}_3 := \frac{1}{2}\mathbf{e}_4$.

Jetzt müssen wir noch die Vektoren in $\mathrm{Ker}(B)$ – bisher nur $\mathbf{v}_1 = \mathbf{e}_1$ – zu einer Basis ergänzen, ich wähle $\mathbf{v}_4 = \mathbf{e}_2 - \mathbf{e}_3$.

Jetzt haben wir $\mathbf{v}_4 \to 0$ und $\mathbf{v}_3 \to \mathbf{v}_2 \to \mathbf{v}_1 \to 0$ gefunden, also die Jordanform:

$${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(B) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Bemerkung. Es gibt verschiedene Algorithmen, um eine Basis zu finden, bezüglich der eine nilpotente Abbildung Jordannormalform hat, genauso wie es verschiedene Argumente gibt, warum solche Basen existieren. Ich habe einfach einen Beweis für die Existenz ausgesucht, den ich persönlich besonders übersichtlich fand, aber das ist Geschmackssache. Unser Beweis liefert das folgende Verfahren zur Bestimmung einer Jordanbasis:

Schritt 1 (Für welches m gilt $N^m = 0$?) Bestimme $\mathrm{Bild}(N) \supsetneq \mathrm{Bild}(N^2) \supsetneq \ldots \mathrm{Bild}(N^{m-1}) \supsetneq \mathrm{Bild}(N^m) = 0$ indem wir die Abbildung N induktiv auf Basen der Vektorräume anwenden.

Schritt 2 Wähle eine Basis $B_{m-1} := \mathrm{Neu}_{m-1}$ von $\mathrm{Bild}(N^{m-1})$.

Für alle diese Elemente gilt $N\mathbf{v} = 0$. Diese Elemente werden im Verlauf des Argumentes zu einer Kette der Länge m ergänzt.

Schritt 3 Wähle für alle Basiselemente aus Neu_{m-1} Urbilder $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in $\mathrm{Bild}(N^{m-2})$. Prüfe ob B_{m-1} schon eine Basis von $\mathrm{Ker}(N\colon \mathrm{Bild}(N^{m-2}) \to \mathrm{Bild}(N^{m-1}))$ enthält, falls nicht, wähle $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$, die die vorhandenen Elemente zu einer Basis dieses Kerns ergänzen. Setze $\mathrm{Neu}_{m-2} := \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \cup \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ und $B_{m-2} := B_{m-1} \cup \mathrm{Neu}_{m-2}$. Das ist eine Basis von $\mathrm{Bild}(N^{m-2})$.

Für die Urbilder gilt $N^2\mathbf{v} = N(N\mathbf{v}) = 0$

Für die $\mathbf{w}_i$ gilt wieder $N\mathbf{w}_i = 0$, diese werden noch zu einer Kette der Länge $m-1$ ergänzt.

Wiederhole: Wähle Urbilder von Neu_{m-2} und ergänze die Elemente des Kerns zu einer Basis des Kerns von $\mathrm{Ker}(N\colon \mathrm{Bild}(N^{m-3}) \to \mathrm{Bild}(N^{m-2}))$, bis wir bei $V = \mathrm{Bild}(N^0) = \mathrm{Bild}(\mathrm{id}_V)$ angekommen sind.

Alle Verfahren funktionieren induktiv und suchen als erstes Basisvektoren für Elemente aus dem längsten Jordanblock. Bei unserem Verfahren passiert das, weil wir mit dem Bild der größten Potenz von N, die gerade noch $\neq 0$ ist beginnen, dieses Bild enthält nur Vektoren aus den Jordanblöcken der maximalen Länge.

Aus dem vorigen Satz können wir nun das Resultat zur Jordannormalform ablesen.

Folgerung 7 (Jordannormalform für nilpotente Abbildungen). *Ist $N\colon V \to V$ eine nilpotente Abbildung eines endlichdimensionalen K-Vektorraums, so existiert eine Basis $\underline{\mathbf{v}}$ von V so dass die Matrix von N bezüglich dieser Basis eine Blockmatrix ist, die aus Jordanblöcken zum Eigenwert 0 entlang der Diagonalen besteht:*

$$ {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(N) = \begin{pmatrix} J_{r_1}(0) & 0 & \cdots & 0 \\ 0 & J_{r_2}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(0) \end{pmatrix}. $$

Die Jordanblöcke sind dabei bis auf Reihenfolge eindeutig bestimmt.

Beweis. Die Existenz haben wir im vorherigen Satz bewiesen. Für die Eindeutigkeit können wir die Bemerkung, dass die Längen der Jordanblöcke die Dimension der Kerne $\mathrm{Ker}(N), \mathrm{Ker}(N^2), \ldots$ bestimmt verwenden. Wir wissen, dass für eine Matrix der Form

$$ \begin{pmatrix} J_{r_1}(0) & 0 & \cdots & 0 \\ 0 & J_{r_2}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(0) \end{pmatrix} $$

gilt dass:

$$\begin{aligned}
\dim(\mathrm{Ker}(N)) &= \#\{\text{Blöcke}\} = m \\
\dim(\mathrm{Ker}(N^2)) &= \#\{\text{Blöcke}\} + \#\{\text{Blöcke der Länge } \geq 2\} = m + \#\{r_i | r_i \geq 2\} \\
\dim(\mathrm{Ker}(N^l)) &= \sum_{j=1}^{l} \underbrace{\#\{i \,|\, r_i \geq j\}}_{\#\text{Blöcke der Länge } \geq j}
\end{aligned}$$

Da in diesem Gleichungssystem in der l-ten Zeile jeweils die Unbekannte $\#\{r_i | r_i \geq l\}$ hinzukommt, können wir umgekehrt aus den Zahlen $\dim(\mathrm{Ker}(N)), \dim(\mathrm{Ker}(N^2)), \ldots, \dim(\mathrm{Ker}(N^n))$ die Anzahle der Blöcke jeder Länge bestimmen. Das zeigt die Eindeutigkeit. □

Frage. *Angenommen Sie haben von einer Matrix $N \colon K^{10} \to K^{10}$ berechnet, dass*

$$\begin{aligned}
\dim(\mathrm{Ker}(N)) &= 4 \\
\dim(\mathrm{Ker}(N^2)) &= 7 \\
\dim(\mathrm{Ker}(N^3)) &= 9 \\
\dim(\mathrm{Ker}(N^4)) &= 10.
\end{aligned}$$

Was ist die Jordannormalform von N?

ANTWORT: 4 Jordanblöcke, davon haben $7 - 4 = 3$ die Länge ≥ 2 also gibt es einen Jordanblock der Länge 1, es haben $9 - 7 = 2$ Jordanblöcke Länge ≥ 3 und also genau einen Jordanblock der Länge 2, und $10 - 9 = 1$ die Länge ≥ 4, also gibt es genau einen Block der Länge 3 und einen der Länge 4, die Jordannormalform ist also

$$\begin{pmatrix} 0 & & & \\ & J_2(0) & & \\ & & J_3(0) & \\ & & & J_4(0) \end{pmatrix}.$$

Frage. *Was ist wenn Sie stattdessen*

$$\begin{aligned}
\dim(\mathrm{Ker}(N)) &= 4 \\
\dim(\mathrm{Ker}(N^2)) &= 6 \\
\dim(\mathrm{Ker}(N^3)) &= 9 \\
\dim(\mathrm{Ker}(N^4)) &= 10
\end{aligned}$$

berechnet haben?

ANTWORT: Dann haben Sie sich verrechnet, da

$$\dim(\mathrm{Ker}(N^2)) - \dim(\mathrm{Ker}(N)) = 6 - 4 = 2$$

bedeutet, dass es 2 Jordanblöcke der Länge ≥ 2 gibt und

$$\dim(\mathrm{Ker}(N^3)) - \dim(\mathrm{Ker}(N^2)) = 3$$

bedeutet, dass es 3 Jordanblöcke der Länge $\geq$ 3. Das passt nicht zusammen.

Die Normalform für allgemeine Abbildungen

Da wir jetzt verstanden haben, wie wir für nilpotente Abbildungen gute Basen finden, ergibt sich der allgemeine Fall mit Hilfe der Hauptraumzerlegung erfreulich leicht. Wüssten wir für eine Abbildung $A\colon V \to V$ statt $A^n = 0$ stattdessen $(A - c\,\mathrm{id}_V)^n = 0$, so könnten wir das Verfahren auf $N := A - c\,\mathrm{id}_V$ anwenden und finden eine Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ so dass

$$
{}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A - c\,\mathrm{id}_V) = \begin{pmatrix} J_{r_1}(0) & 0 & \dots & 0 \\ 0 & J_{r_2}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(0) \end{pmatrix}.
$$

Das bedeutet aber, dass

$$
\begin{aligned}
{}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) &= {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A - c\,\mathrm{id}_V) + {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(c\,\mathrm{id}_V) \\
&= \begin{pmatrix} J_{r_1}(0) & 0 & \dots & 0 \\ 0 & J_{r_2}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(0) \end{pmatrix} + \begin{pmatrix} c & 0 & \dots & 0 \\ 0 & c & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & c \end{pmatrix} \\
&= \begin{pmatrix} J_{r_1}(c) & 0 & \dots & 0 \\ 0 & J_{r_2}(c) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(c) \end{pmatrix}.
\end{aligned}
$$

Ist $\det(A - t\,\mathrm{id}_V) = (c - t)^n$, so wissen wir also wie wir eine Basis finden, bezüglich derer die Matrix eine Blockmatrix der Form

$$
\begin{array}{cccccccccccccc}
c \\
& c & 1 \\
& & c \\
& & & c & 1 \\
& & & & c \\
& & & & & c & 1 \\
& & & & & & c & 1 \\
& & & & & & & c & 1 \\
& & & & & & & & c \\
& & & & & & & & & c & 1 \\
& & & & & & & & & & c & 1 \\
& & & & & & & & & & & c & 1 \\
& & & & & & & & & & & & c & 1 \\
& & & & & & & & & & & & & c
\end{array}.
$$

ist. Die Blöcke sind dann wieder bis auf Reihenfolge eindeutig bestimmt. Für eine allgemeine Abbildung können wir zunächst V in die Haupträume $\mathrm{Hau}(A, c_i) = \ker((A - c_i\,\mathrm{id}_V)^n)$ zerlegen und dann dieses Ergebnis auf jeden Hauptraum anwenden. Das Ergebnis fassen wir im folgenden Satz zusammen.

Satz (Jordannormalform)**.** *Sei $A\colon V \to V$ eine lineare Abbildung mit charakteristischem Polynom*

$$
\det(A - t\,\mathrm{id}_V) = \prod_{i=1}^{r} (c_i - t)^{n_i}
$$

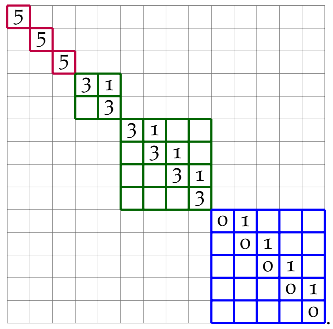
wobei $c_1, \ldots, c_r \in K$ paarweise verschieden sind. Dann existiert eine Basis $\underline{\mathbf{v}}$ von V so dass die Matrix von A bezüglich dieser Basis eine Blockmatrix ist, die aus Jordanblöcken zu den Eigenwerten $c_1, \ldots, c_r$ entlang der Diagonalen besteht:

$$ {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) = \begin{pmatrix} J_{r_1}(c_1) & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(c_r) \end{pmatrix}. $$

Die Jordanblöcke sind dabei bis auf Reihenfolge eindeutig bestimmt.

Die Summe der Größen der Blöcke zu fixiertem Eigenwert c_i ist dabei genau die Vielfachheit n_i der Nullstelle c_i des charakteristischen Polynoms.

Eine Matrix in Jordannormalform:



Zwischenfrage aus dem Publikum: Wie rechnen wir damit die Potenzen von A aus? Das haben wir in der Vorlesung an dieser Stelle beantwortet, im Skript möchte ich es nach dem Beweis einfügen.

Beweis. Die Existenz können wir aus den bereits gezeigten Resultaten ablesen:

1. Nach der Hauptraumzerlegung ist $V = \bigoplus \mathrm{Ker}((A - c_i \,\mathrm{id}_V)^n)$.

2. Für jeden der Haupträume $\mathrm{Hau}(A, c_i) = \mathrm{Ker}((A - c_i \,\mathrm{id}_V)^n)$ können wir den Satz über die Jordannormalform für nilpotente Abbildungen auf die Abbildung

$$ N_i := (A - c_i \,\mathrm{id}_V)|_{\mathrm{Hau}(A,c_i)} \colon \mathrm{Hau}(A, c_i) \to \mathrm{Hau}(A, c_i) $$

anwenden und erhalten eine Basis $B_i = \{\mathbf{v}_{1,i}, \ldots, \mathbf{v}_{m_i,i}\}$, so dass

$$ {}_{\underline{B_i}}\mathrm{Mat}_{\underline{B_i}}(A|_{\mathrm{Hau}(A,c_i)}) = \begin{pmatrix} J_{r_{i,1}}(c_i) & 0 & \ldots \\ 0 & J_{r_{i,1}}(c_i) & 0 \\ \vdots & 0 & \ddots \end{pmatrix} $$

aus Jordanblöcken zum Eigenwert c_i besteht.

3. Zusammen bilden die Basen B_i für die Haupträume $\mathrm{Hau}(A, c_i)$ eine Basis $B = \cup B_i$ von V, bezüglich der die Matrix eine Blockmatrix der gewünschten Form ist.

Die Eindeutigkeit bis auf Reihenfolge ergibt sich aus dem Resultat für nilpotente Abbildungen, denn die Haupträume sind nach Definition $\mathrm{Ker}((A - c_i \,\mathrm{id}_V)^n)$, für eine Jordanmatrix ist dieser Kern nach Konstruktion gerade die lineare Hülle der Basisvektoren die zu den Spalten der Jordanblöcke zum Eigenwert c_i gehören. □

Eine Anwendung der Jordannormalform ist eine Formel für die Potenzen einer linearen Abbildung. Damit können wir insbesondere verstehen, wie die Bilder $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \ldots$ für einen beliebigen Vektor $\mathbf{v}$ aussehen.

Da die Potenz einer Jordanmatrix die Blockmatrix ist, die aus den Potenzen der Jordanblöcke besteht, genügt es, uns zu überlegen, wie wir die Potenzen $J_r(c)^n$ eines Jordanblocks berechnen. Für den Eigenwert 0 wissen wir, dass die Potenzen von $J_r(0)$ dadurch gegeben sind, dass wir die 1sen jeweils auf die nächst höhere Nebendiagonale schreiben:

$$J_r(c) = \begin{pmatrix} c & 1 & 0 & \cdots & 0 \\ & c & 1 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & c \end{pmatrix}$$

$$J_r(0)^2 = \begin{pmatrix} 0 & 0 & 1 & \cdots & 0 \\ & 0 & 0 & 1 & \vdots \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix}, J_r(0)^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 1 \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix}, \ldots$$

Lassen Sie uns damit $J_r(c)^2$ berechnen:

ausmultiplizieren

$$\begin{aligned} J_r(c)^2 &= \left(\begin{pmatrix} c & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & c \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & \ldots & 0 \end{pmatrix} \right)^2 \\ &= \begin{pmatrix} c & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & c \end{pmatrix}^2 + \begin{pmatrix} c & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & c \end{pmatrix} \circ J_r(0) \\ &\quad + J_r(0) \circ \begin{pmatrix} c & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & c \end{pmatrix} + J_r(0)^2 \\ &= \begin{pmatrix} c^2 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & c^2 \end{pmatrix} + 2 \begin{pmatrix} 0 & c & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & c \\ 0 & \ldots & \ldots & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & \ldots & 0 \end{pmatrix} \\ &= \begin{pmatrix} c^2 & 2c & 1 & 0 \\ 0 & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & 2c \\ 0 & \ldots & \ldots & c^2 \end{pmatrix} \end{aligned}$$

Wir haben hier wie bei der 1. binomischen Formel einfach das Produkt

$$(x+y)^2 = (x+y) \cdot (x+y) = x^2 + xy + yx + y^2$$

ausmultipliziert und hier Glück, dass die Matrizen cE_n und $J_r(0)$ kommutieren, d.h. es gilt

$$(cE_n) \circ J_r(0) = J_r(0) \circ (cE_n) = c \cdot J_r(0).$$

Das können wir für die höheren Potenzen genauso berechnen, wenn wir die allgemeine binomische Formel für die höheren Potenzen $(x+y)^n$ kennen.

Die allgemeine binomische Formel können wir uns leicht überlegen, indem wir bemerken, dass wir beim Ausmultiplizieren des Produkts

$$\underbrace{(x+y)\cdot(x+y)\cdot\cdots\cdot(x+y)}_{n\text{-Faktoren}} = x^n + n\cdot x^{n-1}y + \dots$$

einfach die Summe über alle Möglichkeiten ist, in den n Faktoren $(x+y)$ entweder x oder y auszuwählen. Den Term x^n erhalten wir, wenn wir in allen Faktoren x wählen, $x^{n-1}y$ wenn wir in einem der n Faktoren y und in allen anderen x wählen, davon gibt es n Möglichkeiten. Den Term $x^{n-2}y^2$ erhalten wir, wenn wir in 2 der n Faktoren das y wählen. D avon gibt es $\frac{n(n-1)}{2}$ Möglichkeiten, denn für das erste y haben wir n Möglichkeiten, danach bleiben $n-1$ mögliche Faktoren, für das zweite y, macht $n(n-1)$, aber damit zählen wir alle Möglichkeiten doppelt, denn bei der Wahl der y's ist die Reihenfolge egal. Allgemein finden wir also den Term $x^{n-k}y^k$ wenn wir aus k Faktoren den Summanden y wählen, dafür gibt es $\binom{n}{k} = \frac{n(n-1)\cdots(n-(k-1))}{k!}$ viele Möglichkeiten. Damit erhalten wir die Formel:

$$\begin{aligned}(x+y)^n &= x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + y^n \\ &= \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k.\end{aligned}$$

Hierbei ist $\binom{n}{1} = \binom{n}{n-1} = n$, denn $n-1$ mal y auszuwählen ist das gleiche wie einmal x auszuwählen.

Das Symbol $\binom{n}{k}$ heißt *Binomialkoeffizient* weil es in der binmischen Formel vorkommt.

Es bezeichnet einfach die Anzahl der Möglichkeiten k Elemente aus einer Menge mit n Elementen auszuwählen. Auf Englisch wird es darum als „n choose k" gelesen, auf Deutsch sagen wir leider „n über k", obwohl „wähle k aus n" besser daran erinnern würde, was das Symbol bedeutet.

Die Formel $\binom{n}{k} = \frac{n(n-1)\cdots(n-(k-1))}{k!} = \frac{n!}{k!\cdot(n-k)!}$ haben wir uns gerade überlegt, es gibt für die Wahl des ersten Elementes n Möglichkeiten, für das zweite dann $n-1$ usw, für das k-te Element bleiben noch $n-(k-1)$ übrig, was $n(n-1)\cdots(n-(k-1))$ Möglichkeiten ergibt, für die Reihenfolge von k Elementen hatten uns bei der symmetrischen Gruppe schon überlegt, dass es dafür $k!$ viele Möglichkeiten gibt.

Für die Potenzen von $J_r(c)$ erhalten wir darum:

$$\begin{aligned}J_r(c)^n &= (cE_r + J_r(0))^n \\ &= c^nE_r + \binom{n}{1}c^{n-1}J_r(0) + \binom{n}{2}c^{n-2}J_r(0)^2 + \cdots + J_r(0)^n \\ &= \sum_{k=0}^{n}\binom{n}{k}c^{n-k}J_r(0)^k.\end{aligned}$$

Da für $k \geq r$ gilt, dass $J_r(0)^k = 0$ die Nullmatrix ist, können wir in der Summe sogar alle Terme mit $k \geq r$ weglassen, selbst für große n hat diese Summe also nur r Summanden.

Da die k-te Potenz von $J_r(0)$ genau auf der k-ten Nebendiagonalen 1-sen hat, bedeutet diese Formel, dass

$$J_r(c)^n = \begin{pmatrix} c^n & \binom{n}{1}c^{n-1} & \binom{n}{2}c^{n-2} & \cdots & \binom{n}{r-1}c^{n-(r-1)} \\ & c^n & \binom{n}{1}c^{n-1} & \binom{n}{2}c^{n-2} & \vdots \\ & & \ddots & \ddots & \binom{n}{2}c^{n-2} \\ & & & \ddots & \binom{n}{1}c^{n-1} \\ & & & & c^n \end{pmatrix}.$$

Damit haben wir eine Formel für die Potenzen einer Matrix in Jordannormalform gefunden.

Bemerkung (Potenzen von A aus der Jordannormalform berechnen)**.** Die Formel für die Potenzen einer Matrix hatten wir in der linearen Algebra 1 schon zum Lösen linearer Rekursionen verwendet. Dafür benötigen wir nicht nur die Normalform, sondern auch eine Basis $\underline{v}$ bezüglich der die Matrix A dann Normalform hat, denn dann gilt wegen der Formel für den Basiswechsel, dass

$$A^n = B \circ {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)^n \circ B^{-1}$$

für die Matrix B deren Spalten die Vektoren der Basis $\underline{v}$ sind.

In einer Übungsaufgabe werden Sie zeigen, dass für die Matrizen, die lineare Rekursionen beschreiben tatsächlich immer Jordanblöcke maximaler Länge auftreten.

ERINNERUNG: $B = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ ist die Matrix, die $\underline{\mathbf{v}}$-Koordinaten in $\underline{\mathbf{e}}$-Koordinaten übersetzt (wir hatten diese Abbildung $\mathrm{Komb}_{\underline{v}}$ genannt) und umgekehrt rechnet B^{-1} die Standardkoordinaten in $\underline{\mathbf{v}}$-Koordinaten um (das war $\mathrm{Koord}_{\mathbf{v}}$).

Bemerkung (Anwendung bei linearen Differentialgleichungen)**.** Eine andere Anwendung der Formel für Potenzen einer Matrix wird Ihnen vielleicht in der Analysis oder in der Vorlesung „Mathematisches Modellieren" begegnen. Das Verhalten von kontinuierlichen, z.B. physikalischen, Prozessen wird of durch Differentialgleichungen beschrieben, das einfachste Beispiel ist vielleicht die Schwingung einer Feder die durch $f''(t) = -c \cdot f(t)$ beschrieben werden kann.

Die zweite Ableitung f'' ist die Geschwindigkeit und bei einer idealen Federn hängt die Kraft linear von der Auslenkunng f ab, das Newtonsche Gesetz „Kraft ist Masse mal Beschleunigung." liefert dann die Gleichung.

Bei anderen Prozessen treten auch höhere Ableitungen ($f^{(n)}(t) :=$ n-te Ableitung) und Linearkombinationen von Ableitungen auf, zum Beispiel also Gleichungen der Form

$$f^{(n)}(t) = a_0 f(t) + a_1 f'(t) + \dots + a_{n-1} f^{(n-1)}(t),$$

die Sie vielleicht an die linearen Rekursionen erinnern. Mit dem gleichen Trick, den wir dort verwendet hatten, können wir diese Gleichung umschreiben als

$$\begin{pmatrix} f'(t) \\ f''(t) \\ \vdots \\ f^{(n)}(t) \end{pmatrix}' \begin{pmatrix} f(t) \\ f'(t) \\ \vdots \\ f^{(n-1)}(t) \end{pmatrix}' = A \circ \begin{pmatrix} f(t) \\ f'(t) \\ \vdots \\ f^{(n-1)}(t) \end{pmatrix}.$$

Wenn wir uns dann an die Gleichung $(e^{at})' = ae^{at}$ erinnern,

könnten wir versuchen, mit der Formel $e^{At} \circ \begin{pmatrix} f(0) \\ f'(0) \\ \vdots \\ f^{(n-1)}(0) \end{pmatrix}$ eine Lösung aufzuschreiben, wobei

$$e^{At} = \sum_{n=0}^{\infty} \frac{1}{n!} A^n t^n.$$

Das funktioniert tatsächlich – Sie müssen sich hier natürlich über die Konvergenz der unendlichen Summe und die Rechenregeln für Ableitungen Gedanken machen – aber um die unendliche Summe ausrechnen zu können, brauchen Sie in jedem Fall eine Formel für A^n und die haben wir gerade gefunden.

ERINNERUNG: Die Exponentialreihe, bzw. Expnentialfunktion

$$e^x := \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

ist die eindeutige Potenzreihe für die $(e^x)' = e^x$ und $e^0 = 1$ gilt, d.h. die Steigung der zugehörigen Funktion ist an jeder Stelle gleich dem Funktionswert.

Wir werden uns später noch einmal überlegen, dass aus dieser Eigenschaft die zweite nützliche Gleichung

$$e^{x+y} = e^x \cdot e^y$$

folgt, was die Notation als „e hoch x" erklärt.

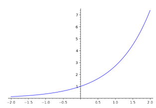


Abbildung 1: Funktionsgraph e^x

Es ist höchste Zeit für ein Beispiel. Lassen Sie uns einmal anschauen, wie wir das Argument aus dem Beweis zur Jordannormalform benutzen können, um die Jordannormalform für eine gegebene Matrix A zu berechnen:

$$A = \begin{pmatrix} 1 & 1 & -1 & -1 & 1 \\ -1 & 3 & -2 & 0 & 2 \\ \mathbf{0} & \mathbf{0} & 1 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & 1 \end{pmatrix}$$

Schritt 1: *Bestimme das charakteristische Polynome und damit die Eigenwerte von A.* Die Matrix ist eine Blockmatrix, also gilt

$$\begin{aligned} \det(A - tE_5) &= \det\left(\begin{pmatrix} 1-t & 1 & -1 & -1 & 1 \\ -1 & 3-t & -2 & 0 & 2 \\ \mathbf{0} & \mathbf{0} & 1-t & 1 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 1-t & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & 1-t \end{pmatrix}\right) \\ &= ((1-t)(3-t)+1) \cdot ((1-t)(1-t)(1-t)) \\ &= (t^2 - 4t + 4) \cdot (1-t)^3 = (t-2)^2(1-t)^3. \end{aligned}$$

Die Eigenwerte sind also der dreifache Eigenwert 1 und der doppelte Eigenwert 2.

Schritt 2: *Bestimme die Bestimme die Haupträume und Jordanbasen für die Haupträume.* Ich berechne dazu die die Kerne von $(A - c\,\mathrm{id})^m$ für $m = 1, 2, 3$, die Dimensionen geben mir dann schon die Jordannormalform an:

Eigenwert 1: Berechne $\mathrm{Ker}(A - 1 \cdot E_5)$ mit Gaußverfahren:

$$\begin{pmatrix} 0 & 1 & -1 & -1 & 1 \\ -1 & 2 & -2 & 0 & 2 \\ \mathbf{0} & \mathbf{0} & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} -1 & 2 & -2 & 0 & 2 \\ 0 & 1 & -1 & -1 & 1 \\ \mathbf{0} & \mathbf{0} & 0 & 1 & 0 \end{pmatrix} \begin{matrix} {}^{+} \\ + \;\; | -2 \\ \end{matrix}$$

$$\text{erhalte} \qquad \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & 0 & 1 & 0 \end{pmatrix}$$

Der Kern ist also 2-dimensional, es gibt also 2 Jordanblöcke zum dreifachen Eigenwert 1, die Jordannormalform zu diesem Eigenwert muss also bis auf Reihenfolge der beiden Blöcke $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ sein.

Berechne $(A - 1 \cdot E_5)^2$:

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

Diese Matrix hat in der Tat Rang 2, also einen 3-dimensionalen Kern. Für eine Basis des Kerns können wir die Matrix auf Zeilenstufenform umschreiben:

$$\begin{pmatrix} 1 & -2 & 2 & 1 & -2 \\ 0 & -1 & 1 & 1 & -1 \end{pmatrix} \begin{array}{l} \longleftarrow\!\rfloor^{+} \\ | -2 \end{array} \sim \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Daran können wir eine Basis ablesen:

$$\mathbf{w}_1 = \begin{pmatrix} 0 \\ 1 \\ \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \mathbf{w}_2 = \begin{pmatrix} 1 \\ 1 \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \end{pmatrix}, \mathbf{w}_3 = \begin{pmatrix} 0 \\ -1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{pmatrix}.$$

Dieser Raum hat die Dimension des Hauptraums, also haben wir den Hauptraum gefunden.

Jetzt bestimmen wir eine Jordanbasis für die nilpotente Abbildung $A - 1 \cdot \mathrm{id}$ auf dem Hauptraum, indem wir die Bilder unserer Basis berechnen. Das Bild des Hauptraumes unter $A - 1 \cdot \mathrm{id}$ ist:

$$\mathrm{Span}\Big((A - 1E_5)\mathbf{w}_1, (A - 1E_5)\mathbf{w}_2, (A - 1E_5)\mathbf{w}_3\Big)$$
$$= \mathrm{Span}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right) = \mathrm{Span}\left(\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right).$$

Damit wählen wir $\mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ als Basis des Bildes. Ein Urbild dieses Vektors unter der Abbildung ist $\mathbf{v}_2 := \mathbf{w}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$. Schließlich müssen wir $\mathbf{v}_1$ noch zu einer Basis des Kerns von $A - \mathrm{id}$ ergänzen.

Dazu schauen wir die Zeilenstufenform der Matrix an und finden

$\mathbf{v}_3 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$. Damit haben wir eine Jordanbasis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ für den Hauptraum zum Eigenwert 1 gefunden.

Eigenwert 2: $\operatorname{Ker}(A - 2E_5)$:

$$\begin{pmatrix} -1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -2 & 0 & 2 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{array}{l} \mid -1 \\ + \quad + \quad + \\ \\ \\ \\ \end{array} \qquad \begin{pmatrix} -1 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Der Kern ist eindimensional, es gibt also nur einen Jordanblock zum Eigenwert 2.

Berechne $(A - 2 \cdot E_5)^2$:

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

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

Ich rechne die letzten Einträge nicht mehr aus, weil ich schon die Basis $\mathbf{w}_4 = \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{w}_5 = \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ des Kerns sehe. Da der Hauptraum 2-dimensional ist, haben wir hiermit eine Basis des Hauptraum bestimmt.

Jetzt bestimmen wir eine Jordanbasis für die nilpotente Abbildung $A - 2 \cdot \mathrm{id}$ auf diesem Hauptraum. Das Bild des Hauptraumes unter $A - 2 \cdot \mathrm{id}$ ist:

$$\operatorname{Span}\Big((A - 2E_5)\mathbf{w}_4, (A - 2E_5)\mathbf{w}_5\Big)$$

$$= \operatorname{Span}\left(\begin{pmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right) = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right).$$

Damit wählen wir $\mathbf{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ als Basisvektor, ein Urbild dieses

Vektors ist $\mathbf{v}_5 = \mathbf{e}_2$.

Damit haben wir eine Jordanbasis für A gefunden:

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_5 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

für die wir wissen

$$A\mathbf{v}_1 = \mathbf{v}_1, A\mathbf{v}_2 = \mathbf{v}_1 + \mathbf{v}_2, A\mathbf{v}_3 = \mathbf{v}_3, A\mathbf{v}_4 = 2\mathbf{v}_4, A\mathbf{v}_5 = 2\mathbf{v}_5 + \mathbf{v}_4$$

und also

$$_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

Probe: $A \cdot (\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\mathbf{v}_4\mathbf{v}_5) =$

$$\begin{pmatrix} 1 & 1 & -1 & -1 & 1 \\ -1 & 3 & -2 & 0 & 2 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

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

$$= (\mathbf{v}_1(\mathbf{v}_1 + \mathbf{v}_2)\mathbf{v}_3(2\mathbf{v}_4)(\mathbf{v}_4 + 2\mathbf{v}_5))$$

Bemerkung (Jordanzerlegung)**.** Jede Jordanmatrix

$$J = \begin{pmatrix} J_{r_1}(c_1) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(c_r) \end{pmatrix}$$

können wir als Summe einer Diagonalmatrix J_{diag} und der nilpoten-

ten Matrix $J_{nilp} = \begin{pmatrix} J_{r_1}(0) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{r_m}(0) \end{pmatrix}$ schreiben, wobei die Matrizen J_{diag}, J_{nilp} kommutieren.

Damit können wir also auch jede $n \times n$ Matrix A als Summe

$$A = A_{diag} + A_{nilp}$$

schreiben, wobei A_{diag} eine diagonalisierbare Matrix und A_{nilp} nilpotent ist und außerdem

$$A_{diag} \circ A_{nilp} = A_{nilp} \circ A_{diag}$$

gilt.

Es ist nicht schwer zu beweisen, dass diese Eigenschaften die Zerlegung eindeutig bestimmen, das möchte ich hier aber nicht tun.

Bemerkung (Geometrische Interpretation der Jordannormalform)**.** Eine Jordanmatrix

$$J_r(c) = \begin{pmatrix} c & 1 & 0 & \cdots & 0 \\ & c & 1 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & c \end{pmatrix}$$

mit $c \neq 0$ können wir als Produkt

$$J_r(c) = \begin{pmatrix} c & 1 & 0 & \cdots & 0 \\ & c & 1 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & c \end{pmatrix} = \begin{pmatrix} c & 0 & 0 & \cdots & 0 \\ & c & 0 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 0 \\ & & & & c \end{pmatrix} \circ \begin{pmatrix} 1 & \frac{1}{c} & 0 & \cdots & 0 \\ & 1 & \frac{1}{c} & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & \frac{1}{c} \\ & & & & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & \frac{1}{c} & 0 & \cdots & 0 \\ & 1 & \frac{1}{c} & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & \frac{1}{c} \\ & & & & 1 \end{pmatrix} \circ \begin{pmatrix} c & 0 & 0 & \cdots & 0 \\ & c & 0 & 0 & \vdots \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 0 \\ & & & & c \end{pmatrix}$$

schreiben. Die Abbildung ist also die Komposition einer Streckung mit dem Faktor c mit einer Scherung. Wir können also jede lineare Abbildung in geeigneten Koordinaten als Komposition von einer Streckung und einer Scherungen verstehen.

Jede Matrix genügt ihrem charakteristischen Polynom

Zum Abschluss dieses Kapitels möchte ich noch ein Resultat nachtragen, das wir bei unserem Zugang zur Jordannormalform nicht verwendet haben, das aber einfach ein schönes Resultat ist und außerdem im Argument noch einmal anschaut was beim wiederholten Anwenden einer linearen Abbildung A auf einen Vektor $\mathbf{v}$ passiert.

Satz 8 (Cayley-Hamilton). *Jede Matrix genügt ihrem charakteristischen Polynom, d.h.: Ist $A\colon V \to V$ eine lineare Abbildung eines n-dimensionalen Vektorraums, und $\operatorname{charpol}_A(t) = \det(A - t\,\mathrm{id}_V)$ das charakteristische Polynom von A, so gilt*

$$\operatorname{charpol}_A(A) = 0.$$

Lassen Sie uns die Notation entschlüsseln: Das charakteristische Polynom, ist ein Polynom mit Koeffizienten in K, also ein Ausdruck der Form

$$\begin{aligned} p(t) &= a_0 + a_1 t + \cdots + a_n t^n \\ &= a_0 t^0 + a_1 t + \cdots + a_n t^n \qquad \text{setze } t^0 := 1. \end{aligned}$$

Für eine Endomorphismen $A\colon V \to V$ (oder $n \times n$-Matrizen) hatten wir $A^m = \underbrace{A \circ \cdots \circ A}_{m-\text{Faktoren}}$ definiert, dafür ist genauso $A^0 := \mathrm{id}_V$ sinnvoll, weil „A 0-mal anwenden=nichts tun". Damit setzen wir

$$\begin{aligned} p(A) &:= a_0 A^0 + a_1 A + \cdots + a_n A^n \\ &= a_0 \,\mathrm{id}_V + a_1 A + \cdots + a_n A^n \in \operatorname{End}(V). \end{aligned}$$

$\operatorname{End}(V) = \{A\colon V \to V \mid A \text{ linear}\}$

Diese Definition liefert also für jedes A eine Abbildung

$$\begin{aligned} K[t] &\to \operatorname{End}(V) \\ p(t) &\mapsto p(A) \end{aligned}$$

die mit $+$ und $\circ$ verträglich ist, d.h. für alle Polynome $p, q \in K[t]$ gilt

$$(p+q)(A) = p(A) + q(A) \text{ und } (p \cdot q)(A) = p(A) \circ q(A).$$

Die Formel $\operatorname{charpol}_A(A) = 0$ bedeutet damit, dass $\operatorname{charpol}_A(A) \in \operatorname{End}(V)$ die 0-Abbildung auf V ist.

Plausibilitätstest: Für diagonalisierbare Matrizen können wir die Aussage einfach ablesen, indem wir eine Basis aus Eigenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ zu Eigenwerten $c_1, \ldots, c_n \in K$ wählen. Dann ist $\operatorname{charpol}_A(t) = \prod_{i=1}^n (c_i - t)$ und für jedes der $\mathbf{v}_j$ gilt daher

$$\begin{aligned}\operatorname{charpol}_A(A)\mathbf{v}_j &= \prod_{i \neq j} (c_i \operatorname{id} - A) \circ (c_j \operatorname{id}_V - A)\mathbf{v}_j \\ &= \prod_{i \neq j} (c_i \operatorname{id} - A)(0) = 0.\end{aligned}$$

Und eine lineare Abbildung die eine Basis auf 0 abbildet, ist die 0-Abbildung.

Beweis des Satzes. Ist $\operatorname{charpol}_A(t) = a_0 + a_1 t + \cdots + a_n t^n$, so müssen wir zeigen, dass für alle $\mathbf{v} \in V$ die Gleichung

$$\begin{aligned}\operatorname{charpol}_A(A)\mathbf{v} &:= a_0 A^0 \mathbf{v} + a_1 A\mathbf{v} + \cdots + a_n A^n \mathbf{v} \\ &= a_0 \mathbf{v} + a_1 A\mathbf{v} + \cdots + a_n A^n \mathbf{v} \stackrel{?}{=} 0 \in V.\end{aligned}$$

gilt.

Dazu wählen wir uns eine geschickte Basis von V: Ist $\mathbf{v} \neq 0$ (für $\mathbf{v} = 0$ ist nichts zu zeigen), so wählen wir die kleinste ganze Zahl m für die die Vektoren $\mathbf{v}, A\mathbf{v}, \ldots, A^m\mathbf{v}$ linear abhängig sind. (Spätestens für $m = \dim V = n$ sind die Vektoren linear abhängig.)

Immerhin wissen wir, dass die $n+1$ Vektoren $\mathbf{v}, A\mathbf{v}, \ldots, A^n\mathbf{v}$ im n-dimensionalen Vektorraum V linear abhängig sein müssen. Die Frage ist nur, wieso die Koeffizienten des charakteristischen Polynoms eine nichttriviale Linearkombination für diese Vektoren angeben.

Dann sind $\mathbf{v}, A\mathbf{v}, \ldots, A^{m-1}\mathbf{v}$ linear unabhängig und wir können also $A^m\mathbf{v}$ als Linearkombination

$$A^m \mathbf{v} = c_0 \mathbf{v} + c_1 A\mathbf{v} + \cdots + c_{m-1} A^{m-1} \mathbf{v}$$

schreiben.

Ergänzen wir die Vektoren $\mathbf{v}, A\mathbf{v}, \ldots, A^{m-1}\mathbf{v}$ zu einer Basis $\underline{B} = \mathbf{v}, A\mathbf{v}, \ldots, A^{m-1}\mathbf{v}, \mathbf{w}_m, \ldots, \mathbf{w}_{n-1}$ von V, so ist die Matrix von A bezüglich dieser Basis eine Blockmatrix:

$$_{\underline{B}}\operatorname{Mat}_{\underline{B}}(A) = \begin{pmatrix} 0 & \ldots & 0 & c_0 & ? \\ 1 & 0 & 0 & c_1 & ? \\ 0 & \ddots & 0 & \vdots & ? \\ \vdots & 0 & 1 & c_{n-1} & ? \\ & & & & M \end{pmatrix}$$

wobei M eine $(n-m) \times (n-m)$-Matrix ist und ? Einträge, die wir nicht kennen. Daher ist

$$\begin{aligned}\operatorname{charpol}_A(t) &= \operatorname{charpol}_M(t) \cdot \det\left(\begin{pmatrix} -t & \ldots & 0 & c_0 \\ 1 & -t & 0 & c_1 \\ 0 & \ddots & -t & \vdots \\ \vdots & 0 & 1 & c_{n-1} - t \end{pmatrix}\right) \\ &= \operatorname{charpol}_M(t)(-1)^m \left(t^m - c_{m-1}t^{m-1} - \cdots - c_0\right).\end{aligned}$$

Darum gilt

Diese Determinante hatten Sie im ersten Semester berechnet.

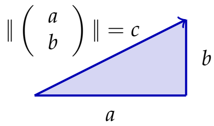
$$\begin{aligned}\operatorname{charpol}_A(A)\mathbf{v} &= \operatorname{charpol}_M(A)\Big((-1)^m \underbrace{A^m\mathbf{v} - c_{m-1}A^{m-1}\mathbf{v} - \cdots - c_0\mathbf{v}}_{=0}\Big)\\ &= 0\end{aligned}$$

□

Skalarprodukte und Abstandsbegriffe

Anschauliche Herleitung des Längenbegriffs

Der Satz des Pythagoras besagt, dass in einem rechtwinkligen Dreieck in der Ebene die Länge der Hypotenuse c durch die Gleichung



$$c^2 = a^2 + b^2$$

bestimmt ist und umgekehrt ein Dreieck genau dann rechtwinklig ist, wenn diese Gleichung gilt. Dieser zweite Teil der Aussage wird in Schulbüchern oft überschlagen.

Im Vektorraum $\mathbb{R}^n$, oder zumindest in der Ebene $\mathbb{R}^2$ und im 3-dimensionalen Raum $\mathbb{R}^3$ können Sie damit den Abstand zwischen zwei Punkten, bzw. die Länge eines Vektors mit dem Satz des Pythagoras berechnen

$$\| \begin{pmatrix} a \\ b \end{pmatrix} \| = \sqrt{a^2 + b^2},$$

wir schreiben dabei $\|\mathbf{v}\| \in \mathbb{R}$ für die Länge des Vektors $\mathbf{v}$. Induktiv können Sie sich im $\mathbb{R}^3$ analog die Formel

$$\| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

anschaulich überlegen. Darum definieren wir im $\mathbb{R}^n$ die Länge eines Vektors als

$$\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \| := \sqrt{x_1^2 + \cdots + x_n^2}.$$

Die Umkehrung des Satzes von Pythagoras, können wir dann benutzen, um zu prüfen, ob zwei Vektoren $\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{w} =$

$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ senkrecht aufeinander stehen, denn

$$\begin{aligned} \|\mathbf{v}+\mathbf{w}\|^2 &= \| \begin{pmatrix} x_1+y_1 \\ \vdots \\ x_n+y_n \end{pmatrix} \|^2 \\ &= (x_1+y_1)^2 + \cdots + (x_n+y_n)^2 \\ &= (x_1^2+2x_1y_1+y_1^2) + \cdots + (x_n^2+2x_ny_n+y_n^2) \\ &= \|\mathbf{v}\|^2 + 2\sum_{j=1}^n x_jy_j + \|\mathbf{w}\|^2 \end{aligned}$$

Definition 9. Das *Standardskalarprodukt* auf $\mathbb{R}^n$ ist die Abbildung

$$\langle \quad , \quad \rangle \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

die durch die Formel

$$\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \rangle := x_1y_1 + \cdots + x_ny_n$$

gegeben ist.

Zwei Vektoren $\mathbf{v}, \mathbf{w}$ heißen *senkrecht* (oder *orthogonal*) bezüglich $\langle \quad , \quad \rangle$, wenn

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0.$$

Wir schreiben $\mathbf{v} \perp \mathbf{w}$ für „$\mathbf{v}$ *ist senkrecht zu* $\mathbf{w}$".

Die *Länge* (oder *Norm*) des Vektors $\mathbf{v} \in \mathbb{R}^n$ ist

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

Wir könnten die Formel auch als $\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \rangle = (x_1, \cdots, x_n) \circ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ also als Multiplikation einer $1 \times n$ mit einer $n \times 1$-Matrix schreiben.

Bemerkung. An der Definition der Länge eines Vektors können wir die folgenden Eigenschaften sofort ablesen:

1. $\|c \cdot \mathbf{v}\| = |c| \cdot \|\mathbf{v}\|$ für alle $c \in \mathbb{R}, \mathbf{v} \in \mathbb{R}^n$.

2. Ein Vektor hat genau dann die Länge 0, wenn $\mathbf{v}$ der 0 Vektor ist:

$$\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = 0.$$

wir werden uns gleich noch die Dreiecksungleichung

$$\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$$

überlegen, die wir für jeden Abstandsbegriff fordern würden.

Bemerkung. An der Definition des Standardskalarprodukts können wir die folgende Eigenschaften ablesen:

1. (Symmetrisch) Es gilt $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ für alle $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.

2. (Bilinear) Die Abbildung $\langle\quad,\quad\rangle$ ist in beiden Variablen linear:

$$\langle a\mathbf{v} + b\mathbf{v}', \mathbf{w}\rangle = a\langle \mathbf{v}, \mathbf{w}\rangle + b\langle \mathbf{v}', \mathbf{w}\rangle \text{ und}$$

$$\langle \mathbf{v}, a\mathbf{w} + b\mathbf{w}'\rangle = a\langle \mathbf{v}, \mathbf{w}\rangle + b\langle \mathbf{v}, \mathbf{w}'\rangle \text{für alle } a, b \in \mathbb{R}, \mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in \mathbb{R}^n.$$

Die Linearität in der zweiten Variablen folgt automatisch, wenn wir Symmetrie und Linearität in der ersten Variablen wissen.

3. (Positiv definit) Für alle $\mathbf{v} \in \mathbb{R}^n$ gilt

$$\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$$

und

$$\langle \mathbf{v}, \mathbf{v}\rangle = 0 \Leftrightarrow \mathbf{v} = 0.$$

WARNUNG: Ab jetzt müssen wir in Beweisen ausschließlich mit der obigen Definition von *senkrecht* argumentieren, dann das ist die einzige, die wir formal festlegen konnten. Die geometrische Vorstellung wird uns helfen, Ideen für Argumente zu finden, aber um einen Beweis aufzuschreiben müssen wir uns auf die Formeln verlassen.

Als erste Probe, sollten wir vielleicht noch einmal nachprüfen, dass der Satz des Pythagoras für unsere Definition jetzt wirklich stimmt. Wir hatten die Definition gerade so eingerichtet, aber die Rechnung können wir mit dem Skalarprodukt übersichtlicher aufschreiben.

Behauptung 10. *(Satz des Pythagoras für senkrechte Vektoren) Zwei Vektoren $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ sind genau dann senkrecht zueinander, wenn*

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 \textit{ gilt.}$$

Beweis. Wir rechnen die Länge der Summe mit der Definition aus:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}\rangle && \text{Definition der Länge} \\ &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{v} + \mathbf{w}, \mathbf{w}\rangle && \text{bilinear} \\ &= \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle && \text{bilinear} \\ &= \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w}\rangle + \|\mathbf{w}\|^2 && \text{symmetrisch.} \end{aligned}$$

Der Letzte Term ist genau dann gleich $\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$ wenn $\langle \mathbf{v}, \mathbf{w}\rangle = 0$ ist und das bedeutet gerade $\mathbf{v} \perp \mathbf{w}$. □

DIE GEOMETRISCHE INTERPRETATION von „senkrecht" liefert uns eine neue Interpretation für die Lösungsmenge einer linearen Gleichung

$$a_1x_1 + \cdots + a_nx_n = 0$$

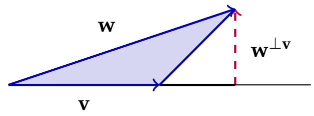
denn wenn wir $\mathbf{a} := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ schreiben ist diese Lösungsmenge

$$\left\{ \mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \mid \langle \mathbf{a}, \mathbf{v}\rangle = 0 \right\}$$

die Menge der Vektoren, die senkrecht zu **a** sind. Der Vektor der Koeffizienten der Gleichung ist also ein Normalenvektor für die Hyperebene die durch die Gleichung definiert wird.

Ende Vorlesung 19.4.

Bemerkung. Eine geometrische Interpretation des Skalarproduktes $\langle \mathbf{v}, \mathbf{w}\rangle$ können wir aus den Rechenregeln und der Interpretation $\langle \mathbf{v}, \mathbf{w}\rangle = 0 \Leftrightarrow \mathbf{v} \perp \mathbf{w}$ ableiten: Sind $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ mit $\mathbf{v} \neq 0$, so können wir $\mathbf{w}$ leicht als Summe $\mathbf{w} = a \cdot \mathbf{v} + \mathbf{w}^{\perp \mathbf{v}}$ schreiben, wobei $\mathbf{w}^{\perp \mathbf{v}} \in \mathbb{R}^n$ ein Vektor ist, der senkrecht zu $\mathbf{v}$ ist, indem wir bemerken, dass



$$\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}\right) \perp \mathbf{v},$$

denn

$$\begin{aligned}\langle \mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}, \mathbf{v}\rangle &= \langle \mathbf{w}, \mathbf{v}\rangle - \langle \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}, \mathbf{v}\rangle \\ &= \langle \mathbf{w}, \mathbf{v}\rangle - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\langle \mathbf{v}, \mathbf{v}\rangle = 0.\end{aligned}$$

Also haben wir

$$\begin{aligned}\mathbf{w} &= \underbrace{\frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}}_{\in \mathbb{R}\cdot \mathbf{v}} + \underbrace{\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}\right)}_{\perp \mathbf{v}} \\ &= \langle \mathbf{w}, \frac{\mathbf{v}}{\|\mathbf{v}\|}\rangle \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} + (\mathbf{w} - \langle \mathbf{w}, \frac{\mathbf{v}}{\|\mathbf{v}\|}\rangle \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|}).\end{aligned}$$

Der Ausdruck

$$\langle \mathbf{w}, \frac{\mathbf{v}}{\|\mathbf{v}\|}\rangle$$

ist also die Länge der orthogonalen Projektion von $\mathbf{w}$ auf $\mathbf{v}$, d.h. $\langle \mathbf{w}, \mathbf{v}\rangle$ ist die Länge der orthogonalen Projektion von $\mathbf{w}$ multipliziert mit der Länge von $\mathbf{v}$, kurz:

Hat $\mathbf{v}$ Länge 1, so ist
$\langle \mathbf{v}, \mathbf{w}\rangle$ die Länge der orthogonalen Projektion von $\mathbf{w}$ auf $\mathbf{v}$.

Anders formuliert sollten wir $\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{v}\|\|w\|} =: \cos(\measuredangle(\mathbf{v}, \mathbf{w}))$ als Cosinus des Winkels zwischen $\mathbf{v}, \mathbf{w}$ interpretieren können und damit auch eine mögliche Definition dafür finden, was wir mit Winkeln genau meinen. Dafür müssten wir aber insbesondere wissen, dass diese Zahl überhaupt zwischen -1 und 1 liegt. Das überlegen wir uns gleich noch.

Jetzt sollten wir uns noch die Dreiecksungleichung überlegen. Wir wissen schon, dass

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w}\rangle + \|\mathbf{w}\|^2$$

für alle $\mathbf{v}, \mathbf{w}$ gilt. Wir müssen für die Dreiecksungleichung also zeigen, dass

$$|\langle \mathbf{v}, \mathbf{w}\rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$$

gilt. Diese Ungleichung heißt Cauchy-Schwarz Ungleichung.

Lemma 11 (Cauchy-Schwarz-Ungleichung). *Für alle* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ *gilt*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$$

und Gleichheit gilt genau dann, wenn die Vektoren $\mathbf{v}, \mathbf{w}$ *linear abhängig sind.*

Die Cauchy-Schwarz-Ungleichung ist für das Skalarprodukt, was die Dreiecksungleichung für die Norm ist.

Beweis. Wenn $\mathbf{v} = 0$ gilt, so stimmt die Aussage, da in diesem Fall beide Seiten der Ungleichung 0 sind und insbesondere $\mathbf{v} = 0\mathbf{w}$. Ist $\mathbf{v} \neq \mathbf{w}$ so können wir $\mathbf{w}$ wie oben in die orthogonale Projektion von $\mathbf{w}$ auf $\mathbf{v}$ und einen dazu senkrechten Vektor zerlegen:

$$\mathbf{w} = \underbrace{\frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}}_{\in \mathbb{R}\cdot \mathbf{v}} + \underbrace{\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)}_{\perp \mathbf{v}}.$$

Merkhilfe: Es gibt viele Beweise für die Ungleichung, manche mit Trick, dieser kommt fast ohne Trick aus. Ich merke mir das als:
„Die Cauchy-Schwarz Ungleichung können wir beweisen, indem wir einen der beiden Vektoren in seine orthogonale Projektion auf den anderen und den dazu senkrechten Vektor zerlegen“.

Da die beiden Summanden senkrecht zueinander sind, können wir damit die Länge von $\mathbf{w}$ berechnen als

$$\begin{aligned}
\|\mathbf{w}\|^2 &= \|\frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\|^2 + \|\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)\|^2 \\
&= \frac{\langle \mathbf{w}, \mathbf{v} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle^2}\|\mathbf{v}\|^2 + \|\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)\|^2 \\
&= \frac{\langle \mathbf{w}, \mathbf{v} \rangle^2}{\langle \mathbf{v}, \mathbf{v} \rangle} + \underbrace{\|\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)\|^2}_{\geq 0}.
\end{aligned}$$

Multiplizieren wir Diese Gleichung mit $\|\mathbf{v}\|^2$ erhalten wir äquivalent:

$$\|\mathbf{v}\|^2\|\mathbf{w}\|^2 = \langle \mathbf{w}, \mathbf{v} \rangle^2 + \underbrace{\|\mathbf{v}\|^2\|\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)\|^2}_{\geq 0}.$$

Also gilt in der Tat:

$$\|\mathbf{v}\|\|\mathbf{w}\| \geq |\langle \mathbf{w}, \mathbf{v} \rangle|$$

und Gleichheit gilt genau dann, wenn $\|\left(\mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}\right)\| = 0$, also $\mathbf{w} = \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}$ gilt. □

Ein kleines Wunder passiert, wenn wir bemerken, dass wir den Beweis zwar mit einer geometrischen Überlegung gefunden haben, im Beweis dann aber nur die Rechenregeln für das Skalarprodukt verwendet haben.

Eigenschaften von $\langle\ ,\ \rangle$:

1. (Symmetrisch) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
2. (Bilinear) Die Abbildung $\langle\ ,\ \rangle$ ist in beiden Variablen linear.
3. (Positiv definit)
$$\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = 0.$$

Wenn wir zum Beispiel das Integral von z.B. stetigen Funktionen, oder einfach von Polynomen über ein Intervall betrachten

$$(f,g) \mapsto \int_0^1 f(x) \cdot g(x)dx =: \langle f, g \rangle$$

dann ist diese Zuordnung ebenfalls sichtbar symmetrisch, bilinear und $\langle f, f \rangle = \int_0^1 f(x)^2dx$ ist als Integral über eine stetige Funktion deren Funktionswerte alle ≥ 0 sind nur dann 0 wenn $f(x) = 0$ für alle $0 \leq x \leq 1$.

Damit können wir die Argumente von oben abschreiben und erhalten:

$$|\langle f,g\rangle| = |\int_0^1 f(x)\cdot g(x)dx| \leq \|f\|\cdot\|g\|$$

wobei $\|f\| = \sqrt{\langle f,f\rangle} = \sqrt{\int_0^1 f(x)^2 dx}$ ist.

Ohne die Geometrie wären wir aber vielleicht nicht darauf gekommen für $f \neq 0$ die „Projektion von g auf f", also $\frac{\langle f,g\rangle}{\langle f,f\rangle}\cdot f$ anzuschauen, zumal dieser Ausdruck ohne die Abkürzende Schreibweise $\langle f,g\rangle$ für das Integral furchtbar aussieht!

Das ist vielleicht ein guter Grund, um Abbildungen mit den Eigenschaften des Standardskalarproduktes einen eigenen Namen zu geben. Das machen wir im nächsten Abschnitt.

Exkurs: Skalarprodukte auf reellen Vektorräumen

Wie gerade gesehen, können wir geometrische Ideen, die wir in Termen der Eigenschaften des Standardskalarproduktes formulieren und beweisen können, auf Problem übertragen, die nicht geometrisch aussehen. Deshalb geben wir dieser Struktur einen eigenen Namen.

Definition 12. Ein *Skalarprodukt* auf einem $\mathbb{R}$-Vektorraum V ist eine Abbildung

$$\begin{aligned}\langle\ ,\ \rangle\colon V\times V &\to \mathbb{R}\\ (\mathbf{v},\mathbf{w}) &\mapsto \langle\mathbf{v},\mathbf{w}\rangle\end{aligned}$$

für die die folgenden Rechenregeln gelten:

1. (Symmetrisch) Für alle $\mathbf{v},\mathbf{w}\in V$ gilt $\langle\mathbf{v},\mathbf{w}\rangle = \langle\mathbf{w},\mathbf{v}\rangle$.
2. (Bilinear) Die Abbildung $\langle\ ,\ \rangle$ ist in beiden Variablen linear:

$$\langle a\mathbf{v}+b\mathbf{v}',\mathbf{w}\rangle = a\langle\mathbf{v},\mathbf{w}\rangle + b\langle\mathbf{v}',\mathbf{w}\rangle \text{ und}$$

$$\langle\mathbf{v},a\mathbf{w}+b\mathbf{w}'\rangle = a\langle\mathbf{v},\mathbf{w}\rangle + b\langle\mathbf{v},\mathbf{w}'\rangle \text{für alle } a,b\in\mathbb{R},\mathbf{v},\mathbf{v}',\mathbf{w},\mathbf{w}'\in V.$$

Die Linearität in der zweiten Variablen folgt automatisch, wenn wir Symmetrie und Linearität in der ersten Variablen wissen.

3. (Positiv definit) Für alle $\mathbf{v}\in V$ gilt

$$\langle\mathbf{v},\mathbf{v}\rangle \geq 0$$

und

$$\langle\mathbf{v},\mathbf{v}\rangle = 0 \Leftrightarrow \mathbf{v} = 0.$$

Ein *euklidischer Vektorraum* ist ein $\mathbb{R}$-Vektorrraum V, zusammen mit einem Skalarprodukt $\langle\ ,\ \rangle$.

Beispiel 13. Wir hatten zwei Beispiele gesehen

1. Unser Standardskalarprodukt $\langle\ ,\ \rangle\colon \mathbb{R}^n\to\mathbb{R}^n$.
2. Das Integral auf dem Raum der reellen Polynome $V=\mathbb{R}[t]$ mit

$$\langle p,q\rangle := \int_0^1 p(x)\cdot q(x)dx$$

Hier könnten wir die Integrationsgrenzen natürlich auch anders wählen $\int_{-1}^{1}$ oder $\int_{0}^{2\pi}$, das würde jeweils andere Skalarprodukte liefern.

Bemerkung. Zu jedem Skalarprodukt $\langle \quad , \quad \rangle$ definieren wir

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

und nennen das die Norm (oder Länge) bezüglich des Skalarproduktes.

Beispiel 14. Die Norm die auf dem Raum der Polynome (oder dem Raum der stetigen Funktionen) durch das Integral

$$\|f\|^2 = \int_a^b f(x)^2 dx$$

definiert wird, liefert einen guten Abstandsbegriff für Funktionen denn

$$\|f - g\|^2 = \int_a^b \big(f(x) - g(x)\big)^2 dx$$

ist nur dann klein, wenn die Fläche zwischen den Funktionsgraphen von f, g klein ist. Daher werden uns Normen dieser Art zum Beispiel helfen, wenn wir eine komplizierte Funktion durch Funktionen einer einfachen Form, z.b. durch Polynome approximieren wollen.

Behauptung 15. *In jedem euklidischen Vektorraum $V, \langle \quad , \quad \rangle$ gilt die Cauchy-Schwarz Ungleichung:*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$$

für alle $\mathbf{v}, \mathbf{w} \in V$ und darum auch die Dreiecksungleichung

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Beweis. Unser Beweis für das Standardskalarprodukt im $\mathbb{R}^n$ hat nur die Rechenregeln für Skalarprodukte verwendet und funktioniert darum wortwörtlich für euklidische Vektorräume. $\square$

Bemerkung. Der Satz des Pythagoras gilt genauso in euklidischen Vektorräumen: Zwei Vektoren $\mathbf{v}, \mathbf{w}$ in einem euklidischen Vektorraum sind genau dann senkrecht zueinander, wenn

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 \text{ gilt.}$$

Denn auch in diesem Beweis haben wir nur die 3 Eigenschaften des Skalarproduktes verwendet.

Orthogonale Vektoren, Basen und Unterräume

Lassen Sie uns nun zum $\mathbb{R}^n$ zurückkehren. Dort ist meine geometrische Vorstellung von linear abhängigen Vektoren so, dass ich glauben würde, dass paarweise orthogonale Vektoren immer linear unabhängig sein sollten. Diese Intuition möchte ich einmal formal überprüfen. Um sicher zu sein, dass ich im Beweis wirklich nur mit den Rechenregeln arbeite formuliere ich die Aussage allgemein.

Behauptung 16 (Orthogonale Vektoren sind linear unabhängig). *Sei $V, \langle\ ,\ \rangle$ ein euklidischer Vektorraum und $\mathbf{v}_1, \dots, \mathbf{v}_k \in V \setminus \{0\}$ für die $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ für alle $i \neq j$ gilt. Dann sind die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_n$ linear unabhängig.*

Beweis. Es ist zu zeigen, dass die Gleichung $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = 0$ nur die triviale Lösung $a_1 = \cdots = a_k = 0$ besitzt. Wir wissen dabei nur $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ für alle $i \neq j$.

Ist $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = 0$ wenden wir darum einmal $\langle\ , \mathbf{v}_i \rangle$ auf den Vektor $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$ an und finden für jedes i einerseits:

$$\langle a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k, \mathbf{v}_i \rangle = \langle 0, \mathbf{v}_i \rangle = 0$$

und andererseits

$$\begin{aligned} \langle a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k, \mathbf{v}_i \rangle &= a_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + \cdots + a_k\langle \mathbf{v}_k, \mathbf{v}_i \rangle && \text{bilinear} \\ &= a_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle && \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \text{ für } i \neq j. \end{aligned}$$

Da aber $\langle \mathbf{v}_i, \mathbf{v}_i \rangle > 0$ gilt, folgt daraus $a_i = 0$. Also sind in der Tat alle $a_i = 0$ und das war zu zeigen. □

Das Argument aus diesem Beweis liefert uns für eine Linearkombination $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$ von orthogonalen Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_k$ eine einfache Formel für die Koeffizienten der Linearkombination

$$\langle \mathbf{v}, \mathbf{v}_i \rangle = a_i \|\mathbf{v}_i\|^2.$$

Für eine Basis aus orthogonalen Vektoren der Länge 1, können wir die Koordinaten bezüglich der Basis sofort als Skalarprodukt $\langle \mathbf{v}, \mathbf{v}_i \rangle$ ausrechnen! Solche Basen bekommen daher einen Namen.

Definition. Eine Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ eines euklidischen Vektorraums $V, \langle\ ,\ \rangle$ heißt *Orthonormalbasis* wenn

1. die Vektoren paarweise orthogonal sind, d.h.

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \text{ für } i \neq j \text{ und}$$

2. alle die Länge 1 haben, d.h.

$$\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1 \text{ für alle } i.$$

Folgerung 17. *Ist $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Orthogonalbasis eines euklidischen Vektorraums $V, \langle\ ,\ \rangle$, so sind die Koordinaten eines Vektors $\mathbf{v}$ bezüglich der Basis die Skalarprodukte $\langle \mathbf{v}, \mathbf{v}_i \rangle$, d.h.*

$$\begin{aligned} \mathrm{Koord}_{\mathbf{v}} \colon V &\to \mathbb{R}^n \\ \mathbf{v} &\mapsto \mathrm{Koord}_{\mathbf{v}}(\mathbf{v}) = \begin{pmatrix} \langle \mathbf{v}, \mathbf{v}_1 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{v}_n \rangle \end{pmatrix}. \end{aligned}$$

Beweis. Das haben wir uns vor der Definition von Orthonormalbasis überlegt. Es war der Grund, warum wir diesen Begriff eingeführt haben. □

Beispiel 18. Für die Ebene in $\mathbb{R}^3$ die durch die Gleichung $x_1 + x_2 + x_3 = 0$ gegeben ist, sind unsere Lieblingsbasen $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ oder $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ leider beide nicht orthogonal, aber $\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$ ist orthogonal zu $\begin{pmatrix} 1 \\ -1 \\ -0 \end{pmatrix}$, also ist $\mathbf{v}_1' = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2' = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$ zusammen mit dem Normalenvektor $\mathbf{v}_3' = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ der Ebene eine orthogonale Basis von $\mathbb{R}^3$. Die Längen der Vektoren sind

$$\|\mathbf{v}_1\| = \sqrt{2}, \|\mathbf{v}_2\| = \sqrt{6}, \|\mathbf{v}_3\| = \sqrt{3},$$

also ist $\mathbf{v}_1 := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}, \mathbf{v}_3 := \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ eine Orthonormalbasis. Die Koordinaten von $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ bezüglich dieser Basis können wir also ohne ein Gleichungssystem zu lösen ablesen und finden

$$\text{Koord}_{\mathbf{v}}(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}) = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{-3}{\sqrt{6}} \\ \frac{6}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{\sqrt{3}}{\sqrt{2}} \\ 2\sqrt{3} \end{pmatrix}.$$

Die Folgerung zur Berechnung der Koordinaten können wir auch in Termen von Matrizen umformulieren:

Folgerung 19 (Orthogonale Matrizen). *Eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ ist genau dann eine Orthonormalbasis bezüglich des Standardskalarproduktes, wenn für die Matrix $A = (\mathbf{v}_1 \ldots \mathbf{v}_n)$ mit Spaltenvektoren $\mathbf{v}_i$ gilt dass*

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

Beweis. Der (i,j)-te Eintrag der Matrix $A^t \circ A = (\mathbf{v}_1 \ldots \mathbf{v}_n)^t \circ (\mathbf{v}_1 \ldots \mathbf{v}_n)$ ist per Definition genau $\mathbf{v}_i^t \circ \mathbf{v}_j = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$. Dieses Produkt ist also genau dann die Einheitsmatrix wenn $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ für $i \neq j$ und $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 1$ für $i = j$ ist.

Das ist genau die Bedingung dafür, dass $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Orthonormalbasis ist. □

Mit Orthonormalbasen rechnet es sich also leichter.

ERINNERUNG: Wir hatten uns überlegt, dass die Lösungsmenge einer homogenen linearen Gleichung genau die Vektoren sind, die senkrecht zum Koeffizientenvektor sind. Genauso sehen wir, dass die Lösungsmenge eines Gleichungssystems

$$A\mathbf{v} = 0$$

genau aus den Vektoren $\mathbf{v}$ besteht, die senkrecht zu allen Zeilenvektoren von A – also den Spalten von A^t – sind. Es wäre also praktisch, ein Verfahren zu haben, dass aus einer Liste von Vektoren eine System orthogonaler Vektoren macht, das den gleichen Vektorraum aufspannt.

Für zwei Vektoren $\mathbf{v}, \mathbf{w}$ hatten wir uns schon überlegt, dass der Vektor $\mathbf{w} - \frac{\langle \mathbf{w},\mathbf{v}\rangle}{\langle \mathbf{v},\mathbf{v}\rangle}\mathbf{v}$ senkrecht zu $\mathbf{v}$ ist und wir wissen auch, dass

$$\mathrm{Span}(\mathbf{v}, \mathbf{w}) = \mathrm{Span}(\mathbf{v}, \mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}).$$

Frage. *Wie können wir das für 3 linear unabhängige Vektoren $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ erreichen?*

Wenn Ihnen allgemein nichts einfällt, könnten Sie ein konkretes Beispiel ausprobieren:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 4 \\ 5 \end{pmatrix} \in \mathbb{R}^4.$$

Bevor Sie weiterlesen sollten Sie das einmal selbst versuchen.

Wenn Sie eine Methode zum Orthogonalisieren von 3 Vektoren gefunden haben, können Sie das Argument höchstwahrscheinlich auch induktiv für m Vektoren verwenden.

Satz 20 (Gram-Schmidt-Verfahren). *Sind $\mathbf{w}_1, \dots, \mathbf{w}_m$ linear unabhängige Vektoren eines euklidischen Vektorraums $V, \langle,\rangle$ dann sind die induktiv definierten Vektoren*

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1 \\
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\langle \mathbf{w}_2, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{w}_3, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 - \frac{\langle \mathbf{w}_3, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 \\
\mathbf{v}_i &:= \mathbf{w}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{w}_i, \mathbf{v}_j\rangle}{\langle \mathbf{v}_j, \mathbf{v}_j\rangle}\mathbf{v}_j \quad \text{für } i = 2, \dots, m
\end{aligned}$$

eine Basis von $W = \operatorname{Span}(\mathbf{w}_1, \dots, \mathbf{w}_m)$, die aus orthogonalen Vektoren besteht und die Vektoren $\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \dots, \frac{\mathbf{v}_m}{\|\mathbf{v}_m\|}$ sind eine Orthonormalbasis von W.

Der Grund, warum wir nicht gleich eine Formel für die Orthonormalbasis angeben ist einfach, dass in den Längen $\|\mathbf{v}_i\|$ Wurzeln versteckt sind und es bei der Berechnung angenehmer ist, wenn erst im letzten Schritt Wurzeln auftauchen.

Beweis. Wir argumentieren induktiv, genau wie im Beispiel: Für $m = 1$ ist nichts zu zeigen. Induktionsschritt: Gilt die Aussage für k Vektoren, so auch für $k + 1$:

Sind $\mathbf{w}_1, \dots, \mathbf{w}_{k+1}$ linear unabhängig, so wissen wir nach Induktionsannahme schon, dass die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_k$ orthogonal sind. Wir müssen darum nur noch $\langle \mathbf{v}_{k+1}, \mathbf{v}_i\rangle$ für $i = 1, \dots k$ berechnen:

$$\begin{aligned}
\langle \mathbf{v}_{k+1}, \mathbf{v}_i\rangle &= \Big\langle \mathbf{w}_{k+1} - \sum_{j=1}^{k} \frac{\langle \mathbf{w}_{k+1}, \mathbf{v}_j\rangle}{\langle \mathbf{v}_j, \mathbf{v}_j\rangle}\mathbf{v}_j, \mathbf{v}_i\Big\rangle && \text{Def } \mathbf{v}_{k+1} \\
&= \langle \mathbf{w}_{k+1}, \mathbf{v}_i\rangle - \sum_{j=1}^{k} \frac{\langle \mathbf{w}_{k+1}, \mathbf{v}_j\rangle}{\langle \mathbf{v}_j, \mathbf{v}_j\rangle}\langle \mathbf{v}_j, \mathbf{v}_i\rangle && \langle,\rangle \text{ bilinear.} \\
&= \langle \mathbf{w}_{k+1}, \mathbf{v}_i\rangle - \frac{\langle \mathbf{w}_{k+1}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}\langle \mathbf{v}_i, \mathbf{v}_i\rangle && \begin{array}{c}\text{nur Summand } j = i \text{ übrig} \\ \text{da } \langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0 \text{ für } i \neq j\end{array} \\
&= 0.
\end{aligned}$$

Also sind die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ orthogonal.

Nach Induktion wissen wir außerdem, dass $\operatorname{Span}(\mathbf{w}_1, \dots, \mathbf{w}_k) = \operatorname{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ und also gilt

$$\begin{aligned}
\operatorname{Span}(\mathbf{w}_1, \dots, \mathbf{w}_{k+1}) &= \operatorname{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_{k+1}) \\
&= \operatorname{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1})
\end{aligned}$$

nach dem Austauschsatz weil $\mathbf{v}_{k+1}$ von der Form $\mathbf{v}_{k+1} = \mathbf{w}_{k+1} + \sum_{j=1}^{k} a_j \mathbf{v}_k$ ist. □

Bemerkung. Der Satz zeigt insbesondere, dass wir aus Basis eines euklidischen Vektorraums eine Orthogonalbasis berechnen können. Insbesondere besitzt jeder endlichdimensionale euklidische Vektorraum eine Orthonormalbasis.

Das ist für unser Standardbeispiel $\mathbb{R}^n, \langle,\rangle$ klar, aber es ist vielleicht ganz schön, dass unser konkretes Problem für Unterräume auch ein Resultat für abstrakte euklidische Vektorräume frei Haus liefert.

Beispiel 21. Lassen Sie uns das Verfahren einmal im euklidischen Vektorraum $\mathbb{R}[x]_{\leq 2}$ mit Skalarprodukt $\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)dx$ ausprobieren indem wir die Basis $\mathbf{w}_1 = 1, \mathbf{w}_2 = x, \mathbf{w}_3 = x^2$ orthogonalisieren:

$$\langle \mathbf{w}_1, \mathbf{w}_1\rangle = \langle 1, 1\rangle = \int_{-1}^{1} 1dx = 2. \langle \mathbf{w}_2, \mathbf{w}_1\rangle \quad = \langle x, 1\rangle = \int_{-1}^{1} 1\cdot xdx = 0.$$

Die Vektoren $\mathbf{v}_1 = \mathbf{w}_1 = 1$ und $\mathbf{v}_2 = \mathbf{w}_2 - 0\cdot \mathbf{w}_1 = \mathbf{w}_2 = x$ sind also bereits orthogonal.

$$\begin{aligned}\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{w}_3, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 - \frac{\langle \mathbf{w}_3, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 \\ &= x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle}1 - \frac{\langle x^2, x\rangle}{\langle x, x\rangle}x.\end{aligned}$$

Die Skalarprodukte in dieser Formel sind:

$$\begin{aligned}\langle x^2, 1\rangle &= \int_{-1}^{1} x^2 dx = \left[\frac{1}{3}x^3\right]_{-1}^{1} = \frac{2}{3} \\ \langle x^2, x\rangle &= \int_{-1}^{1} x^3 dx = \left[\frac{1}{4}x^4\right]_{-1}^{1} = 0 \\ \langle x, x\rangle &= \int_{-1}^{1} x^2 dx = \frac{2}{3}\end{aligned}$$

also

$$\begin{aligned}\mathbf{v}_3 &= x^2 - \frac{\langle x^2, 1\rangle}{\langle 1, 1\rangle}\cdot 1 - \frac{\langle x^2, x\rangle}{\langle x, x\rangle}x \\ &= x^2 - \frac{2}{3\cdot 2}\cdot 1 - 0x \\ &= x^2 - \frac{1}{3}.\end{aligned}$$

Wir haben also die orthogonale Basis:

$$\mathbf{v}_1 = 1, \mathbf{v}_2 = x, \mathbf{v}_3 = x^2 - \frac{1}{3}$$

gefunden. Für eine orthonormale Basis müssten wir noch die Längen der Vektoren ausrechnen:

$$\begin{aligned}\|\mathbf{v}_1\| &= \sqrt{\langle 1, 1\rangle} = \sqrt{2} \\ \|\mathbf{v}_2\| &= \sqrt{\langle x, x\rangle} = \sqrt{\frac{2}{3}} \\ \|\mathbf{v}_3\| &= \sqrt{\int_{-1}^{1}(x^2 - \frac{1}{3})^2 dx} = \sqrt{\int_{-1}^{1} x^4 - \frac{2}{3}x^2 + \frac{1}{9}dx} \\ &= \sqrt{\left[\frac{1}{5}x^5 - \frac{2}{9}x^3 + \frac{1}{9}x\right]_{-1}^{1}} = \sqrt{\frac{2}{5} - \frac{4}{9} + \frac{2}{9}} = \frac{2\sqrt{2}}{3\sqrt{5}}.\end{aligned}$$

Die Vektoren

$$\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{2}}, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{\sqrt{3}}{\sqrt{2}}x, \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{3\sqrt{5}}{2\sqrt{2}}(x^2 - \frac{1}{3})$$

sind also eine Orthonormalbasis.

Sie sehen, dass die normalisierten Vektoren etwas unhandliche Koeffizienten bekommen. Zum Rechnen per Hand sind orthogonale Vektoren meist angenehmer, zum Argumentieren sind orthonormale Vektoren praktischer.

Anwendung: Projektion auf einen Unterraum

Mit einer Orthonormalbasis eines Unterraum $W \subset V$ können wir jetzt analog zum Fall von der Projektion eines Vektors auf die Gerade die durch einen anderen gegeben war, nun für jeden Vektor $\mathbf{v} \in V$ den eindeutigen Vektor $\mathrm{pr}_W(\mathbf{v}) \in W$ berechnen, der den kürzesten Abstand zu $\mathbf{v}$ hat.

Behauptung 22 (Orthogonale Projektion auf einen Unterraum)**.** *Ist $W \subset V$ ein endlichdimensionaler Unterraum eines euklidischen Vektorraums $V\langle,\rangle$ und $\mathbf{w}_1, \ldots, \mathbf{w}_m$ eine Orthonormalbasis von W. Dann berechnet die lineare Abbildung*

$$\begin{aligned} \mathrm{pr}_W \colon V &\to W \\ \mathbf{v} &\mapsto \mathrm{pr}_W(\mathbf{v}) := \sum_{i=1}^{m} \langle \mathbf{v}, \mathbf{w}_i \rangle \mathbf{w}_i \end{aligned}$$

das eindeutige Element von W, für das $\|\mathbf{v} - \mathbf{w}\|$ minimal ist. Insbesondere hängt die Abbildung pr_W nicht von der Wahl der Orthonormalbasis von W ab.

Beweis. Wir müssen für alle $\mathbf{w} \in W$ zeigen, dass $\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - \mathrm{pr}_W(\mathbf{v})\|$ gilt und $=$ nur für $\mathbf{w} = \mathrm{pr}_W(\mathbf{v})$ erfüllt ist.

Per Konstruktion ist der Vektor $\mathbf{v} - \mathrm{pr}_W(\mathbf{v}) = \mathbf{v} - \sum_{i=1}^{m} \langle \mathbf{v}, \mathbf{w}_i \rangle \mathbf{w}_i$ orthogonal zu allen $\mathbf{w}_i$, also auch orthogonal zu allen Elementen von $W = \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_m)$.

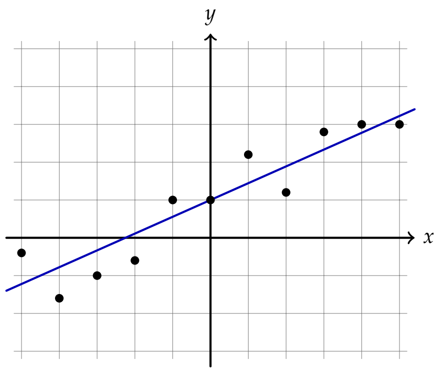
Darum ist nach dem Satz des Pythagoras in euklidischen Vektorräumen:

$$\begin{aligned} \|\mathbf{v} - \mathbf{w}\|^2 &= \|\underbrace{\mathbf{v} - \mathrm{pr}_W(\mathbf{v})}_{\perp W} + \underbrace{\mathrm{pr}_W(\mathbf{v}) - \mathbf{w}}_{\in W}\|^2 \\ &= \|\mathbf{v} - \mathrm{pr}_W(\mathbf{v})\|^2 + \|\mathrm{pr}_W(\mathbf{v}) - \mathbf{w}\|^2. \end{aligned}$$

Da $\|\mathrm{pr}_W(\mathbf{v}) - \mathbf{w}\| \geq 0$ und $=$ nur für $\mathrm{pr}_W(\mathbf{v}) = \mathbf{w}$ gilt zeigt das die Behauptung. $\square$

Anwendung: Approximation von Datenpunkten

Sie haben vielleicht schon einmal eine Grafik einer linearen Regression gesehen, d.h. ein Bild in dem eine Reihe von Messwerten $M(i)$ eingetragen wurde und dazu eine Gerade, die die Messwerte passabel approximiert.



So eine Approximation (durch eine Gerade oder ein Polynom höheren Grades) können wir jetzt leicht ausrechnen:

Zu gegebenen Funktionswerten $f(-n), \dots, f(-1), f(0), f(1), \dots, f(n)$ mit $n \geq 1$ suchen wir eine Gerade $G(x) = ax + b$, so dass

$$\left\| \begin{pmatrix} f(-n) \\ \vdots \\ f(n) \end{pmatrix} - \begin{pmatrix} G(-n)) \\ \vdots \\ G(n) \end{pmatrix} \right\|$$

minimal wird.

Weil die Menge der Geraden $\mathbb{R}[x]_{\leq 1} = \{G(x) = ax + b \,|\, a, b \in \mathbb{R}\} \subseteq \mathbb{R}[x]$ ein Untervektorraum und die Auswertungsabbildung

$$\begin{aligned} \mathrm{ev} \colon \mathbb{R}[x] &\to \mathbb{R}^{2n+1} \\ p(x) &\mapsto \begin{pmatrix} p(-n) \\ \vdots \\ p(n) \end{pmatrix} \end{aligned}$$

linear ist, ist das Bild $W := \{ \mathrm{ev}(G) \in \mathbb{R}^{2n+1} | G(x) \in \mathbb{R}[x]_{\leq 1}\}$ ein Unterraum.

Eine Orthogonalbasis ist zum Beispiel durch die Vektoren

$$\mathbf{v}_1 := \mathrm{ev}(1) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{v}_2 := \mathrm{ev}(x) = \begin{pmatrix} -n \\ -n+1 \\ \vdots \\ n-1 \\ n \end{pmatrix}$$

gegeben, wobei $\|\mathbf{v}_1\| = \sqrt{2n+1}, \|\mathbf{v}_2\| = \sqrt{n^2 + \cdots + 1 + 0 + 1 + \cdots + n^2}$.

Für einen gegebenen Vektor $f = \left(\begin{pmatrix} f(-n) \\ \vdots \\ f(n) \end{pmatrix} \right)$ ist nach der Formel für die Projektion auf den Unterraum $W \subset \mathbb{R}^{2n+1}$ also die best approximierende Gerade durch

$$\frac{\langle f, \mathrm{ev}(1) \rangle}{\langle \mathrm{ev}(1), \mathrm{ev}(1) \rangle} \cdot 1 + \frac{\langle f, \mathrm{ev}(x) \rangle}{\langle \mathrm{ev}(x), \mathrm{ev}(x) \rangle} \cdot x$$

gegeben.

Es ist vielleicht überraschend, dass wir aus dem Resultat ohne großen Aufwand eine einfache Formel für die approximierende Gerade finden konnten. Und noch besser ist, dass die Methode uns genauso erlauben würde durch Polynome höherer Ordnung zu approximieren. Wenn wir zuvor mit dem Orthogonalisierungsverfahren eine Folge von orthogonalen Polynomen konstruieren finden wir damit eine Folge von Approximationen, bei denen in jedem Schritt nur ein Vielfaches des jeweils nächsten Polynoms hinzu kommt.

Es gibt für verschiedene Skalarprodukte einen ganzen Zoo an bekannten Folgen orthogonaler Polynome, die für verschiedene numerische Anwendungen genutzt werden.

Die QR-Zerlegung – Eine Umformulierung der Orthogonalisierungsmethode

Die Formel zur Orthogonalisierung von linear unabhängigen Vektoren $\mathbf{w}_1, \ldots, \mathbf{w}_m$ in einem euklidischen Vektorraum $V, \langle,\rangle$ war:

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1 \\
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\langle \mathbf{w}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\langle \mathbf{w}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{w}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 \\
\mathbf{v}_i &:= \mathbf{w}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{w}_i, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j \quad \text{für } i = 2, \ldots, m.
\end{aligned}$$

D.h. die Abbildung, die die Basis $\mathbf{w}_1, \ldots, \mathbf{w}_i$ aus den $\mathbf{v}_1, \ldots, \mathbf{v}_m$ berechnet, also

$$\mathbf{w}_i = \mathbf{v}_i + \sum_{j=1}^{i-1} \frac{\langle \mathbf{w}_i, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j$$

ist eine obere Dreiecksmatrix und genauso gilt das für die umgekehrte Abbildung. In Termen der Orthonormalbasis $\mathbf{n}_i := \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$ bedeutet das

$$\mathbf{w}_i = \|\mathbf{v}_i\| \cdot \mathbf{n}_i + \sum_{j=1}^{i-1} \langle \mathbf{w}_i, \mathbf{n}_j \rangle \cdot \mathbf{n}_j,$$

also

$$(\mathbf{w}_1 \ldots \mathbf{w}_m) = (\mathbf{n}_1 \ldots \mathbf{n}_m) \circ \begin{pmatrix} \|\mathbf{v}_1\| & \langle \mathbf{w}_2, \mathbf{n}_1 \rangle & \langle \mathbf{w}_3, \mathbf{n}_1 \rangle & \ldots & \langle \mathbf{w}_m, \mathbf{n}_1 \rangle \\ 0 & \|\mathbf{v}_2\| & \langle \mathbf{w}_3, \mathbf{n}_2 \rangle & \ldots & \langle \mathbf{w}_m, \mathbf{n}_2 \rangle \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \|\mathbf{v}_{m-1}\| & \langle \mathbf{w}_m, \mathbf{n}_{m-1} \rangle \\ 0 & \ldots & \ldots & \ldots & \|\mathbf{v}_m\| \end{pmatrix}.$$

Diese Aussage ist so nützlich, dass sie einen eigenen Namen bekommen hat:

Satz 23 (QR-Zerlegung)**.** *Ist A eine reelle $n \times m$-Matrix mit linear unabhängigen Spalten, so existiert eine $m \times m$-obere Dreiecksmatrix mit positiven Diagonaleinträgen R und eine $n \times m$-Matrix Q mit orthonormalen Spaltenvektoren, so dass*

$$A = Q \circ R.$$

Die Matrizen Q, R lassen sich hierbei berechnen, indem wir das Gram-Schmidtsche Orthogonalisierungsverfahren auf die Spaltenvektoren der Matrix A anwenden.

Aufgabe 1. Sie können sich mittels Induktion davon überzeugen, dass die Matrizen Q, R der Zerlegung eindeutig bestimmt sind.

Die QR-Zerlegung ist praktisch, um ähnlich wie bei Regressionsproblemen beste Approximationen von Lösungen für überbestimmte Gleichungssysteme zu finden. Ist A eine reelle $n \times m$-Matrix mit linear unabhängigen Spalten und $\mathbf{b} \in \mathbb{R}^n$ gegeben, so hat das Gleichungssystem

$$A\mathbf{x} = \mathbf{b}$$

in der Regel keine Lösung $\mathbf{x} \in \mathbb{R}^m$, aber wir können ein $\mathbf{x}$ suchen, für das $\|A\mathbf{x} - \mathbf{b}\|$ minimal ist. Wenden wir die Aussage zur Projektion auf Unterräume auf den Unterraum $\mathrm{Bild}(A) = \mathrm{Span}(\mathrm{Spalten}(A))$ und den Vektor $\mathbf{b}$ an, so sehen wir, dass $\|A\mathbf{x} - \mathbf{b}\|$ genau dann minimal ist, wenn $A\mathbf{x} = \mathrm{pr}_{\mathrm{Bild}(A)}(\mathbf{b})$ die Projektion von $\mathbf{b}$ auf das Bild von A ist.

In der QR-Zerlegung $A = Q \circ R$ sind die Spalten $\mathbf{n}_1, \dots, \mathbf{n}_m$ von Q eine Orthonormalbasis des Bildes von A. Die Koordinaten von $\mathrm{pr}_{\mathrm{Bild}(A)}(\mathbf{b})$ bezüglich dieser Basis sind genau durch den Vektor $Q^t\mathbf{b} \in \mathbb{R}^m$ gegeben, denn $\mathrm{pr}_{\mathrm{Bild}(A)}(\mathbf{b}) = \sum_{i=1}^m \langle \mathbf{n}_i, \mathbf{b}\rangle \mathbf{n}_i$ und $\langle \mathbf{n}_i, \mathbf{b}\rangle = \ni^t \circ \mathbf{b}$.

Also sind Lösungen der Gleichung

$$A\mathbf{x} = \mathrm{pr}_{\mathrm{Bild}(A)}(\mathbf{b}) \Leftrightarrow Q \circ R\mathbf{x} = \mathrm{pr}_{\mathrm{Bild}(A)}(\mathbf{b})$$

genau die Lösungen von

$$Q^t \circ Q \circ R\mathbf{x} = Q^t\mathbf{b},$$

denn die Multiplikation mit Q^t berechnet die $\mathbf{n}$-Koordinaten der Elemente von $\mathrm{Bild}(A)$. Da die Spalten von Q orthonormal sind, gilt zudem $Q^t \circ Q = E_m$. Es gilt also:

Behauptung 24 (Beste Approximation für überbestimmte Gleichungssysteme)**.** *Ist A eine reelle $n \times m$-Matrix mit linear unabhängigen Spalten, $\mathbf{b} \in \mathbb{R}^n$ ein Vektor und $A = Q \circ R$ die QR-Zerlegung von A, so ist*

$$\mathbf{x} := R^{-1} \circ Q^t\mathbf{b}$$

der eindeutige Vektor $\mathbf{x} \in \mathbb{R}^m$ für den $\|A\mathbf{x} - \mathbf{b}\|$ minimal ist.

Die Annahme, dass die Spalten von A linear unabhängig sind, bedeutet insbesondere $n \geq m$, d.h A hat höchstens mehr Zeilen als Spalten, nicht weniger.

In den Anwendungen ist die Anzahl der Zeilen meist viel größer als die Anzahl der Spalten. Im Beispiel der Regressionsgerade war die Anzalh der Zeilen die Anzahl der Datenpunkte, die Anzahl der Spalten, die anzahl der Koeffizizenten der linearen Funktionen.

Das Beispiel der Berechnung einer Regressionsgeraden können wir als Beispiel für den obigen Satz auffassen. Hier war A durch die Auswertungsabbildung $\mathrm{ev}\colon \mathbb{R}[x]_{\leq 1} \cong \mathbb{R}^2 \to \mathbb{R}^{2n+1}$ gegeben. Unsere Basis $1, x$ war so gewählt, dass $\mathrm{ev}(1), \mathrm{ev}(x)$ orthogonal waren, darum konnten wir $Q^t \begin{pmatrix} f(-n) \\ \vdots \\ f(n) \end{pmatrix}$ einfach als $\begin{pmatrix} \langle \frac{\mathrm{ev}(1)}{\|\mathrm{ev}(1)\|}, f\rangle \\ \langle \frac{\mathrm{ev}(x)}{\|\mathrm{ev}(1)\|}, f\rangle \end{pmatrix}$ bestimmen.

Ausblick: Abstände und neuronale Netze

Es ist vielleicht, nur ein kleiner Punkt, aber die lineare Algebra, die wir bisher kennengelernt haben, findet sich in den modernen Algorithmen der künstlichen Intelligenz wieder. Auch wenn unsere Zeit knapp ist, möchte ich die Gelegenheit nutzen, das einmal anzureißen, obwohl ich das selbst nicht so gut verstehe wie mir das lieb wäre. Es gibt eine schöne kurze Einführung hierzu von

Geordie Williamson[5], in der Sie weitere Quellen finden, dieses Gebiet entwickelt sich gerade etwas unheimlich schnell.

[5] Geordie Williamson. Is deep learning a useful tool for the pure mathematician? Preprint, 2023. URL https://arxiv.org/abs/2304.12602

Das Grundprinzip haben Sie sicher gesehen. Ein einfaches Modell für die Funktionsweise von Neuronen ist, dass diese einerseits elektrische Signale zur Aktivierung von anderen Neuronen erhalten und geben selbst Signale weiter, wenn insgesamt eine hinreichende Schwelle der Aktivierung überschritten wird. Das Ergebnis eines Lern- oder Trainingsprozesses wird dabei durch die Sensitivität, d.h. welches Neuron auf welche Verbindungen wie stark reagiert festgehalten.

Liegen am Eingang eines Neurons x Impulse von $y_1, \ldots, \mathbf{y}_{m_x}$ an, so wird es die Spannungen gewichten und insgesamt $\sum_{j=1}^{m_x} w_j y_j$ erhalten und je nach Größe dieser Summe selbst das Signal $f(\sum_{i=1}^{m_x} w_i y_i)$ wobei f eine Aktivierungsfunktion ist weitergeben. Hierbei wird häufig einfach $f(x) = \max\{0, x\}$ verwendet, d.h. wir schneiden einfach negative Ergebnisse ab.

Ein einfaches Modell hierfür sortiert das Netz in Schichten:

$$\mathbb{R}^{n_1} \xrightarrow{A_1} \mathbb{R}^{n_2} \xrightarrow{Aktiv?} \mathbb{R}^{n_2} \xrightarrow{A_2} \mathbb{R}^{n_3} \xrightarrow{Aktiv?} \mathbb{R}^{n_3} \xrightarrow{A_3} \mathbb{R}^{n_4}$$

wobei die Koordinaten eines Vektors $\mathbf{y} \in \mathbb{R}^{n_i}$ einfach die bei den Neuronen in Schicht i ankommenden Signale angeben, A_i die Gewichtsmatrix ist, d.h. die Einträge von $A\mathbf{v}$ sind die gewichteten Summen $\sum_{j=1}^{m_x} w_j y_j$ und *Aktiv*? ist die Abbildung, die auf die Koordinaten jeweils die Aktivierungsfunktion f anwendet. Für die Funktion $f(x) = \max\{0, x\}$ liefert das eine stückweise lineare Abbildung.

Die Elemente des ersten Vektorraums beschreiben jeweils die Eingangsdaten (z.B. Helligkeiten von Bildpunkten, ein Tonsignal, etc.) die Elemente des letzten Vektorraums werden als Ergebnis interpretiert, z.B. die Wahrscheinlichkeit, dass das Bild den Buchstaben A,B,C,usw. darstellt.

Da die Daten oftmals nicht natürlich in einem $\mathbb{R}^n$ liegen und insbesondere der 0-Punkt nicht ganz natürlich bestimmt ist (sowohl eine Helligkeitsskala von 0 bis 1 (dunkel-hell) oder -1 bis 1 mit 0 als „so hell wie die Umgebung“ könnten natürlich sein) werden für die A_i oft affine Abbildungen, d.h. Abbildungen die aus linearen Abbildungen und Verschiebungen $\mathbf{v} \mapsto \mathbf{v} + \mathbf{v}_0$ entstehen, also $\mathbf{v} \mapsto A(\mathbf{v} + \mathbf{v}_0) + \mathbf{w}_0$ zugelassen.

Wenn wir ein einfaches Netz zum Beispiel darauf trainieren möchten, um auf einem Bild Buchstaben zu erkennen, könnten wir nun so vorgehen: Wir nehmen ein zufällig initialisiertes Netz und wenden es auf eine Auswahl von Testdaten $\mathbf{v}_1, \ldots, \mathbf{v}_N$ an. Für diese Berechnen wir das Endergebnis und schauen, die Abstände der Ergebnisvektoren vom „richtigen“ Ergebnis an. Daraus bestimmen wir eine Richtung in der wir die Gewichtsabbildungen ändern wollen, um den Abstand zu verkleinern. Das wiederholen wir.

Es ist erstaunlich, dass das für viele Probleme recht gut funktioniert.

Exkurs in die Geometrie: Orthogonale Abbildungen

Die ersten linearen Abbildungen, die Ihnen auf geometrischem Weg begegnet sind, waren wahrscheinlich Drehungen und Spieglungen. Diese Abbildungen erhalten Längen und Winkel, also auch das Skalarprodukt, d.h.

$$\langle A\mathbf{v}, A\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle.$$

Definition 25. Ist $V, \langle , \rangle$ ein euklidischer Vektorraum, dann heißt eine lineare Abbildung $A\colon V \to V$ *orthogonal* (oder Isometrie) genau dann, wenn für alle $\mathbf{v}, \mathbf{w} \in V$ gilt dass

$$\langle A\mathbf{v}, A\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle.$$

Die Menge der orthogonalen $n \times n$-Matrizen wird mit

$$O(n) := \{A \in \mathrm{GL}_n(\mathbb{R}) \mid A \text{ orthogonal}\}$$

bezeichnet.

Beispiel 26 (Orthogonale Abbildungen der Ebene sind Drehungen oder Spiegelungen)**.** Eine Isometrie $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ der Ebene $\mathbb{R}^2$, bildet den Standardbasisvektor $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ auf einen Vektor $A\mathbf{e}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$ der Länge 1 ab, d.h. es gilt $a^2 + b^2 = 1$.

Für den zu $\mathbf{e}_1$ orthogonalen Vektor $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ muss also für $A\mathbf{e}_2 = \begin{pmatrix} c \\ d \end{pmatrix}$ gelten, dass $\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \rangle = ac + bd = 0$ und $c^2 + d^2 = 1$. Die beiden zu $A\mathbf{e}_1 = \begin{pmatrix} a \\ b \end{pmatrix}$ orthonormalen Vektoren sind $\begin{pmatrix} -b \\ a \end{pmatrix}$ und $\begin{pmatrix} b \\ -a \end{pmatrix}$. Die Matrix

$$D = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

können wir als

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

schreiben, diese beschreibt also die Drehung um den Winkel θ. Für diese Matrix haben wir uns auch schon überlegt, dass das charakteristische Polynom $\det(D - tE_2) = t^2 - 2at + (a^2 + b^2) = t^2 - 2at + 1$ keine reellen Nullstellen hat.

Wählen wir den anderen zu $\begin{pmatrix} a \\ b \end{pmatrix}$ orthonormalen Vektor als zweite Spalte, also $A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$, so erhalten wir eine Spiegelung, denn hier ist $\det(A - tE_2) = t^2 - a^2 - b^2 = t^2 - 1 = (t+1)(t-1)$, die Matrix hat also Eigenwerte ± 1, mit orthogonalen Eigenvektoren $\mathbf{v}_1 = \begin{pmatrix} b \\ 1-a \end{pmatrix}$ zum Eigenwert 1 und $\mathbf{v}_2 = \begin{pmatrix} -b \\ 1+a \end{pmatrix}$ zum Eigenwert -1 falls $b \neq 0$, bzw. $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ falls $b = 0$. Die Abbildung beschreibt also die Spiegelung an der Geraden $\mathbb{R} \cdot \mathbf{v}_1 = \mathrm{Span}(\mathbf{v}_1)$.

Auf dem Übungsblatt überlegen Sie sich, dass eine Matrix A genau dann orthogonal ist, wenn die Spalten der Matrix eine Orthonormalbasis von $\mathbb{R}^n$ bilden.

Behauptung 27. 1. *Die orthogonalen Matrizen $O(n)$ besteht genau aus den Matrizen, für die $A^t = A^{-1}$ gilt, d.h. das sind genau die Matrizen, deren Spaltenvektoren eine Orthonormalbasis von $\mathbb{R}^n$ bilden.*

2. *Die Menge $O(n)$ ist eine Untergruppe der Gruppe der invertierbaren Matrizen GL_n, d.h. die Einheitsmatrix E_n ist orthogonal, zu jeder orthogonalen Matrix ist auch die inverse A^{-1} orthogonal und zu je zwei orthogonalen Matrizen $A, B \in O(n)$ ist auch das Produkt $A \circ B$ orthogonal.*

3. *Für jede orthogonale Matrix A gilt* $\det(A) = \pm 1$.

Beweis. Der Beweis von 1. und 2. ist eine Übungsaufgabe auf Blatt 5. Der 3. Teil folgt aus dem 1., denn

Fragen Sie nach, wenn Sie mit der Aufgabe Schwierigkeiten haben sollten!

$$\det(A^t) = \det(A) \text{ und } \det(A^{-1}) = \frac{1}{\det(A)}$$

darum folgt aus $A^t = A^{-1}$, dass $\det(A) = \frac{1}{\det(A)}$, d.h. $\det(A)^2 = 1$. Das bedeutet $\det(A) = \pm 1$. □

Bemerkung. 1. Wir hatten die Determinante einer Matrix durch die Suche nach einem Volumenbegriff im $\mathbb{R}^n$ motiviert. Dass für orthogonale Matrizen $|\det(A)| = 1$ gilt, bedeutet also, dass unsere Vorstellung, dass längen- und winkeltreue Abbildungen auch das Volumen erhalten.

2. Die Teilmenge der orthogonalen Matrizen mit Determinante 1 heißt spezielle Orthogonale Gruppe

$$SO(n) := \{A \in O(n) \mid \det(A) = 1\}.$$

Wegen des Determinanten–Multiplikationssatzes ist das wieder eine Untergruppe von $\mathrm{GL}_n(\mathbb{R})$.

3. Die Eigenwerte c einer orthogonalen Abbildung erfüllen $|c| = 1$, denn für orthogonale Abbildungen gilt nach Voraussetzung $\|A\mathbf{v}\| = \|\mathbf{v}\|$ für alle $\mathbf{v} \in V$, gilt also $A\mathbf{v} = c\mathbf{v}$ so folgt hieraus

$$\|\mathbf{v}\| = \|A\mathbf{v}\| = \|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

und damit $|c| = 1$.

Wir können hiermit nun die orthogonalen Abbildungen in $\mathbb{R}^2$ und $\mathbb{R}^3$ gut verstehen:

Satz 28. *1. Jede orthogonale Matrix $A \in SO(2)$ ist eine Drehung.*

2. Jede orthogonale Matrix $A \in O(2)$ mit $\det(A) = -1$ ist eine Spiegelung.

3. Jede orthogonale Matrix $A \in SO(3)$ ist eine Drehung, d.h. existiert ein Eigenvektor $\mathbf{v}_1$ zum Eigenwert 1 und eine Orthonormalbasis $\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \mathbf{n}_2, \mathbf{n}_3$ so dass

$$_{\underline{\mathbf{n}}}\mathrm{Mat}_{\underline{\mathbf{n}}}(A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & -b \\ 0 & b & a \end{pmatrix}$$

für Zahlen a, b mit $a^2 + b^2 = 1$.

Das Resultat zu $SO(3)$ ist überraschend, es bedeutet, dass die Hinternanderausführung von 2 Drehungen um verschiedene Drehachsen, immer selbst eine Drehung ist – allerdings um eine Achse, die meist mit den ersten beiden Drehachsen wenig zu tun hat.

Probieren Sie das wenigstens einmal mit einem Würfel aus!

In höherer Dimension, z.B. in $\mathbb{R}^4$ ist die entsprechende Aussage nicht mehr richtig.

Beweis. Die Aussagen zu 1. und 2. hatten wir schon gezeigt, als wir die orthogonalen Abbildungen der Ebene beschrieben haben.

Zu 3. ist der wichtigste Trick, zu bemerken, dass das charakteristische Polynom $\det(A - tE_3) \in \mathbb{R}[t]$ der Abbildung $A \in SO(3)$ ein Polynom vom Grad 3 ist. Alle Polynome ungeraden Grades haben wegen des Zwischenwertsatzes eine Nullstelle in $\mathbb{R}$. Also hat A einen reellen Eigenwert c. Da Eigenwerte orthogonaler Abbildungen $|c| = 1$ erfüllen, ist also $c \in \{1, -1\}$.

Es existiert also ein $\mathbf{v}_1 \in \mathbb{R}^3$ mit $\|\mathbf{v}_1\| = 1$ und $A\mathbf{v}_1 = c\mathbf{v}_1$. Da A orthogonal ist, bildet A den Unterraum $(\mathbb{R}\mathbf{v}_1)^\perp = \{\mathbf{w} \in \mathbb{R}^3 \mid \langle \mathbf{w}, \mathbf{v}_1 \rangle = 0\}$ auf sich selbst ab und auf dem Übungsblatt haben Sie gezeigt, dass

$$\mathbb{R}^3 = (\mathbb{R}\mathbf{v}_1) \oplus (\mathbb{R}\mathbf{v}_1)^\perp.$$

Die Einschränkung der Abbildung A auf den Unterraum $(\mathbb{R}\mathbf{v}_1)^\perp$ ist weiterhin orthogonal. Ergänzen wir also $\mathbf{v}_1$ durch eine Orthonormalbasis $\mathbf{v}_2, \mathbf{v}_3$ von $(\mathbb{R}\mathbf{v}_1)^\perp$, so ist

$$_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) = \begin{pmatrix} c & 0 & 0 \\ 0 & a' & b' \\ 0 & c' & d' \end{pmatrix},$$

wobei $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in O(2)$ orthogonal ist.

Außerdem gilt

$$1 = \det(A) = c \cdot \det(\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}).$$

Im Fall $c = 1$ muss darum $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SO(2)$ gelten, d.h. A ist nach 1. eine Drehung um die Achse $\mathbb{R}\mathbf{v}_1$.

Im Fall $c = -1$ ist $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$ nach 2 eine Spiegelung, d.h. es gibt eine ON-Basis $(\mathbf{v}_2', \mathbf{v}_3')$ von $\mathrm{Span}(\mathbf{v}_2, \mathbf{v}_3)$ so dass für die Basis

$\underline{\mathbf{v}}' = (\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3')$ gilt, dass

$$_{\underline{\mathbf{v}}'}\mathrm{Mat}_{\underline{\mathbf{v}}'}(A) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

d.h. A ist eine Drehung um 180° um die Achse $\mathbf{v}_3'$. □

Multiplikation in $\mathbb{C}$ und Drehungen

Die Beschreibung von Drehungen durch Matrizen der Form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ erklärt, was die Multiplikation in $\mathbb{C}$ mit Drehungen zu tun hat. Sie Erinnern sich, dass die komplexen Zahlen

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

die formalen Ausdrücke der Form $a + ib$ sind, bei denen die Rechenregel $i^2 = -1$ für das Symbol i gilt. Diese Menge können wir auch als Ebene $\mathbb{R}^2$, also als $\mathbb{R}$-Vektorraum auffassen, indem wir die Basis $1, i$ wählen. Damit hatten Sie sich im letzten Semester überlegt, dass die Multiplikation mit einer komplexen Zahl $z = a + ib$ bezüglich der Basis $\underline{\mathbf{v}} = (\mathbf{v}_1 = 1, \mathbf{v}_2 = i)$ gerade

$$_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}((a + ib)\cdot) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

ist, da $(a + ib) \cdot 1 = a + ib = a\mathbf{v}_1 + b\mathbf{v}_2$ und $(a + ib) \cdot i = -b + ia = -b\mathbf{v}_1 + a\mathbf{v}_2$. Falls $|z| = \sqrt{a^2 + b^2} = 1$ gilt, ist das also in der Tat die Drehung um den Winkel zwischen 1 und z.

Die Multiplikation mit einer komplexen Zahl vom Betrag 1 ist also eine Drehung.

Den Betrag der Zahl $z = a + ib$ können wir auch als $|z|^2 = z \cdot \overline{z}$ schreiben, wobei $\overline{z} = \overline{a + ib} := a - ib$ die komplexe Konjugation bezeichnet. Die komplexe Konjugation $i \mapsto -i$ ist ein Isomorphismus von $\mathbb{C}$ der mit den Rechenoperationen verträglich ist

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

weil $(-i)^2 = -1$ die gleiche Rechenregel wie i erfüllt.

Geometrisch entspricht die komplexe Konjugation genau der Spiegelung an der x-Achse in $\mathbb{R}^2$.

Komplexe Zahlen der Länge 1 und die Exponentialfunktion

Die komplexen Zahlen der Länge 1 entsprechen nach Definition genau den Punkten auf dem Einheitskreis. Diese lassen sich besonders schön mittels der Exponentialfunkion beschreiben. Das ist wieder ein Beispiel einer mathematischen Zweckentfremdung, d.h.

die Verwendung eines Objekts für etwas für das es ursprünglich gar nicht vorgesehen war.

Da wir nicht in der Analysisvorlesung sind, möchte ich das Grundprinzip einmal algebraisch vorstellen. Die Ableitung ist als lineare Abbildung auf dem Vektorraum der Polynome nilpotent, da die Ableitung den Grad eines Polynoms verkleinert. Es gibt dort also nur den Eigenwert 0. Das ist anders wenn wir stattdessen Potenzreihen anschauen. Für einen Körper K bezeichnen wir mit

$$K[\![t]\!] := \{a_0 + a_1 t + a_2 t^2 + \cdots = \sum_{n=0}^{\infty} a_n t^n \mid a_i \in K\}$$

den K-Vektorraum der formalen Potenzreihen mit Koeffizienten in K. Hier ist es jetzt wichtig, dass wir einfach mit formalen Ausdrücken rechnen, denn wir können in Körpern selten Elemente von K für t einsetzen, da unendliche Summen nicht immer sinnvoll sind. Formal können wir Potenzreichen $p(t), q(t)$ aber wie Polynome addieren und multiplizieren, es gelten die gleichen Formeln für die Koeffizienten wie bei Polynomen. Insbesondere gelten auch die gleichen Rechenregeln (Assoziativ-, Kommutativ-, Distributivgesetz).

Genauso können wir die formale Ableitung

$$\begin{aligned} (\quad)' \colon K[\![t]\!] &\to K[\![t]\!] \\ p(t) = \sum_{n=0}^{\infty} a_n t^n &\mapsto p(t)' := \sum_{n=1}^{\infty} n a_n t^{n-1} \\ &\overset{m=n-1}{=} \sum_{m=0}^{\infty} (m+1) a_{m+1} t^m \end{aligned}$$

definieren. Diese Abbildung ist sichtbar linear und erfüllt die gleichen Rechenregeln wie für Polynome (z.B. die Produktregel) wiederum, weil die Formeln für die Koeffizienten mit den Formeln für Polynome übereinstimmen.

Lemma 29 (Exponentialreihe)**.** *Sei K ein Körper, der die rationalen Zahlen enthält. Dann ist für alle $c \in K$ ist der Eigenraum zum Eigenwert c der Ableitung auf $K[\![t]\!]$ eindimensional:*

$$\dim_K \operatorname{Eig}((\quad)', c) = 1.$$

Der eindeutig bestimmte Eigenvektor $p(t)$ zum Eigenwert c mit konstantem Koeffizient $p(0) = 1$ ist

$$\begin{aligned} p(t) &= \sum_{n=0}^{\infty} \frac{c^n}{n!} t^n \\ &= 1 + ct + \frac{c^2}{2} t^2 + \frac{c^3}{6} t^3 + \frac{c^4}{24} t^4 + \dots . \end{aligned}$$

Wenn Sie den Beweis genau anschauen, sehen Sie, dass wir nur durch Elemente in $\mathbb{Q}$ teilen. Das Ergebnis gilt darum etwas allgemeiner, K muss selbst kein Körper sein.

Beweis. Ein Element $p(t) = \sum_{n=0}^{\infty} a_n t^n \in K[\![t]\!]$ ist nach Definition genau dann ein Eigenvektor zum Eigenwert c wenn

$$p(t)' = c \cdot p(t).$$

Vorbemerkung: Da der Vektorraum $K[\![t]\!]$ nicht endlich dimensional ist, können wir keine Matrix für die Abbildung angeben und damit auch die Eigenvektoren nicht wir üblich ausrechnen. Wenn uns nichts mehr einfällt, hilft nur noch die Definition.

Wir wissen

$$\begin{aligned} p(t)' &= \sum_{n=0}^{\infty} (n+1)a_{n+1}t^n \\ c \cdot p(t) &= \sum_{n=0}^{\infty} c \cdot a_n t^n. \end{aligned}$$

Gleichheit dieser beiden Potenzreihen bedeutet, dass für alle n die Koeffizienten von t^n übereinstimmen, d.h.

$$\begin{aligned} a_1 &= c \cdot a_0 && n = 0 \\ 2 \cdot a_2 &= c \cdot a_1 && n = 1 \\ 3 \cdot a_3 &= c \cdot a_0 && n = 2 \\ (n+1)a_{n+1} &= c \cdot a_n \end{aligned}$$

Damit können wir induktiv, die Koeffizienten a_n aus a_0 berechnen: Setzen wir $a_1 = ca_0$ in die 2. Gleichung ein folgt, dass $2 \cdot a_2 = ca_1 = c^2 a_0$ also

$$a_2 = \frac{1}{2} c^2 a_0.$$

Setzen wir das in die 3. Gleichung ein finden wir $3a_3 = ca_2 = \frac{c^3}{2} a_0$ also

$$a_3 = \frac{1}{2 \cdot 3} c^3 a_0.$$

Wir zeigen darum mit Induktion, dass für alle n gilt, dass

$$a_n = \frac{c^n}{n!} a_0.$$

Den Induktionsanfang haben wir gerade gesehen und gilt die Aussage für eine Zahl m, d.h. $a_m = \frac{c^m}{m!} a_0$ so folgt aus $(m+1)a_{m+1} = c \cdot a_m = c \cdot \frac{c^m}{m!} a_0$ auch

$$a_{m+1} = \frac{1}{m+1} \cdot \frac{c^{m+1}}{m!} a_0 = \frac{c^{m+1}}{(m+1)!} a_0.$$

Also gilt die Aussage dann auch für $m+1$ und somit für alle n. □

Definition. Die Reihe $e(t) = \sum_{n=0}^{\infty} \frac{1}{n!} t^n \in K[\![t]\!]$ heißt Exponentialreihe. Wir schreiben $e^t := e(t)$.

Bemerkung. 1. Aus dem Lemma zur Exponentialreihe folgt, dass

$$e^t \cdot e^{-t} = 1.$$

denn beide Seiten sind Eigenvektoren zum Eigenwert 0, da

$$\begin{aligned} (e^t \cdot e^{-t})' &= (e^t)' e^{-t} + e^t \cdot (e^{-t})' && \text{Produktregel} \\ &= e^t \cdot e^{-t} + e^t \cdot (-e^t) = 0 \end{aligned}$$

und haben konstanten Koeffizienten 1, müssen also gleich sein.

2. Mit einem ähnlichen Argument im Potenzreihenring in 2 Variablen $K[\![t_1]\!][\![t_2]\!] = KK[\![t_1, t_2]\!]$ können wir uns auch überlegen, dass

$$e^{t_1} \cdot e^{t_2} = e^{t_1 + t_2}$$

gilt.

In der Analysis werden Sie sich überlegen, dass die Exponentialreihe für alle $t \in \mathbb{C}$ sehr gut konvergiert (absolut und gleichmäßig auf beschränkten Teilmengen von $\mathbb{C}$) und wir damit für alle $z \in \mathbb{C}$ eine Zahl $e^z \in \mathbb{C}$ erhalten.

Das Argument für die Konvergenz ist nicht sehr schwer, der Punkt ist, dass sich die Potenzen t^n, t^{n+1} jeweils um den Faktor t unterscheiden, aber $n!, (n+1)!$ um den Faktor $n+1$, der Nenner wächst darum für $n > |t|$ viel schneller als t^n.

Aus der ersten Formel $e^z \cdot e^{-z} = 1$ folgt dann sofort, dass e^{-z} ein Inverses von e^z ist, insbesondere also $e^z \neq 0$ für alle $z \in \mathbb{C}$, d.h. e^z hat keine Nullstellen. Die zweite Formel erklärt auch die Schreibweise, denn für $e = e^1$ gilt in der Tat $e^n = e^{1+1+\cdots+1} = \underbrace{e \cdot e \cdot \cdots \cdot e}_{n-\text{mal}}$.

Ist $z = ix$ für eine reelle Zahl $x \in \mathbb{R}$, so gilt außerdem, dass

$$\|e^{ix}\|^2 = e^{ix} \cdot \overline{e^{ix}} = e^{ix} \cdot e^{-ix} = 1.$$

Die Zahlen der Form e^{ix} haben also alle Betrag 1, liegen also auf dem Einheitskreis. Außerdem gilt für die Ableitung der Funktion

$$e^{ix} \colon \mathbb{R} \to \mathbb{C}, x \mapsto e^{ix}$$

dass $|(e^{ix})'| = |ie^{ix}| = |i||e^{ix}| = 1$ und Sie werden in der Analysis lernen, dass Ableitung vom Betrag 1 bedeutet, dass das Kurvenstück $\{e^{ix} \mid x \in [0,t]\}$ die Länge t hat. Damit muss also wenn 2π die Länge des Einheitskreises ist gelten, dass

$$e^{2\pi i} = 1$$

und für die halbe Länge π muss gelten

$$e^{i\pi} = -1.$$

Das ist eine der Formeln für die Top-10 der schönsten mathematischen Formeln, da alle auftretenden Symbole, die negative Zahl -1, die Länge des Halbkreises π, die Eulersche Zahl e und i, die Wurzel aus -1 alle einzeln schwer zu finden waren und aus ganz unterschiedlichen Gründen eingeführt wurden. Dass die 4 in einer Formel zusammenpassen, ist erstaunlich.

Bemerkung (N-te Einheitswurzeln)**.** Für jede natürliche Zahl $N \in \mathbb{N}$ erfüllen

$$\zeta_N = e^{\frac{2\pi i}{N}} \text{ und } (\zeta_N)^k = e^{\frac{2\pi i}{N}k} \text{ für } k = 0, \dots N-1$$

die Gleichung $z^N = 1$ und heißen darum *N-te Einheitswurzeln*. Unsere geometrische Interpretation, dass $e^{i\alpha}$ die komplexe Zahl mit Winkel α zur reellen Achse auf dem Einheitskreis ist besagt für die N-ten Einheitswurzeln also, dass diese den Einheitskreis in N gleiche Stücke teilen.

Zum Abschluss sollten wir noch die Koordinaten $e^{ix} = a + ib$

ausrechnen:

$$\begin{aligned} e^{ix} &= \sum_{n=0}^{\infty} \frac{i^n}{n!} x^n \qquad\qquad i^2 = -1 \text{ sortiere daher nach } n \text{gerade/ungerade} \\ &= \sum_{\substack{n=2m\\ \text{gerade}}} \frac{i^{2n}}{(2n)!} x^{2n} + \sum_{\substack{n=2m+1\\ \text{ungerade}}} \frac{i^{2n+1}}{(2n+1)!} x^{2n+1} \\ &= \underbrace{\sum_{\substack{n=2m\\ \text{gerade}}} \frac{(-1)^n}{(2n)!} x^{2n}}_{=\cos(x)} + i \cdot \underbrace{\sum_{\substack{n=2m+1\\ \text{ungerade}}} \frac{(-1)^n}{(2n+1)!} x^{2n+1}}_{=\sin(x)} \\ &= \cos(x) + i\sin(x) \end{aligned}$$

Die Multiplikation mit e^{ix} ist also tatsächlich die Drehung um den Winkel x!

Diese Formel ist im übrigen die einzige Art, wie ich mir die Additionstheoreme für $\sin(x)$ und $\cos(x)$ merken kann: Lassen Sie uns die Formel

$$e^{i(x+y)} = e^{ix} \cdot e^{iy}$$

einmal in Koordinaten ausschreiben:

$$\begin{aligned} e^{i(x+y)} &= e^{ix} \cdot e^{iy} \\ &= (\cos(x) + i\sin(x)) \cdot (\cos(y) + i\sin(y)) \\ &= (\cos(x)\cos(y) - \sin(x)\sin(y)) + i(\cos(x)\sin(y) + \sin(x)\cos(y)). \end{aligned}$$

Da $e^{i(x+y)} = \cos(x+y) + i\sin(x+y)$ folgt also:

$$\begin{aligned} \cos(x+y) &= \cos(x)\cos(y) - \sin(x)\sin(y) \\ \sin(x+y) &= \cos(x)\sin(y) + \sin(x)\cos(y). \end{aligned}$$

Das ist also nur eine komplizierte Art, die Formel

$$e^{i(x+y)} = e^{ix} \cdot e^{iy}$$

aufzuschreiben. Vielleicht stimmen Sie mir zu, dass das die einfachere Formel ist.

Damit genug zur komplexen Exponentialfunktion, ich hoffe Sie haben dieses nützliche Wunder damit ein wenig schätzen gelernt.

Skalarprodukte für komplexe Vektorräume

Motivation: Der Spektralsatz

Um Endomorphismen $A\colon V \to V$ zu verstehen, haben wir vorzugsweise Basen aus Eigenvektoren gesucht. In euklidischen Vektorräumen haben wir nun gesehen, dass wir besonders leicht mit Orthonormalbasen rechnen können. Damit stellt sich die Frage, für welche lineare Abbildungen wir beides auf einmal erreichen können.

FRAGE: Welche linearen Abbildungen $A\colon V \to V$ eines euklidischen Vektorraums besitzen eine Orthonormalbasis aus Eigenvektoren?

Die Antwort hierauf ist besonders schön und es stellt sich heraus, dass diese Matrizen tatsächlich in der Natur oft vorkommen.

Um uns das zu überlegen, wird es – wie immer bei Eigenwerten – nützlich sein, in den komplexen Zahlen statt den reellen zu rechnen, weil charakteristische Polynome dort immer Nullstellen haben und es darum immer Eigenwerte gibt.

Dafür müssten wir das Skalarprodukt

$$\langle , \rangle \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

auf $\mathbb{C}^n$ ausdehnen. Dafür könnten wir zunächst versuchen, die Eigenschaften „symmetrisch, bilinear und positiv definit" für

$$\langle , \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$

zu übernehmen. Das hat aber das Problem, dass in $\mathbb{C}$ die Eigenschaften „bilinear" und „positiv definit" unverträglich sind, weil dann

$$\langle i\mathbf{v}, i\mathbf{v} \rangle = i^2 \langle \mathbf{v}, \mathbf{v} \rangle = -\langle v, v \rangle$$

gelten würde und darum $\langle i\mathbf{v}, i\mathbf{v} \rangle$ und $\langle v, v \rangle$ nicht beide positiv sein können.

An dieser Stelle fällt uns vielleicht auf, dass das für $n = 1$ schon Blödsinn ist, denn dort haben wir uns schon überlegt dass die Länge $|z|$ von $z = a + ib \in \mathbb{C}$ durch

$$|z|^2 = z \cdot \overline{z}$$

bestimmt ist. Darum verallgemeinern wir die Formel für das Standardskalarprodukt einfach indem wir die Koordinaten des zweiten Vektors konjugieren:

Definition. Das (hermitesche) Standardskalarpdodukt auf $\mathbb{C}^n$ ist die Abbildung

$$\langle \quad , \quad \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$

die durch die Formel

$$\left\langle \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle := z_1 \overline{w}_1 + \cdots + z_n \overline{w}_n$$

gegeben ist.

Wir könnten die Formel auch als

$$\left\langle \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle = (z_1, \cdots, z_n) \circ \begin{pmatrix} \overline{w}_1 \\ \vdots \\ \overline{w}_n \end{pmatrix}$$

also als Multiplikation einer $1 \times n$ mit einer $n \times 1$-Matrix schreiben.

Zwei Vektoren $\mathbf{v}, \mathbf{w}$ heißen *senkrecht* (oder *orthogonal*) bezüglich $\langle \quad , \quad \rangle$, wenn

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0.$$

Wir schreiben $\mathbf{v} \perp \mathbf{w}$ für „$\mathbf{v}$ *ist senkrecht zu* $\mathbf{w}$".

Die *Länge* (oder *Norm*) des Vektors $\mathbf{v} \in \mathbb{C}^n$ ist

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

ZUM BEISPIEL IST $\langle \begin{pmatrix} 1 \\ i \end{pmatrix}, \begin{pmatrix} 1 \\ -i \end{pmatrix} \rangle = 1 \cdot 1 + i \cdot (\overline{-i}) = 1 + i^2 = 0$ und also sind die beiden Vektoren orthogonal, was wir gerade eben noch mit unserer geometrischen Vorstellung vergleichen können, weil die Vektoren in einem 2-dimensionalen reellen Unterraum von $\mathbb{C}^2 \cong \mathbb{R}^4$ liegen.

Bemerkung. An der Definition des Standardskalarprodukts können wir die folgende Eigenschaften ablesen:

1. (hermitesch) Es gilt $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$ für alle $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$. Insbesondere ist für alle $\mathbf{v} \in \mathbb{C}^n$ $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$

Das ist der Ersatz für „symmetrisch"

2. (sesquilinear) Die Abbildung $\langle \ , \ \rangle$ ist in der ersten Variablen linear:

Das ist der Ersatz für „bilinear"

$$\begin{aligned} \langle \mathbf{v} + \mathbf{v}', \mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}', \mathbf{w} \rangle, \\ \langle a\mathbf{v}, \mathbf{w} \rangle &= a \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

   und in der zweiten Variablen konjugiert-linear:

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} + \mathbf{w}' \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w}' \rangle, \\ \langle \mathbf{v}, a\mathbf{w} \rangle &= \overline{a} \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

   für alle $a \in \mathbb{C}, \mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in \mathbb{C}^n$.

3. (positiv definit) Für alle $\mathbf{v} \in \mathbb{C}^n$ gilt $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ und

$$\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = 0.$$

Wie bei reellen Skalarprodukten, ist es wieder nützlich, diesen Begriff nicht nur für $\mathbb{C}^n$ sondern auch für Unterräume $U \subseteq \mathbb{C}^n$, Funktionenräume und andere Vektorräume verfügbar zu haben.

Definition 30. Ein *Skalarprodukt* auf einem $\mathbb{C}$-Vektorraum V ist eine Abbildung

$$\begin{aligned} \langle \ , \ \rangle \colon V \times V &\to \mathbb{C} \\ (\mathbf{v}, \mathbf{w}) &\mapsto \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

für die die folgenden Rechenregeln gelten:

1. (hermitesch) Es gilt $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$ für alle $\mathbf{v}, \mathbf{w} \in V$. Insbesondere ist für alle $\mathbf{v} \in \mathbb{C}^n$ $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$

Das ist der Ersatz für „symmetrisch"

2. (sesquilinear) Die Abbildung $\langle \ , \ \rangle$ ist in der ersten Variablen linear:

Das ist der Ersatz für „bilinear"

$$\begin{aligned} \langle \mathbf{v} + \mathbf{v}', \mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}', \mathbf{w} \rangle, \\ \langle a\mathbf{v}, \mathbf{w} \rangle &= a \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

   und in der zweiten Variablen konjugiert-linear:

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} + \mathbf{w}' \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w}' \rangle, \\ \langle \mathbf{v}, a\mathbf{w} \rangle &= \overline{a} \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

   für alle $a \in \mathbb{C}, \mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in V$.

3. (positiv definit) Für alle $\mathbf{v} \in V$ gilt $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ und

$$\langle \mathbf{v}, \mathbf{v} \rangle = 0 \Leftrightarrow \mathbf{v} = 0.$$

Ein *unitärer Vektorraum* ist ein $\mathbb{C}$-Vektorraum V, zusammen mit einem Skalarprodukt $\langle \quad , \quad \rangle$.

Beispiel 31. 1. Unser Standardskalarprodukt $\langle \quad , \quad \rangle \colon \mathbb{C}^n \to \mathbb{C}^n$.

2. Das Integral auf dem Raum der reellen Polynome $\mathbb{R}[t]$ können wir auf komplexe Polynome $\mathbb{C}[t]$ erweitern:

$$\langle p, q \rangle := \int_0^1 p(x) \cdot \overline{q(x)} dx$$

Hierbei lesen wir die $\mathbb{C}$-wertige Funktion $f(x) = p(x) \cdot \overline{q(x)} = a(x) + ib(x)$ als Summe des reellen und imaginären Anteils und definieren

$$\int_0^1 a(x) + ib(x) dx := \int_0^1 a(x) dx + i \int_0^1 b(x) dx,$$

d.h. wir integrieren Real- und Imaginärteil der Funktion separat.

Für Polynome stellt sich heraus, dass damit tatsächlich wieder die gleichen Rechenregeln wie zuvor gelten, was in dieser Form etwas überraschend ist. Sie sollten das einmal ausprobieren.

In jedem Fall ist das Integral sicher linear, da die Abbildung $p(x), q(x) \mapsto p(x) \cdot q(x)$ sichtbar sesquilinear und hermitesch ist, ist auch $\langle , \rangle$ sesquilinear und hermitesch.

Im Integral haben wir außerdem, wie beim Standardskalarpodukt die zweite Funktion $q(x)$ in der Formel komplex konjugiert, damit

$$\langle p, p \rangle \geq 0$$

sicher gestellt ist.

3. Wie das bei Verallgemeinerungen so ist, erlaubt es die Definition jetzt (leider ?) auch, auf $\mathbb{C}^n$ viele verschiedene Skalarprodukte aufschreiben. Ist $A \in \mathrm{Mat}_{n,n}(\mathbb{C})$ eine $n \times n$-Matrix so ist die Abbildung

$$\begin{aligned} \langle , \rangle_A \colon \mathbb{C}^n \times \mathbb{C}^n &\to \mathbb{C} \\ (\mathbf{v}, \mathbf{w}) &\mapsto \langle \mathbf{v}, \mathbf{w} \rangle_A := \mathbf{v}^t \circ A \circ \overline{\mathbf{w}} \end{aligned}$$

immer sesquilinear, weil die Matrizenmultiplikation das Distributivgesetz erfüllt. Sie können einmal versuchen, herauszufinden, welche Eigenschaft die Matrix besitzen muss, um hermitesch zu sein und dann versuchen Matrizen zu finden für die die Abbildung auch positiv definit ist.

In der Vorlesung haben wir die notwendige Bedingung $A^t = \overline{A}$ gefunden, und außerdem gesehen, dass A zumindest injektiv sein muss, wenn $\langle , \rangle_A$ positiv definit sein soll.

Auf dem Übungsblatt zeigen Sie auch, dass wir auf diese Art alle Skalarprodukte auf $\mathbb{C}^n$ beschreiben können.

Erinnerung: Es gilt immer

$$(AB)^t = B^t A^t,$$

denn der (i,j)-te Eintrag von $A \circ B$ ist (i-te Zeile A) $\circ$ (j-te Spalte B) und das ist (j-te Zeile B^t) $\circ$ (i-te Spalte A^t).

Bemerkung. 1. Zu jedem Skalarprodukt $\langle \quad , \quad \rangle$ definieren wir

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

und nennen das die Norm (oder Länge) bezüglich des Skalarproduktes.

2. In jedem unitären Vektorraum $V, \langle \quad , \quad \rangle$ gilt die Cauchy-Schwarz Ungleichung:

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$$

für alle $\mathbf{v}, \mathbf{w} \in V$ und darum auch die Dreiecksungleichung

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Sie sollten einmal nachprüfen, dass wir unseren Beweis tatsächlich für $\mathbb{C}$-Vektorräume umschreiben können.

3. Auch in unitären Vektorräumen $V, \langle , \rangle$ sind orthogonale Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V \setminus \{0\}$ linear unabhängig, denn wie zuvor gilt

$$\langle a_1 \mathbf{v}_1 + \cdots + a_r \mathbf{v}_r, \mathbf{v}_i \rangle = \sum_{j=1}^{r} a_j \langle \mathbf{v}_j, \mathbf{v}_i \rangle = a_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle.$$

Also hat die Gleichung $a_1 \mathbf{v}_1 + \cdots + a_r \mathbf{v}_r = 0$ nur die triviale Lösung $a_1 = \cdots = a_r = 0$.

4. Auf dem Übungsblatt werden Sie ganz ähnlich überlegen, dass das Orthogonalisierungsverfahren auch für Vektoren in unitären Vektorräumen funktioniert, Sie dabei aber sehr aufpassen müssen, wie Sie die Formel aufschreiben, weil es in jedem Schritt einen Unterschied macht, ob Sie $\langle \mathbf{v}, \mathbf{w} \rangle$ oder $\langle \mathbf{w}, \mathbf{v} \rangle$ schreiben. Ich kann mir das nur merken, indem ich mich an das Argument statt nur an die Formel erinnere.

Lassen Sie uns zur Übung noch eine Aussage zu Orthogonalbasen und orthogonalen Matrizen für $\mathbb{C}$-Vektorräume übertragen:

Lemma 32. *Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{C}^n$ sind genau dann eine Orthonormalbasis von $\mathbb{C}^n$, wenn für die Martix $A = (\mathbf{v}_1 \ldots \mathbf{v}_n)$ gilt dass $\overline{A}^t \circ \overline{A} = E_n$, d.h. wenn*

$$A^{-1} = \overline{A}^t.$$

Matrizen dieser Form heißen *unitär*.

Beispiel: Periodische Funktionen und .mp3

Um aus dem Vektorraum der periodischen Funktionen

$$\mathrm{Fun}_{per}(\mathbb{R}, \mathbb{C}) := \{f \colon \mathbb{R} \to \mathbb{C} \mid f(x+1) = f(x) \text{ für alle } x \in \mathbb{R}\}$$

einen endlich-dimensionalen Vektorraum zu machen, betrachten wir nur die N Funktionswerte bei $\frac{0}{N}, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}$ (und erinnern uns, dass für eine periodische Funktion $f(1) = f(\frac{N}{N})) = f(\frac{0}{N})$.

Um mir die Brüche zu ersparen, fassen wir das als Funktion

$$F\colon \mathbb{Z}/N\mathbb{Z} = \{[0], \dots, [N-1]\} \to \mathbb{C}$$

auf, d.h. wir setzen $F([k]) := f(\frac{k}{N})$ und schreiben

$$\mathrm{Fun}_N = \{f\colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}\}.$$

Diesen Vektorraum statten wir mit dem Skalarprodukt aus, das vom Integral auf periodischen Funktionen motiviert ist:

$$\langle F, G\rangle := \frac{1}{N} \cdot \sum_{k=0}^{N-1} F(k) \cdot \overline{G(k)}.$$

Das ist einfach ein Vielfaches des Standardskalarproduktes und hoffentlich sichtbar hermitesch, sesquilinear und positiv definit.

Auf diesem Vektorraum ist die Verschiebung $F(X) \mapsto F(X+1)$ eine lineare Abbildung, die das Skalarprodukt erhält:

$$\begin{aligned} T\colon \mathrm{Fun}_N &\to \mathrm{Fun}_N \\ F(k) &\mapsto (TF)(k) := F(k+1) \end{aligned}$$

erfüllt

$$\begin{aligned} \langle TF, TG\rangle &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} (TF)(k) \cdot \overline{(TG)(k)} \\ &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} F(k+1) \cdot \overline{G(k+1)} \\ &= \frac{1}{N} \cdot \sum_{\ell=0}^{N-1} F(\ell) \cdot \overline{G(\ell)} = \langle F, G\rangle \end{aligned}$$

für alle $F, G \in \mathrm{Fun}_N$.

Mit der komplexen Exponentialfunktion sehen Sie vielleicht unmittelbar ein oder zwei Eigenvektoren für T:

$$e_1(k) := e^{\frac{2\pi i}{N}\cdot k} \qquad \text{für } k = 0, \dots, N-1$$

erfüllt sicher

$$e_1(k+1) = e^{\frac{2\pi i}{N}\cdot(k+1)} = e^{\frac{2\pi i}{N}\cdot k} e^{\frac{2\pi i}{N}\cdot 1} = e^{\frac{2\pi i}{N}} \cdot e_1(k).$$

Genauso ist für $m = 0, \dots, N-1$ die Abbildung

$$e_m(k) := e^{\frac{2\pi i}{N}\cdot m\cdot k} \qquad \text{für } k = 0, \dots, N-1$$

ein Eigenvektor zum Eigenwert $e^{\frac{2\pi i\cdot m}{N}}$.

Behauptung 33. *Die Vektoren $e_0, \dots, e_{N-1}$, definiert als*

$$e_m(k) := e^{\frac{2\pi i}{N}\cdot m\cdot k} \qquad \textit{für } k = 0, \dots, N-1$$

sind eine Orthonormalbasis des Vektorraums der periodischen Funktionen

$$\mathrm{Fun}_N = \{f\colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}\}.$$

Beweis. Zunächst haben die Vektoren alle Länge 1, denn

$$\begin{aligned}\langle e_m, e_m \rangle &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} e^{\frac{2\pi i}{N} \cdot m \cdot k} \cdot \overline{e^{\frac{2\pi i}{N} \cdot m \cdot k}} \\ &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} e^{\frac{2\pi i}{N} \cdot m \cdot k} \cdot e^{-\frac{2\pi i}{N} \cdot m \cdot k} \\ &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} 1 = \frac{1}{N} \cdot N = 1.\end{aligned}$$

Für die Orthogonalität können sie auf dem Übungsblatt einmal versuchen, die Formel auszurechnen. Um das hier zu vermeiden, erinnern wir uns lieber, dass die Funktionen Eigenvektoren zu unterschiedlichen Eigenwerten sind und wir darum wissen, dass für $m \neq \ell \in \mathbb{Z}/N\mathbb{Z}$ gilt

$$\begin{aligned}\langle Te_m, Te_\ell \rangle &= \langle e^{\frac{2\pi i \cdot m}{N}} e_m, e^{\frac{2\pi i \cdot \ell}{N}} e_\ell \rangle \\ &= e^{\frac{2\pi i \cdot m}{N}} \langle e_m, e^{\frac{2\pi i \cdot \ell}{N}} e_\ell \rangle \\ &= e^{\frac{2\pi i \cdot m}{N}} e^{-\frac{2\pi i \cdot \ell}{N}} \langle e_m, e_\ell \rangle \\ &= e^{\frac{2\pi i \cdot m - \ell}{N}} \langle e_m, e_\ell \rangle.\end{aligned}$$

Da $\langle , \rangle$ positiv definit ist, muss entweder $\langle e_m, e_\ell \rangle = 0$ gelten, oder die Zahl $e^{\frac{2\pi i \cdot m - \ell}{N}}$ muss positive reelle Zahl sein. Die Zahlen der Form e^{ic} haben Betrag 1, also muss dann $e^{\frac{2\pi i \cdot m - \ell}{N}} = 1$ gelten. $e^{ix} = 1$ bedeutet aber, dass x ein ganzzahliges Vielfaches von 2π ist, also muss dann $\ell = m \in \mathbb{Z}/N\mathbb{Z}$ gelten, was im Widerspruch zu unserer Annahme $m \neq \ell \in \mathbb{Z}/N\mathbb{Z}$ steht. Das zeigt die Behauptung. □

Ende Vorlesung 15.5.

DIESE ORTHONORMALBASIS ist unwahrscheinlich nützlich. Eine Alltagsanwendung verwenden Sie, wenn Sie Tondateien auf Ihrem Mobiltelefon speichern oder versenden. Die Idee kam hier aus der Musik. Fassen wir einen Ton als Schwingung auf, so können wir die Schwingung als Funktion

$$f\colon [0, t] \to \mathbb{R}$$

auffassen, wobei t die Dauer bezeichnet – die ich gleich wieder auf 1 normieren werde – und $f(t)$ die Auslenkung des Lautsprechermembrans (oder Ihres Trommelfells) zum Zeitpunkt t beschreibt. Sie haben sicher schon einmal eine Grafik so einer Hüllkurve gesehen. Diese sehen auf den ersten Blick recht chaotisch aus.

Um Klänge von unterschiedlichen Quellen, zum Beispiel die Töne unterschiedlicher Instrumente, zu unterscheiden, war die Idee, zu versuchen, jeden Ton in „reine Töne“ oder „reine Schwingungen“ zu zerlegen, d.h. die Funktion f als (eventuell unendliche Summe) von Funktionen der Form

$$e_m(x) = \cos(2\pi m x) \text{ oder } f_m(x) = \sin(2\pi m x)$$

also als

$$f(x) = \sum_{m=-\infty}^{\infty} a_m e^{2\pi i m x}$$

zu schreiben und das Ergebnis so als Überlagerung von reinen Schwingungen zu interpretieren.

Das funktioniert ganz wunderbar und ist sehr nützlich, um Tonsignale zu analysieren, je größer m, desto höher der Ton, je größer der Koeffizient, desto lauter ist der Ton. Es ist allerdings nicht leicht zu zeigen, dass eine derartige Zerlegung für jede stetige Funktion möglich ist.

Bei digitalen Aufnahmen bekommen wir von der gedachten stetigen Funktion nur endlich viele Messwerte $F(\frac{k}{N})$, wobei N so groß ist, dass das Ohr höhere Töne als die Frequenz N nicht mehr wahrnimmt (üblich sind etwa 44.000 Messwerte pro Sekunde). Die gemessenen Werte sind also eine Funktion in unserem Raum Fun_N und in diesem Raum ist unsere Orthonormalbasis e_m genau durch die Funktionswerte der reinen Schwingungen $e^{2\pi i m x}$ gegeben.

Die Zerlegung in reine Schwingungen entspricht hier also einfach dem Ausdruck

$$F = \sum_{m=0}^{N-1} \langle F, e_m \rangle \cdot e_m,$$

die Koeffizienten $\langle F, e_m \rangle$ werden mit $\hat{F}(m)$ bezeichnet, die Umrechnung $F \to \hat{F}$ heißt *diskrete Fouriertransformation*. In dieser Notation finden wir also die Formel:

$$F(k) = \sum_{m=0}^{N-1} \hat{F}(m) \cdot e^{\frac{2\pi i}{N} \cdot m \cdot k},$$

die einfach die Koordinaten von F bezüglich unserer Orthonormalbasis aufschreibt.

Das .mp3-Format nutzt dies, um Tondaten stark zu komprimieren. Das menschliche Ohr hört nämlich einerseits nur Schwingungen in einem gewissen Frequenzbereich und zudem nehmen wir nur Töne einer gewissen Lautstärke wahr, wobei für sehr hohe und sehr tiefe Töne größere Lautstärken als im Mittelbereich notwendig sind. Zur Komprimierung werden daher einfach alle Koeffizienten weggelassen, die zu reinen Tönen gehören, die das menschliche Ohr nicht wahrnehmen kann.

Das gleiche Prinzip wird genauso zur Kompression von Bild- oder Videodateien verwendet.

Das mathematische Grundprinzip, dass es eine gute Idee ist Funktionenräumen in Eigenräume für eine Symmetriegruppe – hier war das die Gruppe $\mathbb{Z}/N\mathbb{Z}$ – zu zerlegen, wird Ihnen in der Mathematik an sehr vielen Stellen begegnen.

Unitäre Abbildungen

Bei der Berechnung zur diskreten Fouriertransformation ist uns eine Orthonormalbasis und eine lineare Abbildung begegnet, die ein Skalarprodukt erhalten hat. Bei euklidischen Vektorräumen hatten wir gesehen, dass diese beiden Begriffe eng zusammen hängen und wir wollen das nun noch auf unitäre Vektorräume übertragen.

Erinnerung: Version für euklidische Vektorräume:
Folgerung 19: *Eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ ist genau dann eine Orthonormalbasis bezüglich des Standardskalarproduktes, wenn für die Matrix $A = (\mathbf{v}_1 \ldots \mathbf{v}_n)$ mit Spaltenvektoren $\mathbf{v}_i$ gilt dass*

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

Lemma 34. *Eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{C}^n$ ist genau dann eine Orthonormalbasis bezüglich des Standardskalarproduktes, wenn für die Matrix $A = (\mathbf{v}_1 \ldots \mathbf{v}_n)$ mit Spaltenvektoren $\mathbf{v}_i$ gilt dass*

$$A^{-1} = \overline{A}^t.$$

Das Argument können wir einfach aus dem reellen Beweis abschreiben, wir müssen nur beachten, dass im Standardskalarpodukt auf $\mathbb{C}^n$ die Koordinaten des 2. Eintrags konjugiert werden.

Beweis. Der (i, j)-te Eintrag der Matrix $A^t \circ \overline{A} = (\mathbf{v}_1 \ldots \mathbf{v}_n)^t \circ \overline{(\mathbf{v}_1 \ldots \mathbf{v}_n)}$ ist per Definition genau

$$(\text{i-te Zeile von } A^t) \circ (\text{j-te Spalte von } \overline{A}) = \mathbf{v}_i^t \circ \overline{\mathbf{v}_j} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle.$$

Dieses Produkt ist also genau dann die Einheitsmatrix wenn $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ für $i \neq j$ und $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ für alle i ist.

Das ist genau die Bedingung dafür, dass $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Orthonormalbasis ist.

Die Bedingung $A^t \circ \overline{A} = E_n$ ist äquivalent zu $\overline{A}^t \circ A = \overline{E_n} = E_n$ und das bedeutet gerade $A^{-1} = \overline{A}^t$. □

Genauso können wir den Begriff der orthogonalen Abbildung übertragen.

Definition 35. Ist $V, \langle , \rangle$ ein unitärer Vektorraum, dann heißt eine lineare Abbildung $A \colon V \to V$ *unitär* genau dann, wenn für alle $\mathbf{v}, \mathbf{w} \in V$ gilt dass

$$\langle A\mathbf{v}, A\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle.$$

Ist $V = \mathbb{C}^n$ und $\langle , \rangle$ das Standardskalarprodukt, so wird die Menge der unitären $n \times n$-Matrizen mit

$$U(n) := \{A \in \mathrm{GL}_n(\mathbb{C}) \mid A \text{ unitär}\}$$

bezeichnet und heißt *unitäre Gruppe*.

Erinnerung:
Behauptung 27 Die orthogonalen Matrizen $O(n)$ sind genau die Matrizen, für die $A^t = A^{-1}$ gilt, d.h. das eine Matrix ist genau dann orthogonal wenn die Spaltenvektoren eine Orthonormalbasis von $\mathbb{R}^n$ bilden.

Behauptung 36. *Die unitären Matrizen $U(n)$ bestehen genau aus den Matrizen, für die $\overline{A}^t = A^{-1}$ gilt, d.h. das sind genau die Matrizen, deren Spaltenvektoren eine Orthonormalbasis von $\mathbb{C}^n$ bilden.*

Auch hier können wir den Beweis für euklidische Vektorräume wiederverwenden – darum ist es wichtig sich an Beweise zu erinnern. Es geht immer um das „Warum?“.

Beweis. Ist $A \in U(n)$ unitär, so sind die Spaltenvektoren eine Orthonormalbasis, denn die Spaltenvektoren sind gerade die Bilder $A\mathbf{e}_i$ der Einheitsvektoren $\mathbf{e}_i$ und weil A unitär ist, gilt darum

$$\langle A\mathbf{e}_i, A\mathbf{e}_j \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}.$$

Das bedeutet gerade, dass die Spalten eine Orthonormalbasis sind.

Sind umgekehrt die Spalten $A\mathbf{e}_i$ von A eine Orthonormalbasis von $\mathbb{C}^n$, so können wir zeigen, dass $\langle A\mathbf{v}, A\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle$ für alle $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$ gilt, denn wir können die Vektoren $\mathbf{v} = \sum_{i=1}^n a_i\mathbf{e}_i, \mathbf{w} = \sum_{j=1}^n b_j\mathbf{e}_j$ als Linearkombination der Standardbasisvektoren von $\mathbb{C}^n$ schreiben und damit $\langle A\mathbf{v}, A\mathbf{w}\rangle$ ausrechnen:

$$\begin{aligned}
\langle A\mathbf{v}, A\mathbf{w}\rangle &= \langle A(\sum_{i=1}^n a_i\mathbf{e}_i), A(\sum_{j=1}^n b_j\mathbf{e}_j)\rangle && \text{Formel für } \mathbf{v}, \mathbf{w} \\
&= \langle \sum_{i=1}^n a_i A\mathbf{e}_i, \sum_{j=1}^n b_j A\mathbf{e}_j)\rangle && A \text{ linear} \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i\overline{b_j}\langle A\mathbf{e}_i, A\mathbf{e}_j\rangle && \langle,\rangle \text{ sesquilinear} \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i\overline{b_j}\langle \mathbf{e}_i, \mathbf{e}_j\rangle && \text{Spalten ON-Basis} \\
&= \langle \sum_{i=1}^n a_i\mathbf{e}_i, \sum_{j=1}^n b_j\mathbf{e}_j\rangle && \langle,\rangle \text{ sesquilinear} \\
&= \langle \mathbf{v}, \mathbf{w}\rangle. && \text{Formel für } \mathbf{v}, \mathbf{w}
\end{aligned}$$

Das war zu zeigen. □

Im Beweis zur Fouriertransformation hatten wir noch zwei Eigenschaften von unitären Abbildungen bemerkt:

Lemma 37. *Sei $(V, \langle,\rangle)$ ein unitärer Vektorraum und $A\colon V \to V$ eine unitäre Abbildung, dann gilt:*

1. *Die Eigenwerte von A haben Betrag 1.*
2. *Eigenvektoren zu verschiedenen Eigenwerten von A sind orthogonal.*

Beweis. 1. Für die erste Aussage können wir das Argument für orthogonale Abbildungen kopieren: Sei $c \in \mathbb{C}$ ein Eigenwert und $\mathbf{v} \in V \setminus \{0\}$ ein Eigenvektor, dann gilt

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{v}\rangle &= \langle A\mathbf{v}, A\mathbf{v}\rangle && A \text{ unitär} \\
&= \langle c\mathbf{v}, c\mathbf{v}\rangle && \mathbf{v} \text{ Eigenvektor} \\
&= c\bar{c}\langle \mathbf{v}, \mathbf{v}\rangle = |c|^2\langle \mathbf{v}, \mathbf{v}\rangle && \langle,\rangle \text{ sesquilinear.}
\end{aligned}$$

Da $\langle \mathbf{v}, \mathbf{v}\rangle > 0$ folgt daraus, dass $|c| = 1$ und das war zu zeigen.

2. Seien $\mathbf{v}, \mathbf{w} \in V \setminus \{0\}$ Eigenvektoren zu Eigenwerten $c, c' \in \mathbb{C}$, dann können wir wie zuvor rechnen:

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{w}\rangle &= \langle A\mathbf{v}, A\mathbf{w}\rangle && A \text{ unitär} \\
&= \langle c\mathbf{v}, c'\mathbf{w}\rangle && \mathbf{v}, \mathbf{w} \text{ Eigenvektoren} \\
&= c\overline{c'}\langle \mathbf{v}, \mathbf{w}\rangle && \langle,\rangle \text{ sesquilinear.}
\end{aligned}$$

Also gilt entweder $\langle \mathbf{v}, \mathbf{w}\rangle = 0$ wie behauptet, oder $c\overline{c'} = 1$, also $c^{-1} = \overline{c'}$. Da nach 1. gilt dass $|c| = 1$, ist $c^{-1} = \bar{c}$ also folgt dann $c = c'$. Das ist ein Widerspruch zur Voraussetzung, dass $c \neq c'$.

□

Der Spektralsatz

Wir können nun auf die Frage zurück kommen, für welche linearen Abbildungen $A\colon V \to V$ eines euklidischen (oder unitären) Vektorraums wir eine Orthogonalbasis aus Eigenvektoren finden können.

Für $\mathbb{R}^n$ mit dem Standardskalarprodukt können wir eine notwendige Bedingung hierfür aus der Formel für den Basiswechsel ablesen.

Ist Matrix $A\colon \mathbb{R}^n \to \mathbb{R}^n$ eine Matrix und $\underline{\mathbf{n}}$ eine Orthonormalbasis aus Eigenvektoren, $Q = (\mathbf{n}_1 \dots \mathbf{n}_n)$ die zugehörige orthonormale Matrix so ist

$$_{\underline{\mathbf{n}}}\mathrm{Mat}_{\underline{\mathbf{n}}}(A) = Q^{-1}AQ = D$$

eine Diagonalmatrix.

Multiplizieren wir die Gleichung von links mit Q, von rechts mit Q^{-1} und verwenden, dass A orthogonal ist und als $Q^{-1} = Q^t$ gilt, so erhalten wir

$$A = QDQ^{-1} = QDQ^t.$$

Transponieren wir diese Gleichung so finden wir:

Erinnerung: Für alle Matrizen B, C gilt

$$(B \circ C)^t = C^t \circ B^t.$$

$$A^t = (Q \circ D \circ Q^t)^t = (Q^t)^t \circ D^t \circ Q^t = Q \circ D \circ Q^t = A,$$

also

$$A^t = A$$

was für die Einträge $A = (a_{ij})_{i,j=1,\dots n}$ der Matrix bedeutet, dass $a_{ij} = a_{ji}$ gilt. Matrizen diese Form heißen *symmetrisch*.

Eine symmetrische Matrix:

$$\begin{pmatrix} \star & a & b & c \\ a & \star & d & e \\ b & d & \star & f \\ c & e & f & \star \end{pmatrix}$$

Definition 38. 1. Eine Matrix $A = (a_{ij})_{i,j=1,\dots n} \in \mathrm{Mat}_{n,n}(K)$ heißt *symmetrisch*, falls $A = A^t$.

2. Eine komplexe Matrix $A = (a_{ij})_{i,j=1,\dots n} \in \mathrm{Mat}_{n,n}(\mathbb{C})$ heißt *hermitesch*, falls $A = \overline{A}^t$.

Wir werden am Ende dieses Abschnittes zeigen, dass umgekehrt tatsächlich jede symmetrische Matrix eine Orthonormalbasis aus Eigenvektoren besitzt, insbesondere also diagonalisierbar ist. Wir werden dann auch noch sehen, wo symmetrische Matrizen auftauchen, für die das ein nützliches Resultat ist.

Wie bei der Zerlegung in Haupträume ist es einfacher, wenn wir statt mit Matrizen, mit linearen Abbildungen argumentieren, da wir dann in einem Induktionsbeweis das Resultat für Unterräume $W \subset \mathbb{R}^n$ verwenden können.

Das Problem ist, dass wir die transponierte Matrix A^t einfach in Termen der Einträge definiert hatten, wir sollten uns also überlegen, wie wir die Eigenschaft $A = A^t$ als Eigenschaft einer linearen Abbildung des euklidischen Vektorraums $\mathbb{R}^n, \langle,\rangle_{Std}$ umformulieren können.

Frage: Wie können wir die Einträge a_{ij} der Matrix A aus der linearen Abbildung $A\colon \mathbb{R}^n \to \mathbb{R}^n$ berechnen?

In der j-ten Spalte von A stehen die Koordinaten des Bildes des j-ten Standardbasisvektors $A\mathbf{e}_j$. Den i-ten Eintrag des Vektors von $A\mathbf{e}_j$ können wir also als $a_{ij} = \langle A\mathbf{e}_j, \mathbf{e}_i \rangle$ berechnen (Folgerung 17). Die Gleichung $a_{ij} = a_{ji}$ bedeutet weil $\langle , \rangle$ symmetrisch ist also

$$\langle A\mathbf{e}_j, \mathbf{e}_i \rangle = \langle \mathbf{e}_j, A\mathbf{e}_i \rangle.$$

Das motiviert den folgenden Versuch, die Eigenschaft für Abbildungen von euklidischen Vektorräumen zu formulieren:

Definition 39. Sei $V, \langle , \rangle$ ein euklidischer (oder ein unitärer) Vektorraum. Dann heißt eine lineare Abbildung $A\colon V \to V$ *selbstadjungiert* wenn

$$\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle \text{ für alle } \mathbf{v}, \mathbf{w} \in V.$$

Wir sollten zunächst prüfen, dass diese Definition tatsächlich leisten, was wir möchten.

Behauptung 40. *Ist $V, \langle , \rangle$ ein endlichdimensionaler euklidischer (bzw. unitärer) Vektorraum dann sind für eine lineare Abbildung $A\colon V \to V$ die folgenden Eigenschaften äquivalent:*

1. *A ist selbstadjungiert.*
2. *Für jede Orthonormalbasis $\underline{\mathbf{v}} = \mathbf{v}_1, \dots, \mathbf{v}_n$ von V ist die Matrix ${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)$ symmetrisch (bzw. hermitesch).*
3. *Es gibt eine Orthonormalbasis $\underline{\mathbf{v}} = \mathbf{v}_1, \dots, \mathbf{v}_n$ von V für die die Matrix ${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)$ symmetrisch (bzw. hermitesch) ist.*

Insbseondere ist ein lineare Abbildungen $A\colon \mathbb{R}^n \to \mathbb{R}^n$ genau dann selbstadjungiert, wenn die A eine symmetrische Matrix ist.

Beweis. Wir zeigen zunächst „(1)⇒(2)“. Sei also A selbstadjungiert und $\underline{\mathbf{v}} = \mathbf{v}_1, \dots, \mathbf{v}_n$ eine Orthonormalbasis von V. Die Einträge a_{ij} der Matrix $(a_{ij})_{i,j=1,\dots n} = {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)$ sind die i-ten Koordinaten des Vektors $A\mathbf{v}_j$ bezüglich der Basis $\underline{\mathbf{v}}$. Da $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Orthonormalbasis ist, können wir die i-te Koordinate eines Vektors $\mathbf{v}$ durch die Formel $\langle \mathbf{v}, \mathbf{v}_i \rangle$ berechnen, d.h. es gilt für alle i, j dass $a_{ij} = \langle A\mathbf{v}_j, \mathbf{v}_i \rangle$. Da wir vorausgesetzt hatten, dass A selbstadjungiert ist, folgt daraus

$$\begin{aligned} a_{ij} &= \langle A\mathbf{v}_j, \mathbf{v}_i \rangle && \text{Formel für Koordinaten (Folgerung 17)} \\ &= \langle \mathbf{v}_j, A\mathbf{v}_i \rangle && \text{weil } A \text{ selbstadjungiert} \\ &= \langle A\mathbf{v}_i, \mathbf{v}_j \rangle && \text{weil } \langle , \rangle \text{ symmetrisch} \\ &= a_{ji} && \text{Formel für Koordinaten (Folgerung 17).} \end{aligned}$$

Das zeigt die Behauptung, dass ${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)$ symmetrisch ist.

Haben wir statt einem euklidischen einen unitären Vektorraum, d.h. ist $\langle , \rangle$ hermitesch statt symmetrisch, so müssen gilt in der obigen Rechnung stattdessen $\langle \mathbf{v}_j, A\mathbf{v}_i \rangle = \overline{\langle A\mathbf{v}_i, \mathbf{v}_j \rangle}$ also folgt dann $a_{ij} = \overline{a_{ji}}$, d.h. die Matrix ist dann hermitesch.

Die Folgerung „(2)⇒(3)“ ist klar, weil wir gezeigt haben, dass V eine Orthonormalbasis besitzt und darum die Bedingung (2) nicht leer ist.

Wir zeigen schließlich dass „(3)⇒(1)“ gilt. Sei also ${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A)$ symmetrisch, d.h. es gilt für alle i, j, dass $\langle A\mathbf{v}_j, \mathbf{v}_i\rangle = \langle \mathbf{v}_j, A\mathbf{v}_i\rangle$. Zu zeigen ist dann, dass A selbstadjungiert ist, d.h., dass

$$\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle \text{ für alle } \mathbf{v}, \mathbf{w} \in V$$

gilt.

Erster Beweis: Diese Aussage folgt, weil wir $\mathbf{v}, \mathbf{w}$ als Linearkombinationen der Basisvektoren schreiben können, A linear ist und $\langle,\rangle$ bilinear ist: Sind $\mathbf{v}, \mathbf{w} \in V$ gegeben, so können wir $\mathbf{v} = \sum_{i=1}^n a_i\mathbf{v}_i, \mathbf{w} = \sum_{j=1}^n b_j\mathbf{v}_j$ als Linearkombinationen der Basisvektoren $\underline{\mathbf{v}}$ schreiben. Dann ist

$$\begin{aligned}
\langle A\mathbf{v}, \mathbf{w}\rangle &= \langle A(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n), b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n\rangle && \text{Linearkombination eingesetzt}\\
&= \langle (a_1A\mathbf{v}_1 + \cdots + a_nA\mathbf{v}_n), b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n\rangle && A \text{ linear}\\
&= \sum_{i=1}^n a_i\langle A\mathbf{v}_i, \sum_{j=1}^n b_j\mathbf{v}_j\rangle && \langle,\rangle \text{ linear in 1. Variablen}\\
&= \sum_{i=1}^n a_i \sum_{j=1}^n \overline{b_j}\langle A\mathbf{v}_i, \mathbf{v}_j\rangle && \langle,\rangle \text{sesquilinear in 2. Variablen}\\
&= \sum_{i=1}^n a_i \sum_{j=1}^n \overline{b_j}\langle \mathbf{v}_i, A\mathbf{v}_j\rangle && {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) \text{symmetrisch (bzw. hermitesch)}\\
&= \sum_{i=1}^n a_i\langle \mathbf{v}_i, \sum_{j=1}^n b_jA\mathbf{v}_j\rangle && \langle,\rangle \text{sesquilinear in 2. Variablen}\\
&= \langle \sum_{i=1}^n a_i\mathbf{v}_i, \sum_{j=1}^n b_jA\mathbf{v}_j\rangle && \langle,\rangle \text{linear in 1. Variablen}\\
&= \langle \mathbf{v}, A\mathbf{w}\rangle && A \text{ linear und Linearkomb. eingesetzt.}
\end{aligned}$$

Zweiter Beweis: Statt die Gleichung durch Einsetzen von Linearkombinationen nachzurechnen, können wir auch abstrakter das Argument verwenden, dass lineare Abbildungen eindeutig durch ihre Werte auf einer Basis bestimmt sind: Zunächst sind für alle i die Abbildungen

$$\begin{aligned}
&\langle A\mathbf{v}_i, __\rangle : V \to \mathbb{R}\mathbf{w} \mapsto \langle A\mathbf{v}_i, \mathbf{w}\rangle && \text{und}\\
&\langle \mathbf{v}_i, A__\rangle : V \to \mathbb{R}\mathbf{w} \mapsto \langle \mathbf{v}_i, A\mathbf{w}\rangle && \text{linear,}
\end{aligned}$$

da A linear und $\langle,\rangle$ linear in der 2. Variablen ist. Nach Voraussetzung stimmen diese Abbildungen auf der Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ überein, sind also gleich, d.h. es gilt

$$\langle A\mathbf{v}_i, \mathbf{w}\rangle = \langle \mathbf{v}_i, A\mathbf{w}\rangle \text{ für alle } \mathbf{w} \in V, i = 1, \ldots n.$$

Dann stimmen aber für alle $\mathbf{w} \in V$ die linearen Abbildungen

$$\begin{aligned}
&\langle A__, \mathbf{w}\rangle : V \to \mathbb{R}\mathbf{v} \mapsto \langle A\mathbf{v}, \mathbf{w}\rangle && \text{und}\\
&\langle __, A\mathbf{w}\rangle : V \to \mathbb{R}\mathbf{v} \mapsto \langle \mathbf{v}, A\mathbf{w}\rangle
\end{aligned}$$

auf den Basisvektoren $\mathbf{v}_1, \dots, \mathbf{v}_n$ überein, sind also gleich, d.h. $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle$ für alle $\mathbf{v}, \mathbf{w} \in V$. Also ist A selbstadjungiert und das war zu zeigen. $\square$

Damit können wir jetzt den Spektralsatz zeigen:

Satz 41 (Spektralsatz)**.** *Sei $V, \langle,\rangle$ ein endlichdimensionaler euklidischer (oder ein unitärer) Vektorraum. Dann existiert für jede selbstadjungierte Abbildung $A\colon V \to V$ eine Orthonormalbasis aus Eigenvektoren von A.*

Insbesondere sind also symmetrische reelle Matrizen diagonalisierbar.

Beweis. Schritt 1: Wir zeigen zunächst, dass das charakteristische Polynom $\det(A - t\,\mathrm{id}_V)$ von A in reelle Linearfaktoren zerfällt.

Sei dazu $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Orthonormalbasis von V. Dann ist die Matrix $M = {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) \in \mathrm{Mat}_{n,n}(\mathbb{R}) \subset \mathrm{Mat}_{n,n}(\mathbb{C})$ symmetrisch mit reellen Einträgen, also auch hemitesch, definiert also eine lineare Abbildung $M\colon \mathbb{C}^n \to \mathbb{C}^n$, die bezüglich des Standardskalarproduktes selbstadjungiert ist. Da $\mathbb{C}$ algebraisch abgeschlossen ist (das war der Fundamentalsatz der Algebra), zerfällt das charakteristische Polynom $\det(M - tE_n) = \det(A - t\,\mathrm{id}_V) = \prod_{i=1}^n (c_i - t)$ in $\mathbb{C}[t]$ in Linearfaktoren. Wir müssen also zeigen, dass die Eigenwerte $c_i \in \mathbb{C}$ reell sind. Das gilt, weil A selbstadjungiert ist, dann ist $\mathbf{v} \in \mathbb{C}^n$ ein Eigenvektor zum Eigenwert c_i, dann folgt aus $\langle A\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{v}, A\mathbf{v}\rangle$, dass

$$\begin{aligned} c_i\langle \mathbf{v}, \mathbf{v}\rangle = \langle c_i\mathbf{v}, \mathbf{v}\rangle &= \langle A\mathbf{v}, \mathbf{v}\rangle \\ &= \langle \mathbf{v}, A\mathbf{v}\rangle \\ &= \langle \mathbf{v}, c_i\mathbf{v}\rangle = \overline{c_i}\langle \mathbf{v}, \mathbf{v}\rangle. \end{aligned}$$

Daraus folgt aber $c_i = \overline{c_i}$, also $c_i \in \mathbb{R}$. Ende Vorlesung 22.5

Schritt 2: Wir konstruieren nun mittels Induktion über $\dim V$ eine Orthonormalbasis aus Eigenvektoren von A. Für $\dim V = 1$ ist nichts zu zeigen, denn dann ist jeder Vektor $\mathbf{v} \in V \setminus \{0\}$ ein Eigenvektor von A.

Angenommen die Aussage stimmt für alle euklidischen Vektorräume der Dimension k, dann müssen wir zeigen, dass die Aussage auch gilt wenn $\dim V = k+1$ ist. Sei also $\dim V = k+1$ und $A\colon V \to V$ selbstadjungiert. Nach Schritt 1 gibt es einen reellen Eigenwert c_{k+1} von A und also auch einen Eigenvektor $\mathbf{v} \in V$ mit $A\mathbf{v} = c_{k+1}\mathbf{v}$ und $\|v\| = 1$.

Ich behaupte, dass A den k-dimensionalen Unterraum $W := (\mathbf{v})^\perp \subset V$ in sich selbst abbildet, d.h.

$$\begin{aligned} A|_W\colon W &\to W \\ \mathbf{w} &\mapsto A\mathbf{w} \end{aligned}$$

ist wohldefiniert. Um das nachzuprüfen müssen wir zeigen, das für alle $\mathbf{w} \in W = (\mathbf{v})^\perp$ auch $A\mathbf{w} \in (\mathbf{v})^\perp$ gilt, d.h. $\langle A\mathbf{w}, \mathbf{v}\rangle \stackrel{?}{=} 0$. Das gilt aber, weil

$$\langle A\mathbf{w}, \mathbf{v}\rangle \stackrel{A \text{ selbstadj.}}{=} \langle \mathbf{w}, A\mathbf{v}\rangle \stackrel{\mathbf{v}\ \mathrm{EV}}{=} \langle \mathbf{w}, c\mathbf{v}\rangle \stackrel{\langle\rangle \text{ bilin.}}{=} c\langle \mathbf{w}, \mathbf{v}\rangle \stackrel{\mathbf{w}\in(\mathbf{v})^\perp}{=} 0.$$

Da $A|_W$ selbstadjungiert ist, existiert nach Induktionsvoraussetzung eine Orthonormalbasis $\mathbf{v}_1, \dots, \mathbf{v}_k$ von W die aus Eigenvektoren von A besteht.

Die Basis $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}$ sind dann eine Basis von V, die aus Eigenvektoren von A besteht. □

DAS WORT „SPEKTRALSATZ“, beziehungsweise die Gewohnheit, die Menge der Eigenwerte einer linearen Abbildung auch als Spektrum der Abbildung zu bezeichnen, hat Ihren Ursprung sicherlich im Beispiel der Zerlegung von Tönen in reine Schwingungen, das wir im Abschnitt zu unitären Vektorräumen kennengelernt hatten. In diesem Beispiel waren die Frequenzen gerade die Eigenwerte einer linearen Abbildung. Die mathematische Formulierung der Operation eine Funktion als Linearkombination periodischen Funktionen $e^{imx} = \cos(mx) + i\sin(mx)$ zu schreiben wird auch verwendet um die Spektralzerlegung von Lichtwellen zu beschreiben. In der Physik nimmt das zu Grunde liegende Prinzip, eine Funktionen als Linearkombinationen von orthogonalen Eigenfunktionen eines Operatoren zu schreiben und das als Überlagerung von Zuständen zu interpretieren, in der Quantenmechanik einen zentralen Platz ein. Hierfür war es wichtig, eine Version für unendlichdimensionale Vektorräume zu finden, die meist eines der Ziele der Funktionalanalysis Vorlesung ist.

WIR HABEN FÜR DIESES RESULTAT ZWEI NÜTZLICHE STRATEGIEN zur Lösung von mathematischen Probleme genutzt:

1. Zunächst haben wir konkret mit Matrizen gerechnet und dabei eine Eigenschaft „symmetrisch“ von Matrizen bemerkt. Für die Argumentation war es ganz wesentlich, dafür eine koordinatenfreie Beschreibung „selbstadjungiert“ zu finden.

2. Ein technisches Problem (Hier: Wie finden wir reelle Eigenwerte?) ist verschwunden, indem wir unsere Begriffe verallgemeinert haben (Hier: Wir erlauben vorübergehend $\mathbb{C}$-Vektorräume). Dabei kommt $\mathbb{C}$ weder in der ursprünglichen Frage noch in der Antwort vor.

Schon mit der Formel

$$\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle$$

rechnet es sich viel leichter, als mit der Liste der Bedingungen

$$a_{ij} = a_{ji} \text{ für } i, j = 1, \dots n.$$

Anders gesagt: Ich kann nur beschränkt viel gleichzeitig im Kopf haben, darum ist es notwendig, die einzelnen Dinge so kompakt wie möglich zu fassen.

ALS ANWENDUNGEN DES SPEKTRALSATZES möchte ich Ihnen in den nächsten Vorlesungen drei ganz unterschiedlich aussehende vorstellen:

1. Analysis: Extremwerte für Funktionen in mehreren Variablen.

2. Geometrie: Welche Matrizen definieren Skalarprodukte? Wie können wir quadratische Gleichungen in mehreren Variablen verstehen?

3. Datenanalyse: Was ist der „wesentliche“ Teil einer Matrix? (Auch „Singulärwertzerlegung“ genannt, das ist nützlich für Kompression und ein Baustein für die Algorithmen die Ihnen Vorschläge für „Was Sie als nächstes interessieren könnte“ generieren.[6])

[6] Die Blase in die Sie dabei von der Software gesteckt werden, heißt womöglich Hauptachse oder Eigenraum.

Anwendung 1: Mehrdimensionale Extremwerte

Sie wissen aus der Schule wahrscheinlich, dass die zweite Ableitung einer Funktion die Krümmung des Funktionsgraphen beschreibt. In einer Dimension kann Ihnen das Vorzeichen der zweiten Ableitung an den Extremstellen (die x an denen $f'(x) = 0$) sagen, ob Sie ein lokales Minimum, oder ein lokales Maximum gefunden haben.

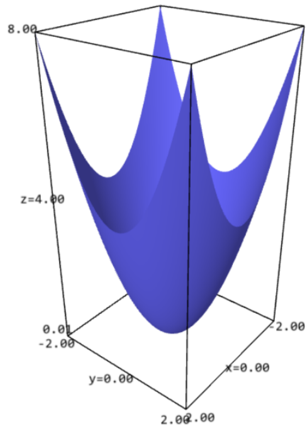


Abbildung 2: $f(x,y) = x^2 + y^2$

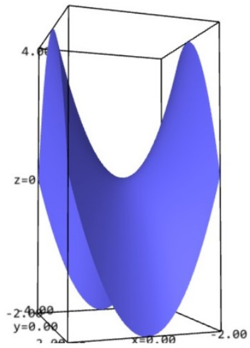


Abbildung 3: $g(x,y) = x^2 - y^2$

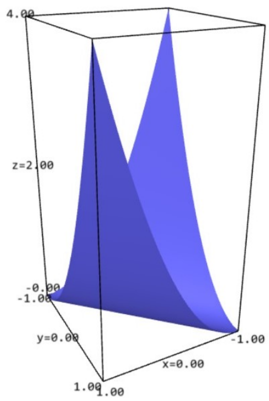


Abbildung 4: $h(x,y) = x^2 + xy + y^2$

Wenn Sie eine (2 mal stetig differenzierbare) Funktion $F\colon \mathbb{R}^n \to \mathbb{R}$ in mehreren Variablen betrachten, so beschreibt ganz analog die Matrix der 2ten Ableitungen $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}F$ das Krümmungsverhalten der Funktion und wenn Sie einen linearen Koordinatenwechsel $\underline{x} \mapsto \underline{\tilde{x}} = Q \cdot \underline{x}$ vornehmen, können Sie die Krümmung mit linearer Algebra umrechnen, weil die Ableitung selbst linear ist.

Es stellt sich heraus, dass die Matrix $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}F$ symmetrisch ist. Daher sagt uns der Spektralsatz, dass wir Koordinaten finden können, so dass die Krümmung eine Diagonalmatrix wird. Die Vorzeichen der Eigenwerte geben dann an, in welche Richtungen der Funktionsgraph nach oben bzw. nach unten gekrümmt ist. Insbesondere werden Sie in der Analysis sehen, dass Sie damit wie in einer Dimension lokale Minima und Maxima charakterisieren können:

1. Gilt an einer kritischen Stelle $\mathbf{v}$ (d.h. $(\frac{\partial}{\partial x_i}F)(\mathbf{v}) = 0$ für alle i), dass alle Eigenwerte der Matrix $(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}F)(\mathbf{v})$ an der Stelle $\mathbf{v} \in \mathbb{R}^n$ positiv sind, so ist $\mathbf{v}$ ein lokales Minimum.
2. Gilt an einer kritischen Stelle $\mathbf{v}$ (d.h. $(\frac{\partial}{\partial x_i}F)(\mathbf{v}) = 0$ für alle i), dass alle Eigenwerte der Matrix $(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j})F(\mathbf{v})$ an der Stelle $\mathbf{v} \in \mathbb{R}^n$ negativ, so ist $\mathbf{v}$ ein lokales Maximum.
3. Finden Sie an einer Extremstelle sowohl positive als auch negative Eigenwerte der Matrix der 2ten Ableitungen, so ist die Extremstelle ein Sattelpunkt.

Anwendung 2: Quadratische Formen und Bilinearformen

Die Anwendung zur Berechnung von Extremwerten, können wir für quadratische Funktionen fast ohne Analysis verstehen und bei der Gelegenheit noch die Frage beantworten, welche Matrizen Skalarprodukte auf $\mathbb{R}^n$ definieren.

Erinnerung wir haben schon gesehen, dass symmetrische Matrizen $A \in \mathrm{Mat}_{n,n}(K)$ mittels der Formel

$$B_A(\mathbf{v},\mathbf{w}) := \mathbf{v}^t \circ A \circ \mathbf{w}$$

symmetrische Bilinearformen definieren, d.h. die Abbildung

$$B_A(\,,\,)\colon K^n \times K^n \to K(\mathbf{v},\mathbf{w}) \qquad \mapsto B_A(\mathbf{v},\mathbf{w}) = \mathbf{v}^t \circ A \circ \mathbf{w}$$

ist in beiden Variablen linear und erfüllt $B_A(\mathbf{v},\mathbf{w}) = B_A(\mathbf{w},\mathbf{v})$ wegen der Rechenregel $(B \circ C)^t = C^t \circ B^t$. In Koordinatenschreibweise können wir $B_A(\,,\,)$ für eine Matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ wie folgt ausrechnen:

$$B_A\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right) = (x_1 \dots x_n) \circ (a_{ij})_{i,j=1,\dots,n} \circ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$
$$= \sum_{i=1}^{n} x_i \Big(\sum_{j=1}^{n} a_{ij} y_j\Big) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j.$$

Bemerkung (Koordinatenwechsel). Wenn wir statt der Standardbasis auf K^n eine beliebige Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ verwenden, können wir für Vektoren $\mathbf{v} = \sum_{i=1}^{n} x'_i \mathbf{v}_i$ und $\mathbf{w} = \sum_{i=1}^{n} y'_i \mathbf{v}_i$, d.h. $\mathbf{v}, \mathbf{w}$ haben $\underline{\mathbf{v}}$-Koordinaten $\underline{x}' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \mathrm{Koord}_{\mathbf{v}}(\mathbf{v}), \underline{y}' = \begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix} = \mathrm{Koord}_{\mathbf{v}}(\mathbf{w})$ die Bilinearform auch in $\underline{v}$-Koordinaten umrechnen, denn ist $P = (\mathbf{v}_1 \dots \mathbf{v}_n)$ die Matrix der neuen Basisvektoren, dann gilt:

$$\begin{aligned} B_A(\mathbf{v}, \mathbf{w}) &= B_A(P\underline{x}', P\underline{y}') \\ &= (P\underline{x}')^t \circ A \circ (P\underline{y}') \\ &= \underline{x}'^t \circ (P^t A P) \circ \underline{y}\prime. \end{aligned}$$

Vorsicht: Der Basiswechsel für die Bilinearform ist also

$$A \to P^t A P,$$

aber für die Matrix der linearen Abbildung A hätten wir

$${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A) = P^{-1} A P.$$

Diese Formeln stimmen nur für orthogonale P überein.

Da es bisher praktisch war, Begriffe ohne Koordinatenwahl für allgemeine Vektorräume zu formulieren, sollten wir das auch für bilineare Abbildungen tun.

Definition. Eine *Bilinearform* B auf einem K-Vektorraum V ist eine Abbildung

$$\begin{aligned} B \colon V \times V &\to K \\ (\mathbf{v}, \mathbf{w}) &\mapsto B(\mathbf{v}, \mathbf{w}) \end{aligned}$$

die in beiden Argumenten linear ist, d.h, für die

$$\begin{aligned} B(a\mathbf{v} + b\mathbf{v}', \mathbf{w}) &= aB(\mathbf{v}, \mathbf{w}) + bB(\mathbf{v}', \mathbf{w}) \quad &&\text{und} \\ B(\mathbf{v}, a\mathbf{w} + b\mathbf{w}') &= aB(\mathbf{v}, \mathbf{w}) + bB(\mathbf{v}, \mathbf{w}') \quad &&\text{für alle } a, b \in K, \mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in V. \end{aligned}$$

Eine *symmetrische Bilinearform*, ist eine Bilinearform für die außerdem

$$B(\mathbf{v}, \mathbf{w}) = B(\mathbf{w}, \mathbf{v}) \qquad \text{für alle } \mathbf{v}, \mathbf{w} \in V \text{ gilt.}$$

Sie hatten auf dem 6. Übungsblatt schon nachgerechnet, dass wir Bilinearformen B nach Wahl einer Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ durch die Matrix der Werte $(B(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1\dots,n}$ beschreiben können, d.h. der abstrakte Begriff der Bilinearform ist tatsächlich eine kompakte Formulierung der Rechenregeln, die wir für $B_A(\mathbf{v}, \mathbf{w})$ gesehen haben. Das können wir hier noch einmal formal festhalten.

Lemma/Definition 42. *Sei $B\colon V \times V \to K$ eine Bilinearform auf einem K-Vektorraum V und $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Basis von V. Dann bezeichnen wir die $n \times n$-Matrix*

$$\mathrm{Mat}_{\underline{\mathbf{v}}}(B) := (B(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1,\dots n} \in \mathrm{Mat}_{n,n}(K)$$

als Matrix der Bilinearform bezüglich der Basis $\underline{v}$.

Sind $\mathbf{v} = \sum_{i=1}^n x_i \mathbf{v}_i$ und $\mathbf{w} = \sum_{i=1}^n y_i \mathbf{v}_i$, Vektoren mit $\underline{\mathbf{v}}$-Koordinaten $\mathrm{Koord}_{\underline{\mathbf{v}}}(\mathbf{v}) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\mathrm{Koord}_{\underline{\mathbf{v}}}(\mathbf{w}) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ *so gilt*

$$\begin{aligned} B(\mathbf{v}, \mathbf{w}) &= \begin{pmatrix} x_1 \dots x_n \end{pmatrix} (B(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1,\dots n} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \mathrm{Koord}_{\underline{\mathbf{v}}}(\mathbf{v})^t \circ \mathrm{Mat}_{\underline{\mathbf{v}}}(B) \circ \mathrm{Koord}_{\underline{\mathbf{v}}}(\mathbf{w}). \end{aligned}$$

Beweis. Die Formel folgt wie in der Übungsaufgabe aus der Bilinearität von B und den Rechenregeln für die Matrixmultiplikation:

$$\begin{aligned} B(\mathbf{v}, \mathbf{w}) &= B(x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n, y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n) && \text{linear in 1. Eintrag} \\ &= \sum_{i=1}^n x_i B(\mathbf{v}_i, y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n) && \text{linear in 2. Eintrag} \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i y_j B(\mathbf{v}_i, \mathbf{v}_j) \\ &= \begin{pmatrix} x_1 \dots x_n \end{pmatrix} \circ (B(\mathbf{v}_i, \mathbf{v}_j))_{i,j=1,\dots n} \circ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \end{aligned}$$

□

Zurück zu quadratischen Gleichungen. Wie bei Skalarprodukten sind symmetrische Bilinearformen schon durch die Werte $B(\mathbf{v}, \mathbf{v})$ für alle $\mathbf{v} \in V$ eindeutig bestimmt wenn $2 \neq 0$ in K, denn dann ist:

$$\begin{aligned} B(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) &= B(\mathbf{v}, \mathbf{v} + \mathbf{w}) + B(\mathbf{w}, \mathbf{v} + \mathbf{w}) \\ &= B(\mathbf{v}, \mathbf{v}) + B(\mathbf{v}, \mathbf{w}) + B(\mathbf{w}, \mathbf{v}) + B(\mathbf{w}, \mathbf{w}) \\ &= B(\mathbf{v}, \mathbf{v}) + 2 \cdot B(\mathbf{v}, \mathbf{w}) + B(\mathbf{w}, \mathbf{w}). \end{aligned}$$

Also

$$2 \cdot B(\mathbf{v}, \mathbf{w}) = B(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) - B(\mathbf{v}, \mathbf{v}) - B(\mathbf{w}, \mathbf{w}).$$

Beispiel 43. Für unser Beispiel der Bilinearform B_A zu einer symmetrischen Matrix A ist

$$\begin{aligned} B_A(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i y_j \\ &= \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \cdot (\sum_{i<j} a_{ij} x_i x_j). \end{aligned}$$

Die quadratische Funktion $q(x_1, x_2) = x_1^2 + 3x_1x_2 + x_2^2$ erhalten wir also aus der Bilinearform auf K^2 die durch $A = \begin{pmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & 1 \end{pmatrix}$ gegeben ist.

Der Spektralsatz sagt uns also, dass wir zu jeder quadratischen Funktion

$$\begin{aligned} q \colon \mathbb{R}^n &\to \mathbb{R} \\ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto q(x_1, \dots, x_n) = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \cdot (\sum_{i<j} a_{ij} x_i x_j) \end{aligned}$$

eine Orthonormalbasis $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$ finden können, so dass q in $\underline{\mathbf{v}}$-Koordinaten $(x'_1, \dots, x'_n)$ durch eine Formel der Form

$$q(\mathbf{v}) = c_1 x_1'^2 + \cdots + c_n x_n'^2 \text{ wobei } \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix} = \mathrm{Koord}_{\mathbf{v}}(\mathbf{v})$$

gegeben ist. Hierbei sind $c_1, \dots, c_n$ die Eigenwerte der Matrix $A = (a_{ij}) \in \mathrm{Mat}_{n,n}(\mathbb{R})$.

Bemerkung. 1. In den Wirtschaftswissenschaften treten Extremwertaufgaben für quadratische Funktionen in der Form: Maximieren Sie $q(\mathbf{v})$ für $\mathbf{v} \in \mathbb{R}^n$ mit $\|v\| = 1$ auf.

Der Spektralsatz sagt uns hierfür, dass das Maximum hier durch die Eigenvektoren von A zum größten Eigenwert gegeben ist

2. In der Zahlentheorie ist die Frage was die rationalen Lösungen von Gleichungen der Form

$$q(\mathbf{v}) = c \text{ mit } c \in \mathbb{Q}, \mathbf{v} \in \mathbb{Q}^n$$

sind. Schon für $x^2 + y^2 = 1$ und $x^2 + y^2 = 3$ ist das interessant. Die rationalen Punkte auf dem Einheitskreis, also

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{Q}^2 \mid x^2 + y^2 = 1 \right\}$$

sind schon sehr lange bekannt:

Wenn Sie eine ganzzahlige Lösung der Gleichung $a^2 + b^2 = c^2$ haben (d.h. a, b, c sind die Seitenlängen eines rechtwinkligen

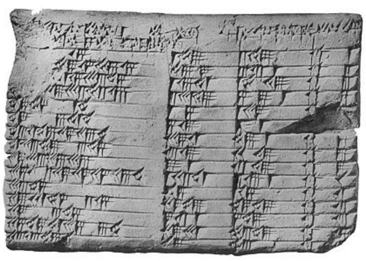


Abbildung 5: Tontafel Plimpton 322 aus der Sammlung der Columbia University, New York (Bild aus Wikipedia)

Dreiecks, dann ist $(\frac{a}{c})^2 + (\frac{b}{c})^2 = 1$ und umgekehrt können sie aus jeder rationalen Lösung $x^2 + y^2 = 1$ durch Multiplikation mit dem Hauptnenner eine ganzzahlige Lösung von $a^2 + b^2 = c^2$ erhalten. Auf einer 3800 Jahre alten Tontafel finden sich so viele ganzzahlige Lösungen dieser Gleichung, dass die Vermutung nahe liegt, dass damals schon das Prinzip wie sich alle Lösungen finden lassen bekannt gewesen sein könnte.

Die Gleichung $x^2 + y^2 = 3$ hat hingegen gar keine rationale Lösung, obwohl die reelle Lösungsmenge genauso ein Kreis ist.

Sie können beide Aussagen (wir können alle rationalen Punkte auf dem Einheitskreis angeben und die zweite Gleichung hat keine rationale Lösung) gern als Knobelaufgabe auffassen. Es gibt jeweils eine kurze Lösung. Sie können natürlich auch nachfragen, wenn Sie wissen möchten, wie das geht.

Es gibt auch eine koordinatenfreie Definition von quadratischen Formen auf einem K-Vektorraum V:

Eine quadratische Form auf einem Vektorraum V ist eine Abbildung $q\colon V \to K$ für die gilt:

1. $q(c\mathbf{v}) = c^2 q(\mathbf{v})$ für alle $\mathbf{v} \in V$ und
2. Die Abbildung $(\mathbf{v}, \mathbf{w}) \mapsto q(\mathbf{v} + \mathbf{w}) - q(\mathbf{v}) - q(\mathbf{w})$ ist eine (sichtbar symmetrische) Bilinearform auf V.

Wenn Sie aufgepasst haben werden Sie sich erinnern dass die 2. Bedingung genau die Formel war, mit der wir $2B(\mathbf{v}, \mathbf{w})$ aus $B(\mathbf{v}, \mathbf{v})$ zurück bekommen können. Die 1. Bedingung wird nur hinzugefügt, um auch für Körper in denen $2 = 0$ ist, einen Begriff zu erhalten, der quadratischen Polynomen entspricht. Das nur, falls Sie dem Begriff noch einmal in einem zahlentheoretischen Zusammenhang begegnen.

Das Minorantenkriterium für positiv definit

Wir hatten gesehen, dass eine Bilinearform, die durch eine symmetrische Matrix A gegeben ist, genau dann positiv definit ist, wenn alle Eigenwerte von A positiv sind. Leider ist dieses Resultat für größere Matrizen unpraktisch, da wir, um die Eigenwerte zu bestimmen zunächst das charakteristische Polynom ausrechnen würden und es dann nicht so leicht ist, zu sehen, was die Nullstellen dieses Polynoms sind. Das Minorantenkriterium ist eine Alternative, die uns diese Rechnung erspart.

Für Extremwertaufgaben sind sowohl nur positive Eigenwerte als auch nur negative Eigenwerte relevant, daher bekommen diese Eigenschaften eigene Namen.

Definition 44. Eine symmetrische Bilinearform $B\colon V \times V \to \mathbb{R}$ auf einem reellen Vektorraum V heißt

1. *positiv definit* wenn $B(\mathbf{v},\mathbf{v}) > 0$ für alle $\mathbf{v} \in V \setminus \{0\}$.
2. *positiv semidefinit* wenn $B(\mathbf{v},\mathbf{v}) \geq 0$ für alle $\mathbf{v} \in V$.
3. *negativ definit* wenn $B(\mathbf{v},\mathbf{v}) < 0$ für alle $\mathbf{v} \in V \setminus \{0\}$.
4. *indefinit* wenn es $\mathbf{v},\mathbf{w} \in V$ gibt, mit $B(\mathbf{v},\mathbf{v}) > 0$ und $B(\mathbf{w},\mathbf{w}) < 0$.

Beispiel 45. Auf $\mathbb{R}^2$ sind die Formen

1. $B_p(\mathbf{v},\mathbf{w}) = \mathbf{v}^t \circ \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \circ \mathbf{w}$ positiv definit,
2. $B_s(\mathbf{v},\mathbf{w}) = \mathbf{v}^t \circ \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \circ \mathbf{w}$ positiv semidefinit,
3. $B_n(\mathbf{v},\mathbf{w}) = \mathbf{v}^t \circ \begin{pmatrix} -3 & 0 \\ 0 & -5 \end{pmatrix} \circ \mathbf{w}$ negativ definit und
4. $B_{in}(\mathbf{v},\mathbf{w}) = \mathbf{v}^t \circ \begin{pmatrix} -3 & 0 \\ 0 & 5 \end{pmatrix} \circ \mathbf{w}$ indefinit.

Eine notwendige Bedingung dafür, dass eine symmetrische Matrix $A \in \mathrm{Mat}_{n,n}(\mathbb{R})$ nur positive Eigenwerte hat, ist dass $\det(A) > 0$ gilt, denn die Determinante ist das Produkt der Eigenwerte.

Das allein ist kein Kriterium, da $\det(\begin{pmatrix} -3 & 0 \\ 0 & -5 \end{pmatrix}) = 15 > 0$ ist, obwohl die entsprechende Bilinearform negativ definit ist.
Für das Kriterium werden wir zusätzlich die Determinanten von Untermatrizen benötigen.

Notation: Für eine $n \times n$-Matrix $A = (a_{ij})_{ij,=1,\ldots n}$ bezeichnen wir für $k = 1,\ldots,n$ mit

$$A_{\mathbf{k}} := (a_{ij})_{ij,=1,\ldots\mathbf{k}}$$

den k-ten Hauptminor von A.

Zum Beispiel sind die Hauptminoren der Matrix $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ also

$$A_1 = (2),\ A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \text{ und } A_3 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Damit können wir jetzt das Resultat formulieren.

Satz 46. *Für eine reelle symmetrische $n \times n$ Matrix A sind die folgenden Aussagen äquivalent:*

1. *Alle Eigenwerte von A sind positiv.*
2. *Die durch A definierte Bilinearform $B_A\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ ist positiv definit.*
3. *Die Determinanten der Hauptminoren sind positiv, d.h. $\det(A_k) > 0$ für alle $k = 1,\ldots,n$.*

Bevor wir das Resultat nachweisen, lassen Sie uns an einem Beispiel testen, dass das Kriterium viel einfacher zu testen ist, als zu versuchen die Eigenwerte von A zu bestimmen.

Beispiel 47. Für die Matrix $A_4 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ finden wir

$$\det(A_1) = \det(2) = 2 > 0,$$

$$\det(A_2) = \det(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}) = 4 - 1 = 3 > 0,$$

$$\begin{aligned} \det(A_3) &= \det(\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}) && \text{Entwicklung nach 1. Spalte} \\ &= 2\det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - 1 \cdot \det\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} \\ &= 2 \cdot 3 - 2 = 4 > 0. \end{aligned}$$

Die Matrix ist also positiv definit.

Genauso finden wir für

$$\begin{aligned} \det(\begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}) &= 2 \cdot \det(\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}) - 1 \cdot \det(\begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix})) \\ &= 2\det(A_3) - 1\det(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}) = 2 \cdot 4 - 3 = 5. \end{aligned}$$

Also ist auch diese 4×4 Matrix positiv definit. Diese Rechnung zeigt auch dass allgemein für die $n \times n$-Matrix

$$A_n = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

gilt dass

$$\det(A_n) = 2\det(A_{n-1}) - \det(A_{n-2}).$$

Und da wir für die ersten Determinanten die Ergebnisse $2, 3, 4, 5$ gefunden haben erwarten wir vielleicht, dass $\det(A_n) = n + 1$ gilt. Das können wir leicht mit Induktion beweisen, denn die Aussage stimmt für $n = 1, 2$. Das gibt den Induktionsanfang. Gilt die Aussage für $k - 1, k$ so auch für $k + 1$, denn dann gilt

$$\begin{aligned} \det(A_{k+1}) &= 2 \cdot \det(A_k) - \det(A_{k-1}) && \text{Formel von oben} \\ &= 2(k+1) - k && \text{Induktions Annahme} \\ &= k + 2 = (k+1) + 1. \end{aligned}$$

Also sind die Matrizen A_n für alle $n \geq 1$ positiv definit.

Beweis des Minorantenkriteriums. Die Äquivalenz der Aussagen 1. und 2. haben wir schon gesehen:

1. $\Rightarrow$ 2. Nach dem Spektralsatz gibt es eine Othrnormalbasis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von A, die aus Eigenvektoren von A besteht, wir bezeichen die zugehörigen Eigenwerte mit c_i d.h. $A\mathbf{v}_i = c_i\mathbf{v}_i$. Hat A positive Eigenwerte, d.h. $c_1, \ldots, c_n \in \mathbb{R}_{>0}$ so ist B_A positiv definit, denn jeder Vektor $\mathbf{v} \in V \setminus \{0\}$ lässt sich als Linearkombination $\mathbf{v} = \sum_{i=1}^n a_i\mathbf{v}_i$ schreiben wobei nicht alle $a_i = 0$ sind.

   Dann ist aber

$$\begin{aligned}
B_A(\mathbf{v}, \mathbf{v}) &= \mathbf{v}^t A\mathbf{v} && \text{Def. } B_A(,) \\
&= \langle \mathbf{v}, A\mathbf{v} \rangle && \text{Def. Standardskalarprodukt} \\
&= \langle \sum_{i=1}^n a_i\mathbf{v}_i, A\sum_{j=1}^n a_j\mathbf{v}_j \rangle && \langle,\rangle \text{bilinear, } A \text{ linear} \\
&= \sum_{i=1}^n a_i \langle \mathbf{v}_i, \sum_{j=1}^n a_j A\mathbf{v}_j \rangle && A\mathbf{v}_j = c_j\mathbf{v}_j \\
&= \sum_{i=1}^n a_i \sum_{j=1}^n a_j c_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle && \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \text{ für } i \neq j, = 1 \text{für } i = j \\
&= \sum_{i=1}^n a_i^2 c_i > 0 && \text{da } c_j > 0.
\end{aligned}$$

   Das war zu zeigen.

2.$\Rightarrow$1. Ist B_A positive definit und $\mathbf{v} \in V \setminus \{0\}$ ein Eigenvektor von A, d.h. $A\mathbf{v} = c\mathbf{v}$, so gilt

$$0 < B_A(\mathbf{v}, \mathbf{v}) = \langle \mathbf{v}, A\mathbf{v} \rangle = \langle \mathbf{v}, c\mathbf{v} \rangle = c \cdot \underbrace{\langle \mathbf{v}, \mathbf{v} \rangle}_{>0}.$$

   Also muss $c > 0$ sein, d.h. alle Eigenwerte von A sind positiv und das war zu zeigen.

2.$\Rightarrow$3. Ist $B_A(,)$ positiv definit, so wissen wir schon, dass alle Eigenwerte von A positiv sind und darum $\det(A) > 0$ gilt. Ist B_A positiv definit, so ist auch die Einschränkung von B_A auf den

   Unterraum $\mathbb{R}^k = \mathrm{Span}(\mathbf{e}_1, \ldots, \mathbf{e}_k) = \{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \end{pmatrix} \mid x_i \in \mathbb{R}\} \subseteq \mathbb{R}^n$

   positiv definit.

   Diese Bilinearform auf $\mathbb{R}^k$ wird aber durch den k-ten Hauptminor A_k von A beschrieben. Also gilt auch $\det(A_k) > 0$ für alle $k = 1, \ldots, n$. Das war zu zeigen.

3$\Rightarrow$2. Das ist der Knackpunkt des Satzes: Wir haben gerade gesehen, dass die Hauptminoren, Bilinearformen auf $\mathbb{R}^k$ mit $k \leq n$ beschreiben. Daher zeigen wir die Aussage mit Induktion.

Für $n = 1$ ist die Aussage klar, denn $A = (a_{11})$ ist dann nur eine Zahl, d.h. $\det(A) = a_{11}$ und ist diese Zahl > 0, so ist die Bilinearform B_A auf $\mathbb{R}^1$ positiv definit.

Angenommen die Aussage stimmt für $m \times m$-Matrizen und A ist eine $m+1 \times m+1$ Matrix für die die Determinanten der Hauptminoren alle positiv sind. Nach Induktionsannahme sind dann alle Eigenwerte der Matrix A_m positiv, d.h. es gibt nach dem Spektralsatz eine Orthonormalbasis $\mathbf{v}_1, \dots, \mathbf{v}_m$ von

$$\mathbb{R}^m = \mathrm{Span}(\mathbf{e}_1, \dots, \mathbf{e}_m) = \{ \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ 0 \\ \vdots \end{pmatrix} \mid x_i \in \mathbb{R}\} \subseteq \mathbb{R}^{m+1}$$

die aus Eigenvektoren von A besteht und die zugehörigen Eigenwerte sind positiv.

Ergänzen wir diese Basis durch den letzten Standardbasisvektor $e_{m+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ zu einer Basis von $\mathbb{R}^{m+1}$ so ist diese leider nicht immer bezüglich B_A orthogonal, aber wir können das reparieren, indem wir

$$\mathbf{v}_{m+1} := \mathbf{e}_{m+1} - \sum_{i=1}^{m} \frac{B_A(\mathbf{e}_{m+1}, \mathbf{v}_i)}{B_A(\mathbf{v}_i, \mathbf{v}_i)} \mathbf{v}_i$$

definieren. Diese Formel ist sinnvoll, da nach Voraussetzung $B_A(\mathbf{v}_i, \mathbf{v}_i) > 0$ für alle i.

Setzten wir $P = (\mathbf{v}_1 \dots \mathbf{v}_{m+1})$ so ist die Matrix der Bilinearform B_A bezüglich der Basis $\mathbf{v}_1, \dots, \mathbf{v}_{m+1}$ nach der Basiswechselformel für Bilinearformen durch $P^t AP$ gegeben und nach Konstruktion ist das eine Diagonalmatrix der Form:

Warnung: Die c_i's haben nichts mit den Eigenwerten von A zu tun, z.B. ist 2 sicher kein Eigenwert von $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, da $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ invertierbar ist.

$$\mathrm{Mat}_{\underline{\mathbf{v}}}(B_A) = \begin{pmatrix} c_1 & & & \\ & \ddots & & \\ & & c_m & \\ & & & B_A(\mathbf{v}_{m+1}, \mathbf{v}_{m+1}) \end{pmatrix}.$$

Wir wissen

$$\begin{aligned} \det(P^t AP) &= \det(P^t)\det(A)\det(P) && \text{Multiplikationssatz für det} \\ &= \det(P)^2 \det(A). && \text{denn } \det(P^t) = \det(P) \end{aligned}$$

) Nach Voraussetzung ist $\det(A) > 0$, also ist auch $\det(P)^2\det(A) = \det(P^t AP) = (\prod_{i=1}^m c_i) B_A(\mathbf{v}_{m+1}, \mathbf{v}_{m+1}) > 0$. Da die $c_i > 0$ sind ist darum $B_A(\mathbf{v}_{m+1}, \mathbf{v}_{m+1}) > 0$ und also ist B_A positiv definit.

□

Bemerkung. Sie hatten in der Vorlesung richtig angemerkt, dass es für den Satz nicht wichtig ist, die Hauptminoren von links oben nach rechts unten zu wählen, Sie könnten genauso gut mit irgend einem Diagonaleintrag anfangen und diesen nach einander zu größeren quadratischen Matrizen erweitern.

Bemerkung. Das Argument für den Satz ist bemerkenswert, denn wir zeigen eine Aussage über die Eigenwerte einer Matrix, indem wir die Matrix als Bilinearform auffassen. Das obwohl Eigenwerte für Bilinearformen gar keinen Sinn machen.

Umgekehrt haben wir für den Beweis der Aussage, dass es für jede quadratische Form auf $\mathbb{R}^n$ eine Orthonormalbasis gibt, so dass die Formel bezüglich dieser Koordinaten keine gemischten Terme enthält die zugehörige Matrix als lineare Abbildung aufgefasst und darauf den Spektralsatz angewendet.

Der Wechsel zwischen verschiedenen Interpretationen von Matrizen – als lineare Abbildungen oder Bilinearform – erlaubt es also Ergebnisse die in einer Interpretation einfach sind zu überraschenden Ergebnissen in der anderen Interpretation zu übertragen.

Anwendung 3: Die Singulärwertzerlegung

Als ersten Anwendungen des Spektralsatzes hatten wir mehrdimensionale Extremwerte und die Diagonalisierung quadratischer Gleichungen in mehreren Variablen gesehen. Eine Kombination dieser Anwendungen war die Maximierung von $\|A\mathbf{v}\|$ für eine symmetrische Matrix A auf der Menge der Vektoren $\mathbf{v}$ der Länge 1. Das wurde genau durch die Eigenvektoren zu den Eigenwerten mit größtem Absolutbetrag maximiert.

Als dritte und vielleicht aktuell wichtigste Anwendung des Spektralsatzes geht es nun um die Singulärwertzerlegung von reellen, nicht notwendig quadratischen Matrizen. Diese Zerlegung liefert eine Methode, einen besonders wesentlichen Anteil einer $N \times M$ Matrix zu finden. Das wird in manchen Zusammenhängen auch Hauptkomponentenanalyse genannt.

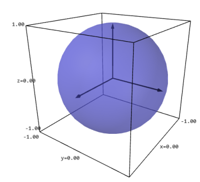
Die Großbuchstaben in $N \times M$ sollten Sie darauf hinweisen, dass große Zahlen vorkommen können.

DIE IDEE ist an einem Beispiel leicht zu motivieren. Wenn wir für eine lineare Abbildung

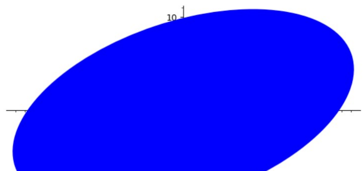
$$A\colon \mathbb{R}^3 \to \mathbb{R}^2$$

(d.h. A ist durch eine 2×3-Matrix beschrieben) geometrisch verstehen möchten, was die Abbildung tut, so können wir zunächst das Bild der Einheitskugel im $\mathbb{R}^3$ unter der Abbildung A anschauen, d.h. wir können die Bilder $A\mathbf{v}$ für Vektoren $\mathbf{v}$ mit $\|v\| = 1$ betrachten. Dabei können wir beobachten, dass das Bild immer eine Ellipse in $\mathbb{R}^2$ ist.

Die Einheitskugel mit Standardbasis:



und das Bild unter $A\colon \mathbb{R}^3 \to \mathbb{R}^2$



FRAGE:Wie finden wir die Hauptachsen dieser Ellipse? Was sind insbesondere die Koordinaten der längsten Achse der Ellipse?

Lassen Sie uns das in Formeln aufschreiben:

$$\text{Suche } \mathbf{v} \in \mathbb{R}^3 \text{ mit } \|v\| = 1, \text{ so dass } \|A\mathbf{v}\| \text{ maximal ist.}$$

Das sieht aus wie das Maximierungsproblem für symmetrische Matrizen, nur ist A leider noch nicht einmal quadratisch, d.h. das Konzept von Eigenvektoren ist für A sinnlos.

Wenn wir nicht weiter wissen, schreiben wir als erstes Definitionen auf: Der Ausdruck $\|A\mathbf{v}\|$ war definiert als

$$\begin{aligned} \|A\mathbf{v}\|^2 &= \langle A\mathbf{v}, A\mathbf{v}\rangle && \text{Definition } \|\,\| \\ &= (A\mathbf{v})^t \circ (A\mathbf{v}) && \text{Def. Standardskalarprodukt} \\ &= \mathbf{v}^t A^t A\mathbf{v} && (B \circ C)^t = C^t \circ B^t \\ &= B_{A^tA}(\mathbf{v}, \mathbf{v}) && \text{Bilinearform zu } A^tA. \end{aligned}$$

Hier ist die Matrix A^tA eine $m \times m$-Matrix ($m \times n$ mal $n \times m$-Matrix), also quadratisch und sogar symmetrisch, denn

$$(A^tA)^t = A^t(A^t)^t = A^tA.$$

Damit haben wir das Problem nun doch auf ein Optimierungsproblem für quadratische Formen zurückgeführt. Dafür haben wir schon gezeigt, dass die normierten Eigenvektoren $\mathbf{v}$ zum größten Eigenwert von A^tA die Vektoren der Länge 1 sind für die $\|A^tA\mathbf{v}\|$ maximal ist.

Beispiel 48. Lassen Sie uns das einmal für die explizite Matrix

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^2$$

ausrechnen.

Zunächst müssen wir A^tA berechnen:

Rechnen Sie die behaupteten Formeln selbst nach und lassen Sie mich wissen, wenn Sie Tippfehler finden!

$$A^tA = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \circ \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

Diese Matrix macht mir keine Lust auf die Berechnung des charakteristischen Polynoms. Darum denke ich einen Moment nach, ob ich einen Eigenwert kenne. Mir fällt auf, dass $A \colon \mathbb{R}^3 \to \mathbb{R}^2$ wegen der Dimensionsformel einen nichttrivialen Kern hat und darum hat A^tA den Eigenwert 0. Den Kern von A bestimme ich mit dem Gauß-Algorithmus und finde

Prüfen Sie das!

$$\mathrm{Ker}(A) = \mathrm{Span}(\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}).$$

Da $\|\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}\| = 3$ gilt, ist $\mathbf{v}_0 := \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$ ein normierter Basisvektor, also insbesondere ein Eigenvektor von A^tA zum Eigenwert 0.

Symmetrische Matrizen bilden das orthogonale Komplement eines Eigenvektors in sich ab. Darum berechnen wir

$$\mathrm{Span}(\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix})^{\perp} = \mathrm{Span}(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}),$$

denn die beiden Vektoren sind linear unabhängig und orthogonal zu $\mathbf{v}_0$. Die Matrix von A auf diesem Unterraum können wir bezüglich dieser Basis ausrechnen:

$$A^tA \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 180 \\ 270 \\ 180 \end{pmatrix} = 90 \cdot \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix} = 90 \cdot (2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix})$$
$$A^tA \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 180 \\ 450 \\ 540 \end{pmatrix} = 90 \cdot \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix} = 90 \cdot (2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}).$$

Damit finden wir die Matrix $90 \cdot \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$ von der ich den Eigenvektor $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ zum Eigenwert $90 \cdot 4$ sehe (denn $\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$), da die Spur einer Matrix die Summe der Eigenwerte ist, sehe ich damit dass der zweite Eigenwert 1 ist, und dafür kann ich den Eigenvektor $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ sehen (da $\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} - 1 \cdot E_2 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$) und $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ im Kern hiervon liegt).

Also hat A^tA die Eigenvektoren $\tilde{\mathbf{v}}_{360} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ zum Eigenwert 360 (dieser entspricht dem Vektor $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$) und $\tilde{\mathbf{v}}_{90} = -2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix}$ zum Eigenwert 90 (dieser entspricht dem Vektor $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$).

Damit sind die Vektoren

$$\mathbf{v}_{360} := \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \mathbf{v}_{90} := \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}, \mathbf{v}_0 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

eine Orthonormalbasis aus Eigenvektoren von A^tA. Damit ist $P = (\mathbf{v}_{360}, \mathbf{v}_{90}, \mathbf{v}_0) \in O(3)$, erfüllt also $P^{-1} = P^t$ und es gilt

$$P^t(A^tA)P = \begin{pmatrix} 360 & 0 & 0 \\ 0 & 90 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Diese Formel können wir auch als

$$P^t(A^tA)P = (AP)^t(AP)$$
$$= (A\mathbf{v}_{360}A\mathbf{v}_{90}A\mathbf{v}_0)^t(A\mathbf{v}_{360}A\mathbf{v}_{90}A\mathbf{v}_0) = \begin{pmatrix} 360 & 0 & 0 \\ 0 & 90 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

schreiben. Insbesondere gilt also $0 = A\mathbf{v}_{360}^t A\mathbf{v}_{90} = \langle A\mathbf{v}_{360}, A\mathbf{v}_{90}\rangle = 0$, die Vektoren $A\mathbf{v}_{360}, A\mathbf{v}_{90}$ sind also eine orthogonale Basis von $\mathrm{Bild}(A)$. Wir wissen außerdem, dass

$$\begin{aligned} \|A\mathbf{v}_{360}\| &= \sqrt{\langle A\mathbf{v}_{360}, A\mathbf{v}_{360}\rangle} = \sqrt{\langle \mathbf{v}_{360}, A^tA\mathbf{v}_{360}\rangle} \\ &= \sqrt{360\langle \mathbf{v}_{360}, \mathbf{v}_{360}\rangle} = \sqrt{360} \end{aligned}$$

und analog

$$\|A\mathbf{v}_{90}\| = \sqrt{90\langle \mathbf{v}_{90}, \mathbf{v}_{90}\rangle} = \sqrt{90}.$$

Damit sind also

$$\mathbf{w}_{360} = \frac{1}{\sqrt{360}}A\mathbf{v}_{360} = \frac{1}{6\sqrt{10}}\begin{pmatrix} 18 \\ 6 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix} \text{ und}$$

$$\mathbf{w}_{90} = \frac{1}{\sqrt{90}}A\mathbf{v}_{90} = \frac{1}{3\sqrt{10}}\begin{pmatrix} 3 \\ -9 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{pmatrix}$$

eine Orthonormalbasis von $\mathrm{Bild}(A) = \mathbb{R}^2$ und es gilt für $Q = (\mathbf{w}_{360}\mathbf{w}_{90}) \in O(2)$, dass

$$Q^tAP = \begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}, \text{ bzw.}$$

$$A = Q\begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}P^t.$$

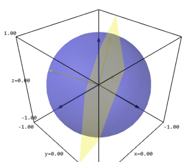
Diese Zerlegung heißt Singulärwertzerlegung der Matrix A, die Diagonaleinträge $\sqrt{360}, \sqrt{90}$ heißen Singulärwerte von A.

Die Darstellung $Q^tAP = \begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}$ erklärt, wie wir die Abbildung A geometrisch verstehen können, denn

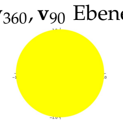
$$\begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix} = \begin{pmatrix} \sqrt{360} & 0 \\ 0 & \sqrt{90} \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

ist die Komposition der Projektion auf die ersten beiden Koordinaten (d.h. die $\mathbf{v}_{360}, \mathbf{v}_{90}$-Koordinaten)), diese bildet die Einheitskugel auf die Einheitskreisscheibe ab und der Streckung durch die Singulärwerte, diese bildet die Einheitskreisscheibe auf eine Ellipse ab.

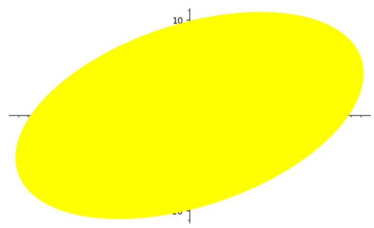
Die Einheitskugel mit $\mathbf{v}_{360}, \mathbf{v}_{90}$ Ebene:



und das Bild der Projektion auf die $\mathbf{v}_{360}, \mathbf{v}_{90}$ Ebene ist ein Kreis:



und dieser wird auf eine Ellipse gestreckt:



Bevor wir ausformulieren, wie sich das Argument für unser Beispiel genauso für eine allgemeine $n \times m$-Matrix übertragen lässt, möchte ich kurz festhalten, wieso es kein Zufall ist, dass die Eigenwerte von A^tA größer oder gleich 0 waren.

Bemerkung. Für jede reelle $n \times m$-Matrix A ist die durch A^tA definierte Bilinearform B_{A^tA} positiv semidefinit, d.h. die Eigenwerte der Matrix sind A^tA sind alle ≥ 0.

Beweis. Ist $\mathbf{v}$ ein Eigenvektor zum Eigenwert c, so ist einerseits

$$\langle \mathbf{v}, A^tA\mathbf{v} \rangle = \langle A\mathbf{v}, A\mathbf{v} \rangle \geq 0$$

und andererseits

$$\langle \mathbf{v}, A^tA\mathbf{v} \rangle = \langle \mathbf{v}, c\mathbf{v} \rangle = c\,\underbrace{\langle \mathbf{v}, \mathbf{v} \rangle}_{>0}.$$

Darum muss dann $c \geq 0$ gelten. $\square$

Lassen Sie uns nun die Singulärwertzerlegung für allgemeine Matrizen formulieren.

Satz 49 (Singulärwertzerlegung)**.** *Sei $A \in \mathrm{Mat}_{n.m}(\mathbb{R})$ eine reelle Matrix und $r := \mathrm{Rang}(A)$. Dann existieren orthogonale Matrizen $P = (\mathbf{v}_1 \dots \mathbf{v}_m) \in O(m), Q = (\mathbf{w}_1 \dots \mathbf{w}_n) \in O(n)$ und reelle Zahlen $s_1 \geq s_2 \geq \dots \geq s_r > 0$, so dass*

$$A = Q \begin{pmatrix} s_1 & 0 & \dots & \dots & 0 \\ 0 & s_2 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & s_r & 0 \\ 0 & \dots & \dots & 0 & 0 \end{pmatrix} P^t.$$

Die Zahlen $s_1, \dots, s_r$ sind hierbei eindeutig bestimmt und heißen Singulärwerte *der Matrix A.*

Beweis. Der Beweis folgt einfach unserer Rechnung für die Beispielmatrix. Sei A eine $n \times m$ Matrix.

Schritt 1 Die Matrix A^tA ist symmetrisch. Nach dem Spektralsatz für selbstadjungierte Abbildungen, existiert also eine Orthonormalbasis von $\mathbb{R}^n$, die aus Eigenvektoren von A^tA besteht.

Wir haben außerdem oben gezeigt, dass die Eigenwerte $c_1, \dots, c_m$ von A^tA alle ≥ 0 sind, also können wir diese so sortieren, dass $c_1 \geq c_2 \geq \dots \geq c_{r'} > 0 = c_{r'+1} = \dots = c_m$ gilt.

Sei $\mathbf{v}_1, \dots, \mathbf{v}_m$ eine zugehörige Orthonormalbasis aus Eigenvektoren, d.h. $A^tA\mathbf{v}_i = c_i\mathbf{v}_i$ für alle $i = 1, \dots m$. Dann ist die Matrix $P := (\mathbf{v}_1 \dots \mathbf{v}_m)$ orthogonal und es gilt

$$\underline{\mathbf{v}}\mathrm{Mat}_{\underline{\mathbf{v}}}(A^tA) = P^tA^tAP = \begin{pmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & \dots & c_m \end{pmatrix}.$$

Da

$$P^tA^tAP = (AP)^tAP = (Av_1 \dots A\mathbf{v}_m)^t \circ (A\mathbf{v}_1 \dots A\mathbf{v}_m)$$

ist, bedeutet die obige Formel

$$\begin{aligned}
\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle &= (A\mathbf{v}_i)^t A\mathbf{v}_j = 0 \text{für } i \neq j\\
\langle A\mathbf{v}_i, A\mathbf{v}_i\rangle &= c_i > 0 \text{für } i = 1,\dots,r'\\
\langle A\mathbf{v}_i, A\mathbf{v}_i\rangle &= 0 \text{für } i > r'.
\end{aligned}$$

Also sind die Vektoren $A\mathbf{v}_1,\dots,A\mathbf{v}_{r'}$ eine orthogonale Basis von $\mathrm{Bild}(A)$, insbesondere ist $r' = \mathrm{Rang}(A) = r$ und $\mathbf{v}_{r+1},\dots,\mathbf{v}_m$ sind eine Basis von $\mathrm{Ker}(A)$.

Außerdem ist

$$\begin{aligned}
\|A\mathbf{v}_i\| &= \sqrt{\langle A\mathbf{v}_i, A\mathbf{v}_i\rangle} = \sqrt{\langle \mathbf{v}_i, A^t A\mathbf{v}_i\rangle} && A^t \text{ adjungierte Abbildung}\\
&= \sqrt{c_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle} && \mathbf{v}_i \text{ Eigenvektor}\\
&= \sqrt{c_i} =: s_i && \mathbf{v}_i\text{normiert.}
\end{aligned}$$

Setzen wir also $w_i := \frac{1}{\sqrt{c_i}} A\mathbf{v}_i$, so sind die Vektoren $\mathbf{w}_1,\dots,\mathbf{w}_r$ also eine Orthonormalbasis von $\mathrm{Bild}(A)$ und es gilt $A\mathbf{v}_i = \sqrt{c_i}\mathbf{w}_i = s_i\mathbf{w}_i$ für $i = 1\dots r$.

Ergänzen wir $\mathbf{w}_1,\dots,\mathbf{w}_r$ zu einer Orthonormalbasis $\mathbf{w}_1,\dots,\mathbf{w}_n$ von $\mathbb{R}^n$, so ist die Matrix $Q = (\mathbf{w}_1 \dots \mathbf{w}_n)$ orthogonal und es gilt

$$Q^t AP = {}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{w}}}(A) = \begin{pmatrix} s_1 & 0 & \dots & \dots & 0\\ 0 & s_2 & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & \vdots\\ \vdots & & & s_r & 0\\ 0 & \dots & \dots & 0 & 0 \end{pmatrix}.$$

Multiplizieren wir diese Gleichung von links mit Q und von rechts mit P^t erhalten wir wie gewünscht

$$A = Q\begin{pmatrix} s_1 & 0 & \dots & \dots & 0\\ 0 & s_2 & \ddots & & \vdots\\ \vdots & & \ddots & \ddots & \vdots\\ \vdots & & & s_r & 0\\ 0 & \dots & \dots & 0 & 0 \end{pmatrix}P^t.$$

Das zeigt die Existenz einer Singulärwertzerlegung.

Die Singulärwerte sind eindeutig, denn ist $A = Q'TP'^t$ eine andere Zerlegung, wobei $T = \begin{pmatrix} t_1 & 0 & \dots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & & t_\ell & 0\\ 0 & \dots & 0 & 0\end{pmatrix}$ wie im Satz,

Vergleichen Sie den Beweis noch einmal mit unserer Rechnung für das Beispiel $A = \begin{pmatrix} 4 & 11 & 14\\ 8 & 7 & -2\end{pmatrix}$. Stimmen Sie mir zu, dass der allgemeine Beweis transparenter ist?

Das ist eine weitere Funktion von Beweisen, nämlich uns klar zu machen, was die Kernargumente sind, die im Rechensalat von Beispielen versteckt waren.

so ist

$$\begin{aligned}
A^tA &= (Q'TP'^t)^tQ'TP'^t \\
&= P'T^tQ'^tQ'TP'^t \\
&= P'T^tTP'^t \\
&= P'\begin{pmatrix} t_1^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & t_\ell^2 & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix} P'^t.
\end{aligned}$$

Das zeigt aber, dass $t_1^2, \dots, t_\ell^2, 0$ die Eigenwerte von A^tA sein müssen, d.h. die Singulärwerte sind in jeder Darstellung die positiven Wurzeln der Eigenwerte von A^tA.

□

Bemerkung. In der Literatur ist für die $n \times m$-Matrix in der Singulärwertzerlegung, in der die einzigen von 0 verschiedenen Einträge die Diagonaleinträge $s_1, \dots, s_r$ sind die Bezeichnung Σ üblich. Außerdem werden die orthogonalen Matrizen Q, P mit den Buchstaben U, V bezeichnet, d.h. Sie finden die Formel $A = U\Sigma V^t$. In dieser Vorlesung bleiben die Buchstaben V, U für Vektorräume und Untervektorräume reserviert.

Bemerkung. Die Singulärwertzerlegung hat viele Anwendungen.

1. Die Darstellung $A = Q^t \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & s_r & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix} P^t$ erklärt, dass wir die Abbildung A in P Koordinaten als Hintereinanderausführung der Projektion auf $\mathrm{Ker}(A)^\perp = \mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_r) \subset \mathbb{R}^m$ (das entspricht der Matrix) $\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & 1 & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix}$ und einer Streckung um die Singulärwerte auffassen können.

2. In der Praxis ist die Singulärwertzerlegung nützlich um Daten die in Form von $N \times M$-Matrizen A gegeben sind (z.B. digitale Bilder, oder Tabellen) zu komprimieren oder zu analysieren. Ist $A = QSP^t$ die Singulärwertzerlegung und sind die ersten k Singulärwerte $s_1, \dots, s_k$ deutlich größer als $s_{k+1}, \dots, s_r$, so ist die $N \times M$ Matrix $(\mathbf{w}_1 \dots \mathbf{w}_k) \begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_k \end{pmatrix} (\mathbf{v}_1 \dots \mathbf{v}_k)^t$ eine gute Approximation von A.

Eine berühmte Anwendung hiervon war der Netflix-Wettbewerb für einen besseren Algorithmus für Filmvorschläge. Alle Preisträger haben die Singulärwertzerlegung verwendet.

Nützliche Vektorraumkonstruktionen

Ein roter Faden durch die Vorlesung war, kompliziert aussehende Probleme dadurch übersichtlich in den Griff bekommen, dass wir nachdem wir ein paar einfache Beispiele angeschaut hatten, eine kompakte abstrakte Formulierung gefunden haben, die uns eine einfache Argumentation ermöglicht hat. In der Linearen Algebra 1 hatten wir das zunächst mit den Körperaxiomen gesehen, die einerseits für die lange Liste von Rechenregeln für die 4 Grundrechenarten auf eine sehr kurze Liste von einfachen Regeln gefunden hat, aus denen sich alle anderen ergeben und damit andererseits zeigt, dass die Rechenregeln auch in anderen Zahlbereichen wie $\mathbb{C}, \mathbb{Z}/p\mathbb{Z}$ genauso gelten. Genauso war der Begriff des Vektorraums dazu da, allgemeine Aussagen so formulieren zu können, dass wir Induktionsargumente über die Dimension eines Vektorraums anwenden können, da in diesen Argumenten Unterräume auftreten, für die wir keine „Standardbasis" haben. Das Grundprinzip haben wir immer wieder verwendet, zuletzt im Argument für die Singulärwertzerlegung.

In diesem Kapitel möchte ich nun einerseits noch einmal einen Rückblick auf einige Konstruktionen werfen, die wir von einem abstrakten Standpunkt noch einmal neu betrachten können, andererseits möchte ich das auch noch einmal als Anlass nehmen, um logische Argumentation und den Umgang mit unseren sehr kompakten, abstrakten Darstellungsformen zu üben.

Der Dualraum eines Vektorraums

Für das Standardskalarprodukt auf $\mathbb{R}^n$ war die Formel $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t \mathbf{w}$ sehr nützlich, die aus einem Spaltenvektor $\mathbf{v} \in \mathbb{R}^n$ einen Zeilenvektor $\mathbf{v}^t \in \mathrm{Mat}_{1,n}(\mathbb{R}) = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$ macht, diese als lineare Abbildung $\mathbf{v}^t \colon \mathbb{R}^n \to \mathbb{R}$ auffasst und auf $\mathbf{w}$ anwendet.

Ist $v \in V$ ein Element eines abstrakten K-Vektorraums V, dann ist der Ausdruck $\mathbf{v}^t$ zunächst sinnlos, aber statt $1 \times n$-Matrizen können wir immer den Vektorraum

$$V^\vee := \mathrm{Hom}(V, K) = \{ f \colon V \to K \mid f \text{ linear } \}$$

verwenden.

Definition. Der *Dualraum* eines K Vektorraums V ist der Raum der linearen Abbildungen von V nach K:

$$V^\vee := \mathrm{Hom}(V,K) = \{f\colon V \to K \mid f \text{ linear }\}.$$

Die Notation „$V^\vee$" lesen wir als „V *dual*".

Die Notation V^t würde besser daran erinnern, dass wir statt Vektoren nun Abbildungen nach K betrachten, aber $V^\vee$ hat sich durchgesetzt, in manchen Büchern finden Sie auch V^*.

Wir werden sehen, dass es 2 unterschiedliche Zusatzdaten auf V (entweder eine Basis oder eine Bilinearform) gibt, die es ermöglichen aus Elementen von V, Elemente von $V^\vee$ zu machen.

Wie bei jedem neuen Objekt, sollten wir uns zunächst einmal mit $V^\vee$ vertraut machen. Im Beispiel $V = K^n$ wissen wir schon, dass $\mathrm{Hom}(K^n, K) = \mathrm{Mat}_{1,n}(K)$ durch $1 \times n$-Matrizen beschrieben werden kann, insbesondere wieder n-dimensional ist.

Behauptung 50. *Ist V ein endlichdimensionaler K-Vektorraum, so gilt* $\dim V = \dim V^\vee$.

Ohne die Voraussetzung, dass V endlichdimensional ist, wäre das nicht richtig.

In der Vorlesung haben wir das Argument für den Beweis mit einer Reihe von Vorüberlegungen erarbeitet, im Skript füge ich diese im Anschluss an den Beweis als „Wie komme ich darauf?" ein.

Beweis. Da die Dimension eines Vektorraums als die Anzahl der Elemente einer Basis definiert ist, müssen wir eine Basis von $V^\vee$ konstruieren.

Schritt 1: Konstruiere Elemente $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n \in V^\vee$: Jede lineare Abbildung $A\colon V \to W$ ist eindeutig durch die Werte auf einer Basis bestimmt, insbesondere gilt das für lineare Abbildungen $f\colon V \to K$.

Sei also $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ eine Basis von V, dann definieren wir

$$\check{\mathbf{v}}_i\colon V \to K$$
$$\mathbf{w} \mapsto \check{\mathbf{v}}_i(w) := a_i \text{ wobei } \mathbf{w} = \sum_{j=1}^n a_j \mathbf{v}_j$$

die eindeutige Schreibweise von $\mathbf{w}$ als Linearkombination der $\mathbf{v}_i$ ist.

Wir haben bereits gezeigt, dass die Abbildung $\check{\mathbf{v}}_i$ linear ist, denn die Koordinatenabbildung $\mathrm{Koord}_{\mathbf{v}}\colon V \to K^n$ ist linear und $\check{\mathbf{v}}_i$ ist die Komposition der Koordinatenabbildung mit der Projektion auf die i-te Koordinate $p_i\colon K^n \to K; p_i(\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}) := a_i$. Da p_i ebenfalls linear ist, ist auch $\check{\mathbf{v}}_i = p_i \circ \mathrm{Koord}_{\mathbf{v}}$ linear.

Schritt 2: Die Abbildungen

$$\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n \in V^\vee = \{f\colon V \to K \mid f \text{ linear}\}$$

sind linear unabhängig, denn sind $b_1, \dots, b_n \in K$ so dass die Linearkombination

$$b_1\check{\mathbf{v}}_1 + \cdots + b_n\check{\mathbf{v}}_n = 0$$

die Nullabbildung $0 \in V^\vee = \{f\colon V \to K \mid f \text{ linear}\}$ ist, so gilt für alle $i = 1, \ldots n$, dass

$$\begin{aligned} 0 = 0(\mathbf{v}_i) &= (b_1\check{\mathbf{v}}_1 + \cdots + b_n\check{\mathbf{v}}_n)(\mathbf{v}_i) && \text{Vor. Lin.komb. ist 0-Abb} \\ &= \sum_{j=1}^{n} b_j\check{\mathbf{v}}_j(\mathbf{v}_i) && \text{Def. Lin.komb von Abbildungen} \\ &= b_i && \check{\mathbf{v}}_j(\mathbf{v}_i) = 0 \text{ wenn } j \neq i, \check{\mathbf{v}}_i(\mathbf{v}_i) = 1. \end{aligned}$$

Also gilt $b_1 = \cdots = b_n = 0$. Die Abbildungen $\check{\mathbf{v}}_1, \ldots, \check{\mathbf{v}}_n \in V^\vee$ sind also linear unabhängig.

Schritt 3: Die Abbildungen $\check{\mathbf{v}}_1, \ldots, \check{\mathbf{v}}_n \in V^\vee$ sind ein Erzeugendensystem von $V^\vee$: Ist $f\colon V \to K$ ein beliebiges Element von $V^\vee$, so gilt

$$f = \sum_{j=1}^{n} f(\mathbf{v}_j)\check{\mathbf{v}}_j,$$

denn für alle $i = 1, \ldots n$ gilt

$$\begin{aligned} \Big(\sum_{j=1}^{n} f(\mathbf{v}_j)\check{\mathbf{v}}_j\Big)(\mathbf{v}_i) &= \sum_{j=1}^{n} f(\mathbf{v}_j)\check{\mathbf{v}}_j(\mathbf{v}_i) \\ &= f(\mathbf{v}_i)\check{\mathbf{v}}_i(\mathbf{v}_i) \qquad = f(\mathbf{v}_i). \end{aligned}$$

Die linearen Abbildungen f und $\sum_{j=1}^{n} f(\mathbf{v}_j)\check{\mathbf{v}}_j$ stimmen also auf einer Basis überein, sind also gleich. Damit haben wir gezeigt, dass wir alle Elemente $f \in V^\vee$ als Linearkombination der Abbildungen $\check{\mathbf{v}}_1, \ldots, \check{\mathbf{v}}_n$ schreiben können.

Das zeigt, dass die Elemente $\check{\mathbf{v}}_1, \ldots, \check{\mathbf{v}}_n \in V^\vee$ eine Basis von $V^\vee$ sind, insbesondere ist $\dim V^\vee = n = \dim V$. □

KURZE ZUSAMMENFASSUNG DES ARGUMENTES: Lineare Abbildungen $F\colon V \to W$ sind eindeutig durch die Werte auf einer Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V definiert. Wenden wir das auf $W = K$ an, so sehen wir dass lineare Abbildungen $f\colon V \to K$ eindeutig durch die n-Zahlen $b_1 = f(\mathbf{v}_1), \ldots, b_n = f(\mathbf{v}_n) \in K$ definiert sind. Für den Raum dieser Abbildungen sind also die Abbildungen $f_i = \check{\mathbf{v}}_i$, mit $f_i(\mathbf{v}_i) = 1$ und $f(\mathbf{v}_j) = 0$ für $j \neq i$ eine Basis.

WIE KOMME ICH AUF DAS ARGUMENT?

1. *Wie immer, ist der erste Schritt, sich zu erinnern was die Dimension exakt bedeutet: Die Dimension ist die Anzahl der Elemente einer Basis. Wir müssen also eine Basis von $V^\vee$ basteln.*

2. *Wie immer muss ich außerdem die Definition parat haben:*

$$V^\vee = \{f\colon V \to K \mid f \text{ linear}\}$$

ist ein Vektorraum von Abbildungen. Elemente konstruieren, heißt hier also Abbildungen angeben!

3. *Wenn mir nichts einfällt, schaue ich das Beispiel $V = K^n$ an. Hier ist $(K^n)^\vee = \mathrm{Mat}_{1,n}(K)$, denn lineare Abbildungen von K^n nach K sind gerade $1 \times n$-Matrizen, also Zeilenvektoren. Für den Raum der Zeilenvektoren sehen Sie die Basis $\check{e}_1 = (10\dots0), \check{e}_2 = (010\dots0), \dots, \check{e}_n = (0\dots01)$.*

4. *Was für Abbildungen entsprechen den Zeilenvektoren $\check{e}_i = (0\dots010\dots0)$? Das ist die Abbildung $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_i$, also die Projektion auf die $i-$te Koordinate.*

5. *Neue Frage: Kann ich für einen allgemeinen Vektorraum sinnvoll eine „Projektion auf die i-te Koordinate“ definieren? Nicht wirklich, aber wir hatten Koordinaten bezüglich einer Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ von V eingeführt, d.h. die i-te Koordinate von $\mathbf{v} \in V$ bezüglich einer Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ von V ist sinnvoll definiert und das liefert Abbildungen $\check{\mathbf{v}}_i \colon V \to K$.*

6. *Jetzt kann ich ausprobieren, ob die Abbildungen $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n$ ein linear unabhängiges Erzeugendensystem sind. So funktioniert das Argument des Beweises.*

Bemerkung. 1. Für unendlichdimensionale Vektorräume ist die Dimension des Dualraums $V^\vee$ größer als die von V. Da wir uns noch nicht überlegt haben was wir genau mit „Dimension“ meinen, wenn diese nicht endlich ist, möchte ich das hier nur erwähnen, in der Funktionalanalysis werden Sie dazu mehr lernen und auch eine Reparaturmöglichkeit für dieses Problem sehen.

2. Die aus einer Basis $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ konstruierte Basis $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n$ heißt (zu $\mathbf{v}_1, \dots, \mathbf{v}_n$) *duale Basis* von $V^\vee$.

   Warnung: Die Abbildungen $\check{\mathbf{v}}_i$ hängen von der ganzen Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ ab, nicht nur von $\mathbf{v}_i$.

Beispiel 51. Ist $V = \mathbb{R}^2$ und $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, so ist

$$\check{\mathbf{v}}_1(\begin{pmatrix} 2 \\ 1 \end{pmatrix}) = \check{\mathbf{v}}_1(1 \cdot \mathbf{v}_1 + 1 \cdot \mathbf{v}_2) = 1.$$

Aber für $\mathbf{e}_1 = \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ gilt

$$\check{\mathbf{e}}_1(\begin{pmatrix} 2 \\ 1 \end{pmatrix}) = \check{\mathbf{e}}_1(2 \cdot \mathbf{e}_1 + 1 \cdot \mathbf{e}_2) = 2.$$

Eine weitere Möglichkeit aus Vektoren, Abbildungen zu machen haben wir mit dem Skalarprodukt, oder allgemeiner mit Bilinearformen kennen gelernt. Für Bilinearformen haben Sie auf dem Übungsblatt schon den folgenden Begriff kennen gelernt.

Definition. Eine Bilinearform $B\colon V \times V \to K$ auf einem K-Vektorraum V heißt *nicht ausgeartet* wenn es zu jedem von 0 verschiedenen Vektor $\mathbf{v} \in V \smallsetminus \{0\}$ einen Vektor $\mathbf{w} \in V$ gibt so dass $B(\mathbf{v}, \mathbf{w}) \neq 0$.

Behauptung 52. *Ist $B\colon V \times V \to K$ eine Bilinearform auf einem K-Vektorraum V, dann ist die Abbildung*

$$\begin{aligned} F_B\colon V &\to V^\vee = \{f\colon V \to K \mid f \text{ linear}\} \\ \mathbf{v} &\mapsto F_B(\mathbf{v})\colon V \to K, \text{ definiert durch} \\ &\qquad \mathbf{w} \mapsto F_B(\mathbf{v})(\mathbf{w}) := B(\mathbf{v}, \mathbf{w}) \end{aligned}$$

linear.

Ist V endlich dimensional, so ist die Abbildung F_B ist genau dann ein Isomorphismus, wenn B nicht ausgeartet ist.

Bitte lesen Sie die Aussage der Behauptung einmal ganz in Ruhe und versuchen Sie dabei, die komplizierten Formeln einmal zu verdauen. Ist Ihnen zum Beispiel klar, dass die Abbildung F_B im Fall dass $V = \mathbb{R}^n$ und $B = \langle , \rangle$ das Standardskalarprodukt ist, einfach wieder die Abbildung $\mathbf{v} \mapsto \mathbf{v}^t$ ist? Nehmen Sie sich einmal die Zeit, sich das klar zu machen bevor Sie weiter lesen. Wenn Ihnen das nicht gelingt, fragen Sie nach, zum Beispiel im anonymen Forum auf Moodle. Das Verstehen von mathematischen Aussagen braucht Zeit und die müssen Sie sich selbst nehmen.

Eine kürzere Notation für die Abbildung F_B wäre

$$\mathbf{v} \mapsto B(\mathbf{v}, \_\_) : V \to K,$$

wenn wir uns darauf einigen, dass wir mit __ die Stelle in einer Formel für eine Abbildung markieren, an der wir Werte einsetzen. Das kann ich schneller lesen, als die sperrigere Notation $F_B(\mathbf{v})(\mathbf{w}) := B(\mathbf{v}, \mathbf{w})$ aus der Behauptung.

Beweis. Die Abbildung F_B ist wohldefiniert, d.h. für jedes $\mathbf{v} \in V$ definiert die Formel $\mathbf{w} \mapsto B(\mathbf{v}, \mathbf{w})$ tatsächlich eine lineare Abbildung, da die Abbildung B nach Voraussetzung in der 2. Variablen linear ist.

Die Abbildung F_B ist linear, d.h. es gelten die Gleichheiten von Abbildungen $F_B(\mathbf{v} + \mathbf{v}') = F_B(\mathbf{v}) + F_B(\mathbf{v}')$ und $F_B(c\mathbf{v}) = cF_B(\mathbf{v})$ für alle $\mathbf{v}, \mathbf{v}' \in V$, $c \in K$, da die Abbildung B in der ersten Variablen linear ist.

Die lineare Abbildung F_B ist genau dann injektiv wenn $\operatorname{Ker}(F_B) = \{0\}$ gilt. Aber $v \in \operatorname{Ker}(F_B)$ bedeutet, dass $F_B(\mathbf{v}) \in V^\vee$ die 0-Abbildung ist. Das ist gleichbedeutend mit der Aussage, dass $B(\mathbf{v}, \mathbf{w}) = 0$ für alle $\mathbf{w} \in V$ gilt. Diese Bedingung gilt genau dann nur für $\mathbf{v} = 0$ wenn B nicht ausgeartet ist. Also ist F_B genau dann injektiv, wenn B nicht ausgeartet ist.

Ist V endlich dimensional, so gilt nach der vorigen Behauptung, dass $\dim V = \dim V^\vee$. In diesem Fall ist F_B aber nach der Dimensionsformel genau dann injektiv, wenn F_B ein Isomorphismus ist. Also ist F_B für endlich dimensionale V genau dann ein Isomorphismus wenn B nicht ausgeartet ist. $\square$

Bemerkung. Die Konstruktion aus der Behauptung lässt sich umdrehen: Ist umgekehrt $F\colon V \to V^\vee$ linear, so ist die Abbildung

$$\begin{aligned} B_F\colon V \times V &\to K \\ (\mathbf{v}, \mathbf{w}) &\mapsto B_F(\mathbf{v}, \mathbf{w}) := F(\mathbf{v})(\mathbf{w}) \end{aligned}$$

eine Bilinearform, denn $B_F(,)$ ist in der zweiten Variablen linear, da für alle $\mathbf{v}$ die Abbildung $F(\mathbf{v}) \in V^\vee = \{f\colon V \to K \mid f \text{ linear}\}$ nach Definition linear ist und B_F ist in der ersten Variablen linear, weil F linear ist.

Die Konstruktionen $B \to F_B$ und $F \mapsto B_F$ sind invers zueinander, definieren also Bijektionen

$$\begin{aligned} \{B\colon V \times V \to K \mid B \text{ bilinear}\} &\longleftrightarrow \{F\colon V \to V^\vee \mid F \text{ linear}\} \\ B &\mapsto F_B \\ B_F &\leftarrow\!\shortmid F \end{aligned}$$

die sogar Isomorphismen von Vektorräumen sind.

Diese Bemerkung erklärt dass, wie die Grammatrix einer Bilinearform tatsächlich als lineare Abbildung auffassen können. Sei nämlich $B\colon V \times V \to K$ eine Bilinearform, $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Basis von V. Wir bezeichnen die dazu duale Basis von $V^\vee$ wieder mit $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n$.

Lassen Sie uns damit einmal die Matrix ${}_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\check{\mathbf{v}}}}(F_B)$ berechnen: Die Regel hierfür war, dass in den Spalten der Matrix die Koordinaten der Bilder der Basisvektoren stehen. Sei also $j \in \{1, \dots, n\}$, dann müssen wir $F_B(\mathbf{v}_j) \in V^\vee$ als Linearkombination der $\check{\mathbf{v}}_i$ schreiben. Da wir den i-ten Koeffizienten b_i einer Linearkombination $\sum_{k=1}^n b_k \check{\mathbf{v}}_k$ erhalten, indem wir $\mathbf{v}_i$ in die Abbildung einsetzen, ist die $\check{\mathbf{v}}_i$-Koordinate von $F_B(\mathbf{v}_j)$ gleich $F_B(\mathbf{v}_j)(\mathbf{v}_i) = B(\mathbf{v}_j, \mathbf{v}_i)$. Das ist genau der $(j,i)-$te Eintrag der Grammatrix von B.

Merke: Matrizen gehören immer zu linearen Abbildungen!

Mit der Behauptung können wir uns jetzt ohne Rechnung noch einmal überlegen, dass es adjungierte Abbildungen gibt. Dieses Argument funktioniert jetzt sogar für beliebige Bilinearformen.

Folgerung 53. *Sei V ein endlich dimensionaler K-Vektorraum, $B\colon V \times V \to K$ eine nicht ausgeartete Bilinearform und $A\colon V \to V$ eine lineare Abbildung.*

Dann existiert genau eine lineare Abbildung $A^*\colon V \to V$ so dass für alle $\mathbf{v}, \mathbf{w} \in V$ gilt, dass

$$B(A^*\mathbf{v}, \mathbf{w}) = B(\mathbf{v}, A\mathbf{w}).$$

Die Abbildung A^* heißt die adjungierte Abbildung von A bezüglich der Bilinearform B.

Beweis. Wir müssen uns nur überlegen, dass es für jedes $\mathbf{v} \in V$ genau einen Vektor $\mathbf{v}' =: A^*\mathbf{v} \in V$ gibt, so dass $B(\mathbf{v}', \mathbf{w}) = B(\mathbf{v}, A\mathbf{w})$. Dafür bemerken wir zunächst, dass die Abbildung

$$f_{\mathbf{v}} \colon V \to K; \mathbf{w} \mapsto B(\mathbf{v}, A\mathbf{w})$$

linear, also ein Element von $V^\vee$ ist, da A linear ist und B in der 2. Variablen linear ist. Da B nicht ausgeartet ist, gibt es aber zu jedem Element $f \in V^\vee$ genau ein $\tilde{\mathbf{v}} \in V$ so dass $f(\_\_) = B(\tilde{\mathbf{v}}, \_\_)$. Insbesondere existiert also genau ein $\mathbf{v}' \in V$ so dass $B(\mathbf{v}', \_\_) = B(\mathbf{v}, A\_\_)$. Es muss also $A^*\mathbf{v} := \mathbf{v}'$ sein.

Die Abbildung $\mathbf{v} \mapsto A^*\mathbf{v}$ ist linear, denn nach Definition ist diese Abbildung die Komposition der linearen Abbildungen:

$$V \xrightarrow{\mathbf{v} \mapsto B(\mathbf{v}, A\_\_)} V^\vee \xrightarrow{F_B^{-1}} V.$$

□

Sie brauchen vielleicht einen Moment, um den letzten Schritt des Argumentes zu verstehen. Hier hätten wir genauso gut wieder die Definition von „linear“ nachrechnen können, aber es ist auf Dauer eine große Ersparnis, wenn wir das was wir uns schon früher überlegt haben auch benutzen.

Das braucht Gewöhnungszeit, aber wir sollte anfangen, das hin und wieder zu versuchen.

Beispiel 54. Für das Standardskalarprodukt auf K^n hatten wir uns schon überlegt, dass $A^* = A^t$ durch die transponierte Matrix gegeben ist.

Für die Bilinearform $B \colon K^4 \to K^4$ die durch die symplektische Bilinearform $B(\mathbf{v}, \mathbf{w}) := \mathbf{v}^t \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ gegeben ist, können wir A^* genauso ausrechnen. Für die Rechnung schreiben ich $G := \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, dann ist $G^{-1} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} = G^t$. Ist $A \colon K^4 \to K^4$ eine 4×4 Matrix dann gilt:

$$\begin{aligned} B(\mathbf{v}, A\mathbf{w}) &= \mathbf{v}^t GA\mathbf{w} = \mathbf{v}^t GAG^{-1}G\mathbf{w} \\ &= (\mathbf{v}^t GAG^{-1})G\mathbf{w} = \left(G^{-1^t} A^t G^t \mathbf{v}\right)^t G\mathbf{w} = B(GA^tG^t\mathbf{v}, \mathbf{w}). \end{aligned}$$

Also ist dann $A^* = GA^tG^t$.

Bitte rechnen Sie selbst einmal nach, was die obige Gleichung $A^* = GA^tG^t$ für eine 4×4-Matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$ konkret bedeutet.

Bemerkung. Wenn wir uns die Aussage der Behauptung noch einmal anschauen, sehen sie, dass wir umgekehrt für jede lineare Abbildung $F \colon V \to V^\vee$ eine Bilinearform

$$\begin{aligned} B_F \colon & V \times V \to K \\ & (v, w) \mapsto B_F(\mathbf{v}, \mathbf{w}) := F(\mathbf{v})(\mathbf{w}) \end{aligned}$$

definieren können. Diese Abbildung ist in der ersten Variablen $\mathbf{v}$ linear, weil F linear ist und in der 2. Variablen, weil $F(\mathbf{v}) \in V^\vee$ eine lineare Abbildung von V nach K ist.

Die duale Abbildung

Wenn Sie aufgepasst haben, haben Sie gerade gelernt, dass Bilinearformen auf V das gleiche sind wie lineare Abbildungen $V \to V^\vee$. Bei Bilinearformen hatten wir uns aber insbesondere für symmetrische Bilinearformen interessiert. Damit sollten Sie fragen, was nun symmetrisch für eine Abbildung $F\colon V \to V^\vee$ bedeuten soll.

Die Antwort dazu liefert eine einfache, abstrakte Konstruktion: Haben wir eine lineare Abbildung $A\colon V \to W$ zwischen K-Vektorräumen und ein Element $f \in W^\vee$, also eine lineare Abbildung $f\colon W \to K$, so ist die Komposition $f(A\_\_) = f \circ A\colon V \to K$ ein Element von $V^\vee$, d.h. A definiert eine Abbildung von $W^\vee \to V^\vee$. Diese Abbildung heißt duale Abbildung.

Definition. Die *duale Abbildung* einer linearen Abbildung $A\colon V \to W$ von K-Vektorräumen ist die lineare Abbildung

$$\begin{aligned} A^\vee\colon W^\vee &\to V^\vee \\ f &\mapsto f \circ A = f(A\_\_)\colon V \to K. \end{aligned}$$

$W^\vee = \{f\colon W \to K \mid f \text{ linear}\}$
$V^\vee = \{g\colon V \to K \mid g \text{ linear}\}$

Als Übung sollten wir die Matrix der dualen Abbildung berechnen. Sei also $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ eine Basis von V und $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n$ die dazu duale Basis von $V^\vee$, , $\mathbf{w}_1, \dots, \mathbf{w}_m$ eine Basis von W und $\check{\mathbf{w}}_1, \dots, \check{\mathbf{w}}_m$ die duale Basis von $W^\vee$. Wir wollen

$$_{\underline{\check{\mathbf{v}}}}\mathrm{Mat}_{\underline{\check{\mathbf{w}}}}(A^\vee)$$

bestimmen.

Wir bekommen keine Panik, sondern überlegen uns einfach, was die ganzen Symbole bedeuten. In den Spalten der Matrix stehen die Koordinaten der Bilder der Basisvektoren. Also müssen wir $A^\vee(\mathbf{w}_j)$ berechnen und als Linearkombination der $\check{\mathbf{v}}_i$ schreiben.

Nach Definition ist $A^\vee(\check{\mathbf{w}}_j) = \check{\mathbf{w}}_j(A\_\_)$. Das ist eine Abbildung von V nach K. Die Koordinate einer Abbildung $f\colon V \to K$ bezüglich der dualen Basis, ist wie zuvor $f(\mathbf{v}_i)$. Also ist der i, j-te Eintrag der Matrix gleich $\check{\mathbf{w}}_j(A\mathbf{v}_i)$.

Aber $\check{\mathbf{w}}_j(\mathbf{w})$ ist die $j-$te Koordinate des Vektors $\mathbf{w}$ bezüglich unserer Basis $\underline{\mathbf{w}}$, d.h. $\check{\mathbf{w}}_j(A\mathbf{v}_i)$ ist der j, i-te Eintrag der Matrix $_{\underline{\mathbf{v}}}\mathrm{Mat}_{\underline{\mathbf{w}}}A)$!

Folgerung 55 (Die transponierte Matrix ist die Matrix der dualen Abbildung). *Ist $F\colon V \to W$ linear, $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ eine Basis von V und $\check{\mathbf{v}}_1, \dots, \check{\mathbf{v}}_n$ die dazu duale Basis von $V^\vee$, , $\mathbf{w}_1, \dots, \mathbf{w}_m$ eine Basis von W und $\check{\mathbf{w}}_1, \dots, \check{\mathbf{w}}_m$ die duale Basis von $W^\vee$. Ist $A = {}_{\underline{\mathbf{w}}}\mathrm{Mat}_{\underline{\mathbf{v}}}(F)$ die Matrix von F bezüglich der gewählten Basen, so ist die Matrix der dualen Abbildung $A^\vee$ bezüglich der dualen Basen, die Transponierte A^t.*

Dieses Resultat erklärt uns, zu welcher linearen Abbildung, die transponierte einer Matrix gehört! In der linearen Algebra 1 war unklar geblieben was die transponierte Matrix bedeutet, denn für Gleichungssysteme war transponieren nicht unmittelbar sinnvoll.

WIR KÖNNEN DAMIT noch ein Resultat neu verstehen: Um eine neue lineare Abbildung $A^\vee\colon W^\vee \to V^\vee$ kennenzulernen, könnten wir zunächst schauen, was der Kern der Abbildung ist. Dafür nehmen wir die Definition und schauen, dass wir in der Definition alle dual-Symbole „$\vee$“ entschlüsseln:

Aufpassen: Die Zeichenkette „$= 0$“ ist tückisch! In $V^\vee$ ist das 0-Element die Nullabbildung.

$$\begin{aligned}
\operatorname{Ker}(A^\vee) &= \{f \in W^\vee \mid A^\vee(f) = 0 \in V^\vee\} && \text{Def. Kern}\\
&= \left\{ f\colon W \to K \;\middle|\; \begin{array}{c} f \text{ linear} \\ f(A__) = 0 \in V^\vee \end{array} \right\} && \text{Def. } W^\vee, A^\vee\\
&= \left\{ f\colon W \to K \;\middle|\; \begin{array}{c} f \text{ linear} \\ f(A\underline{\mathbf{v}}) = 0 \in K\,\forall \mathbf{v} \in V \end{array} \right\} && 0 \in V^\vee \text{ bedeutet: Nullabbildung}\\
&= \left\{ f\colon W \to K \;\middle|\; \begin{array}{c} f \text{ linear} \\ f(\mathbf{w}) = 0 \in K\,\forall \mathbf{w} \in \operatorname{Bild}(A) \end{array} \right\}. && \text{Def. } \operatorname{Bild}(A)
\end{aligned}$$

Also ist der Kern von $A^\vee$ die Menge der linearen Abbildungen $W \to K$, die auf dem Bild von A verschwinden.

AUF BLATT 11 werden Sie allgemein für einen Unterraum $U \subseteq V$ eine Formel für den Unterraum

$$U^\top := \{f \in V^\vee \mid f(\mathbf{u}) = 0\,\forall \mathbf{u} \in U\} \subseteq V^\vee$$

finden.

Wenn alles gut geht sehen Sie, wenn Sie diese Formel für $\operatorname{Ker}(A^\vee) = \operatorname{Bild}(A)^\top$ einsetzen, dass $\dim(\operatorname{Bild}(A^\vee)) = \dim(\operatorname{Bild}(A))$. Das zeigt:

Folgerung 56. *Für die duale Abbildung gilt*

$$\operatorname{Rang}(A^\vee) = \operatorname{Rang}(A).$$

DAS IST EIN NEUER BEWEIS VON „Zeilenrang=Spaltenrang“, denn die Matrix der dualen Abbildung ist die transponierte der Matrix von A. Wenn Sie den Beweis aus der linearen Algebra 1 anschauen, sehen Sie, das wir dort dafür ein wenig gerechnet hatten. Der neue Beweis kam jetzt ganz ohne Rechnen aus.

IN DER GLOBALÜBUNG haben wir uns einen Spezialfall der Folgerung noch einmal per Hand überlegt:

Folgerung 57. *Ist $A\colon V \to W$ eine lineare Abbildung endlichdimensionaler K-Vektorräume, so gilt*

1. *A ist genau dann surjektiv, wenn $A^\vee$ injektiv ist.*

2. *A ist genau dann injektiv, wenn $A^\vee$ surjektiv ist.*

Beweis. Version 1 (verwende Folgerung 56): Das folgt aus der Aussage über den Rang, denn $A\colon V \to W$ ist genau dann surjektiv, wenn

$\mathrm{Rang}(A) = \dim W$ und $A^\vee\colon W^\vee \to V^\vee$ ist genau dann injektiv, wenn $\mathrm{Rang}(A^\vee) = \dim W^\vee = \dim W$. Damit folgt 1. aus der Gleichheit $\mathrm{Rang}(A) = \mathrm{Rang}(A^\vee)$.

Für 2. können wir genauso argumentieren: $A\colon V \to W$ ist genau dann injektiv, wenn $\mathrm{Rang}(A) = \dim V$ und $A^\vee\colon W^\vee \to V^\vee$ ist genau dann surjektiv, wenn $\mathrm{Rang}(A^\vee) = \dim V^\vee = \dim V$. Damit folgt auch 2. aus der Gleichheit $\mathrm{Rang}(A) = \mathrm{Rang}(A^\vee)$.

Version 2 (Direktes Argument): Die Abbildung $A^\vee\colon W^\vee \to V^\vee$ ist genau dann injektiv, wenn $\mathrm{Ker}(A^\vee) = \{0\} \subseteq W^\vee$ ist. Den Kern hatten wir schon berechnet als:

$$\begin{aligned}
\mathrm{Ker}(A^\vee) &= \{f \in W^\vee \mid A^\vee(f) = 0 \in V^\vee\} && \text{Def. Kern}\\
&= \left\{ f\colon W \to K \mid \begin{matrix} f \text{ linear} \\ f(A__) = 0 \in V^\vee \end{matrix} \right\} && \text{Def. } W^\vee, A^\vee\\
&= \left\{ f\colon W \to K \mid \begin{matrix} f \text{ linear} \\ f(A\underline{\mathbf{v}}) = 0 \in K\, \forall \mathbf{v} \in V \end{matrix} \right\} && 0 \in V^\vee \text{ bedeutet: Nullabbildung}\\
&= \left\{ f\colon W \to K \mid \begin{matrix} f \text{ linear} \\ f(\mathbf{w}) = 0 \in K\, \forall \mathbf{w} \in \mathrm{Bild}(A) \end{matrix} \right\}. && \text{Def. } \mathrm{Bild}(A)
\end{aligned}$$

Also gilt $\mathrm{Ker}(A^\vee) = \{0\}$ dass eine $f\colon W \to K$ genau dann die 0-Abbildung ist, wenn $f(\mathbf{w}) = 0$ für alle $\mathbf{w} \in \mathrm{Bild}(A) \subseteq W$. Das gilt sicher, wenn $\mathrm{Bild}(A) = W$. Ist umgekehrt $\mathrm{Bild}(A) \subsetneq W$, so können wir eine Basis $\mathbf{w}_1, \dots, \mathbf{w}_r$ von $\mathrm{Bild}(A)$ zu einer Basis $\mathbf{w}_1, \dots, \mathbf{w}_r, \mathbf{w}_{r+1}, \dots, \mathbf{w}_m$ von W ergänzen und dann ist $\check{\mathbf{w}}_{r+1}$ ein Element von $W^\vee \smallsetminus \{0\}$ für das $\check{\mathbf{w}}_{r+1}(\mathbf{w}) = 0$ für alle $\mathbf{w} \in \mathrm{Bild}(A)$. Also ist dann $\mathrm{Ker}(A^\vee) \neq \{0\}$. Das zeigt 1.

Die 2. Aussage sollten Sie sich zur Übung noch einmal selbst überlegen. □

Quotientenvektorräume

Der Dualraum hat uns erklärt, wie wir der transponierten Abbildung auch ohne die Wahl eines Skalarproduktes einen Sinn geben können. Der Quotientenvektorraum liefert ähnlich eine Erklärung, wie wir ohne Wahlen einen Vektorraum finden können, der die Nützlichkeit eines Komplementes eines Unterraums hat. Dabei werden wir uns noch einmal an Äquivalenzrelationen aus dem ersten Semester erinnern. Als Bonus erhalten wir damit eine besonders einfache Konstruktion der reellen Zahlen $\mathbb{R}$, die sehr schön erklärt, wie wir die Idee der Vervollständigung der Zahlengerade mit linearer Algebra exakt formulieren können.

Motivation: Es ist leicht, eine beliebige lineare Abbildung $F\colon V \to W$ zu einer surjektiven Abbildung zu machen, indem wir einfach W durch den Unterraum $\mathrm{Bild}(F) \subseteq W$ ersetzen: Die Abbildung

$$\begin{aligned} F\colon V &\to \mathrm{Bild}(F)\\ \mathbf{v} &\mapsto F(\mathbf{v}) \end{aligned}$$

ist nach Definition surjektiv. Spätestens nachdem wir gesehen haben, dass die duale Abbildung genau dann surjektiv ist, wenn die Abbildung selbst injektiv ist, werden Sie erwarten, dass es eine ähnliche Konstruktion gibt, die aus einer linearen Abbildung eine injektive Abbildung macht.

FRAGE: Wie machen wir aus einer linearen Abbildung $F\colon V \to W$ eine injektive Abbildung?

ANTWORT: Betrachte die Elemente von V nur bis auf Addition von Elementen aus $\mathrm{Ker}(F)$.

ERINNERUNG: In der linearen Algebra 1 hatten wir das Prinzip, verschiedene Symbole als gleichbedeutend aufzufassen, das Sie in der Schule in der Bruchrechnung kennengelernt hatten, mit dem Begriff der Äquivalenzrelation formal exakt formuliert. Die Definition war:

Definition. Eine *Relation* auf einer Menge M ist eine Teilmenge $R \subset M \times M$. Für zwei Elemente $a, b \in M$ schreiben wir

$$a \sim_R b :\Leftrightarrow (a, b) \in R.$$

Ich stelle mir $M \times M$ hier als Kreuztabelle vor, R gibt die angekreuzten Paare (a, b) an.

$\sim_R$ lese ich als „a ist in Relation zu b“.

Eine *Äquivalenzrelation* auf einer Menge M ist eine Relation $R \subset M \times M$ für die gilt, dass

- (reflexiv) Für alle $m \in M$ gilt $m \sim_R m$.
- (transitiv) Gilt $m \sim_R m'$ und $m' \sim_R m''$ so folgt dass $m \sim_R m''$.
- (symmetrisch) Gilt $m \sim_R m'$ so gilt auch $m' \sim_R m$.

m sollte gleichwertig mit sich selbst sein

Ist $\sim$ eine Äquivalenzrelation, so bezeichnen wir:

- Für ein Element $m \in M$ die Menge

  $$[m] := \{m' \in M \mid m \sim_R m'\}$$

  als *Äquivalenzklasse* des Elementes m.

  Jedes Element $m' \in [m]$ heißt *Vertreter der Äquivalenzklasse* $[m]$.

- Die Menge der Äquivalenzklassen:

  $$M/\sim_R := \{A \subseteq M \mid A = [m] \text{ für ein } m \in M\}$$

  als *Quotientenmenge* der Äquivalenzrelation.

$[m]$ ist die Menge der gleichwertigen Symbole für m.

WICHTIGE BEISPIELE von Quotientenmengen waren die rationalen Zahlen $\mathbb{Q}$, die wir als Quotientenmenge der Bruchsymbole bezüglich der Relation $\frac{a}{b} \sim \frac{a'}{b'} :\Leftrightarrow ab' = a'b$ — also der Relation, die die Werte der Bruchsymbole vergleicht indem es die Bruchsymbole auf den Hauptnenner bringt — aufgefasst hatten. Ein zweites Beispiel war $\mathbb{Z}/n\mathbb{Z}$, die Menge der Restklassen bei ganzzahliger Division durch n.

Lemma/Definition 58 (Quotientenvektorraum). *Der Quotientenvektorruam V/U eines K-Vektorraums V nach einem Unterraum $U \subseteq V$, ist die Quotientenmenge*

$$V/U := V/\sim_U$$

bezüglich der Äquivalenzrelation

$$\mathbf{v} \sim_U \mathbf{v}' \Leftrightarrow \mathbf{v}' = \mathbf{v} + \mathbf{u} \text{ für ein } \mathbf{u} \in U.$$

Diese Menge ist zusammen mit den Verknüpfungen

$$\begin{aligned} +\colon V/U \times V/U &\to V/U & \cdot\colon K \times V/U &\to V/U \\ (\mathbf{v}_1, \mathbf{v}_2) &\mapsto [\mathbf{v}_1] + [\mathbf{v}_2] := [\mathbf{v}_1 + \mathbf{v}_2] & (c, \mathbf{v}) &\mapsto c \cdot [\mathbf{v}] := [c\mathbf{v}] \end{aligned}$$

ein K-Vektorraum und die Quotientenabbildung

$$q\colon V \to V/U \tag{1}$$
$$\mathbf{v} \mapsto q(\mathbf{v}) := [\mathbf{v}] \tag{2}$$

ist eine surjektive, lineare Abbildung mit $\mathrm{Ker}(q) = U$.

Beweis. Das Lemma ist wiederum „nur" eine Übung darin, die Bedeutung der Symbole und Aussagen zu verstehen, auszuschreiben und dann kompakt damit zu argumentieren, um unnötige Schreibarbeit zu vermeiden.

1. Die Relation $\mathbf{v} \sim_U \mathbf{v}' \Leftrightarrow \mathbf{v}' = \mathbf{v} + \mathbf{u}$ für ein $\mathbf{u} \in U$ ist gleichbedeutend mit $\mathbf{v} \sim_U \mathbf{v}' \Leftrightarrow \mathbf{v}' - \mathbf{v} \in U$. Damit ist die Relation reflexiv, da $0 \in U$, symmetrisch, da $\mathbf{v}' - \mathbf{v} \in U$ auch bedeutet, dass $\mathbf{v} - \mathbf{v}' = -1 \cdot (\mathbf{v}' - \mathbf{v}) \in U$ und transitiv, da aus $\mathbf{v}' - \mathbf{v} \in U, \mathbf{v}'' - \mathbf{v}' \in U$ folgt, dass
$$(\mathbf{v}'' - \mathbf{v}) = (\mathbf{v}'' - \mathbf{v}') + (\mathbf{v}' - \mathbf{v}) \in U.$$
Damit ist $\sim_U$ wie behauptet eine Äquivalenzrelation.

2. Aus den Unterraumaxiomen für U folgt, dass die Verknüpfungen $+, \cdot$ wohldefiniert sind, d.h. dass die genannten Formeln tatsächlich Abbildungen definieren, denn: Ist $[\mathbf{v}_1] = [\mathbf{v}_1'], [\mathbf{v}_2] = [\mathbf{v}_2'] \in V/U$, so ist $\mathbf{v}_1' = \mathbf{v}_1 + \mathbf{u}_1$ für ein $\mathbf{u}_1 \in U$ und $\mathbf{v}_2' = \mathbf{v}_2 + \mathbf{u}_2$ für ein $\mathbf{u}_2 \in U$. Dann ist aber auch
$$[\mathbf{v}_1' + \mathbf{v}_2'] = [\mathbf{v}_1 + \mathbf{u}_1 + \mathbf{v}_2 + \mathbf{u}_2] = [\mathbf{v}_1 + \mathbf{v}_2] \in V/U,$$
weil $\mathbf{u}_1 + \mathbf{u}_2 \in U$.

   Genauso ist für $c \in K$ dann $[c(\mathbf{v}_1' + \mathbf{u}_1)] = [c\mathbf{v}_1 + c\mathbf{u}_1] = [c\mathbf{v}_1]$, weil mit $\mathbf{u}_1 \in U$ auch $c\mathbf{u}_1 \in U$ gilt.

   Also sind die Verknüpfungen wohldefiniert. Da die Verknüpfungen $+, \cdot$ auf V die Vektorraumaxiome erfüllen übertragen sich diese automatisch auf V/U, da alle Formeln auf V/U aus den entsprechenden Formeln auf V dadurch entstehen, dass wir [] um alle Symbole setzen.

3. Das zeigt auch dass die Abbildung $\mathbf{v} \mapsto [\mathbf{v}]$ linear ist, Quotientenabbildungen sind nach Konstruktion surjektiv und es gilt $[\mathbf{v}] = [0] \Leftrightarrow \mathbf{v} \in U$, also ist $\mathrm{Ker}(q) = U$.

□

DIE ERSTEN FRAGEN an einen neuen Vektorraum sind vielleicht: Was ist die Dimension? Wie finden wir eine Basis? Gibt es Beispiele in denen wir uns den Raum geometrisch vorstellen können?

Die Dimension können wir schon ablesen, denn mit dem Lemma können wir die Dimensionsformel auf die surjektive Quotientenabbildung $q\colon V \to V/U$ anwenden und finden

$$\dim V/U = \dim V - \dim U.$$

Wenn wir uns an den Beweis der Dimensionsformel erinnern, können wir das Argument auch benutzen, um eine Basis von V/U zu finden.

Bemerkung. Ergänzen wir eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$ von U zu einer Basis $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ von V so sind die Elemente $[\mathbf{v}_{m+1}], \ldots, [\mathbf{v}_n]$ eine Basis von V/U.

Beweis. Da die Elemente $\mathbf{v}_1, \ldots, \mathbf{v}_n$ den Vektorraum V erzeugen und die Quotientenabbildung surjektiv ist, sind die Elemente $q(\mathbf{v}_1) = [\mathbf{v}_1], \ldots, q(\mathbf{v}_n) = [\mathbf{v}_n]$ ein Erzeugendensystem von V/U und nach Konstruktion gilt $q(\mathbf{v}_1) = \cdots = q(\mathbf{v}_m) = [0]$, also sind auch die Vektoren $[\mathbf{v}_{m+1}], \ldots, [\mathbf{v}_n]$ ein Erzeugendensystem und da die Anzahl der Elemente dieses Erzeugendensystems gleich $\dim V/U$ ist, ist es eine Basis. □

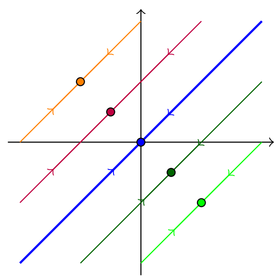
Beispiel 59. Ist $U = \operatorname{Span}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \subset \mathbb{R}^2$ die Diagonale, so können wir $\mathbb{R}^2/U$ gut verstehen.

Zunächst können wir hier verstehen, wieso wir im Beweis oben mit einer Basis von U begonnen haben, nicht einfach mit einer beliebigen Basis von V, denn sind $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ die Standardbasis von $\mathbb{R}^2$, so gilt

$$[\mathbf{e}_1] = [\begin{pmatrix} 1 \\ 0 \end{pmatrix}] = [\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}] = [\begin{pmatrix} 0 \\ -1 \end{pmatrix}] = -[\mathbf{e}_2] \in \mathbb{R}^2/U.$$

Die Bilder der Standardbasis sind also in $\mathbb{R}^2/\operatorname{Span}(\begin{pmatrix} 1 \\ 1 \end{pmatrix})$ linear abhängig.

Wir wissen auch, dass $[\mathbf{e}_1] \neq [0], [\mathbf{e}_2] \neq [0] \in \mathbb{R}^2/U$ da keiner der Standardbasisvektoren in U liegt. Da $\dim \mathbb{R}^2/U = 1$ ist, können wir also jeden der beiden Vektoren als Basis von $\mathbb{R}^2/U$ verwenden.



$\mathbb{R}^2$ mit Unterraum U, die Diagonalen werden in $\mathbb{R}^2/U$ zu jeweils einem Element identifiziert („zusammengedrückt").

Wir können uns den Quotienten $\mathbb{R}^2/U$ hier gut geometrisch vorstellen:Die Äquivalenzrelation identifiziert Vektoren $\mathbf{v}_1, \mathbf{v}_2$ in $\mathbb{R}^2$, wenn sich diese um ein Element der Diagonalen unterscheiden, d.h. die Parallelen zur Diagonalen werden jeweils zu einem Punkt zusammengedrückt.

Daran sehen wir auch, warum sowohl $[\mathbf{e}_1]$ als auch $[\mathbf{e}_2]$ als Basisvektor für den Quotienten genutzt werden können, denn die jeweiligen Koordinatenachsen schneiden alle Parallelen zur Diagonalen in genau einem Punkt.

Damit können wir jetzt wie versprochen Abbildungen injektiv machen. Das geht einfach so, dass wir alle Elemente miteinander identifizieren, die sich nur um ein Element des Kerns unterscheiden, geometrisch bedeutet das, einfach die Richtungen die im Unterraum $\mathrm{Ker}(F)$ einer Abbildung liegen zu ignorieren. Formal können wir das so formulieren:

Behauptung 60 (Charakterisierende Eigenschaft von Quotienten)**.** *Sei V ein K-Vektorraum und $U \subseteq V$ ein Unterraum dann gilt:*

1. *Ist $F\colon V \to W$ eine lineare Abbildung, so ist die Abbildung*

$$\begin{aligned} \overline{F}\colon V/U &\to W \\ [\mathbf{v}] &\mapsto \overline{F}([\mathbf{v}]) := F(\mathbf{v}) \end{aligned}$$

$$\begin{tikzcd} V \arrow[r, "F"] \arrow[dr, "q"'] & W \\ & V/U \arrow[u, dotted, "\overline{F}"'] \end{tikzcd}$$

   genau dann eine wohldefinierte lineare Abbildung wenn $U \subseteq \mathrm{Ker}(F)$.

2. *Ist $F\colon V \to W$ eine lineare Abbildung, so ist für $U := \mathrm{Ker}(F)$ die Abbildung $\overline{F}\colon V/\,\mathrm{Ker}(F) \to W$ injektiv.*

Beweis. Der Beweis besteht wieder nur darin, nachzuschauen, dass unsere Definitionen tatsächlich das leisten, was wir vorher in Worten formuliert hatten, d.h. die Definitionen sind gerade so gemacht, dass dieses Resultat stimmt.

Darum möchte ich Ihnen das formale Argument als Übungsaufgabe stellen, denn es hilft Ihnen mehr die Definitionen zu verdauen, wenn Sie die noch einmal nachschauen, als wenn ich das für Sie tue.

In der Vorlesung haben wir 1. nachgeschaut: Um zu zeigen, dass $\overline{F}$ wohldefiniert ist, müssen wir zeigen, dass wenn $[\mathbf{v}] = [\mathbf{v}']$ ist, auch gilt dass $F(\mathbf{v}) = F(\mathbf{v}')$.

Die Formel $[\mathbf{v}] = [\mathbf{v}']$ bedeutet, dass $\mathbf{v} \sim_U \mathbf{v}'$, d.h. es existiert ein $\mathbf{u} \in U$ mit $\mathbf{v}' = \mathbf{v} + \mathbf{u}$. Dann gilt aber

$$F(\mathbf{v}') = F(\mathbf{v}+\mathbf{u}) \overset{F \text{ linear}}{=} F(\mathbf{v}) + F(\mathbf{u}) \overset{!}{=} F(\mathbf{v})$$

die letzte Gleichung gilt also nur wenn $F(\mathbf{u}) = 0$ ist, also $\mathbf{u} \in \mathrm{Ker}(F)$. Also folgt aus $U \subset \mathrm{Ker}(F)$, dass $\overline{F}([\mathbf{v}]) := F(\mathbf{v})$ wohldefiniert ist.

Gilt umgekehrt $U \not\subseteq \mathrm{Ker}(F)$, so existiert $\mathbf{u} \in U$ mit $\mathbf{u} \notin \mathrm{Ker}(F)$, d.h. $F(\mathbf{u}) \neq 0$. Dann gilt $[u] = [0]$ und $F(\mathbf{u}) \neq 0 = F(0)$, also ist $\overline{F}$ dann nicht wohldefiniert.

Das zeigt, dass $\overline{F}$ genau dann wohldefiniert ist, wenn $U \subseteq \mathrm{Ker}(F)$ gilt. □

Fazit: Ist $U \subset V$ ein Untervektorraum und $q\colon V \to V/U$ die Quotientenabbildung, so ist für jeden K-Vektorraum W die Zuordnung

$$\begin{aligned} \{F \in \mathrm{Hom}(V,W) \mid U \subseteq \mathrm{Ker}(F)\} &\to \mathrm{Hom}(V/U, W) \\ F &\mapsto \overline{F} \end{aligned}$$

bijektiv, die Umkehrabbildung ist durch $\overline{F} \mapsto F := \overline{F} \circ q$ gegeben. Wir fassen die Mengen $\mathrm{Hom}(V/U, W)$ und $\{F \in \mathrm{Hom}(V,W) \mid U \subseteq \mathrm{Ker}(F)\}$ darum als gleich auf, um uns zu Erinnern, dass es eine natürliche bijektive Abbildung zwischen den Mengen gibt schreiben wir

$$\operatorname{Hom}(V/U, W) \cong \{F \in \operatorname{Hom}(V, W) \mid U \subseteq \operatorname{Ker}(F)\}.$$

„$\cong$“ = „natürlich isomorph“, wobei „natürlich“ für uns zunächst einfach bedeutet, dass wir uns auf eine bijektive Abbildung zwischen den beiden Seiten geeinigt haben, die wir nicht mehr mit angeben möchten.

Als Anwendung, können wir jetzt besser verstehen, wie wir das Bild der dualen Abbildung einer linearen Abbildung beschreiben können. Ist $F\colon V \to W$ linear, so hatten wir $F^\vee\colon W^\vee \to V^\vee$ definiert als $g \mapsto g \circ F = g(F(\_\_))$. Damit ist das Bild von $F^\vee$

$$\begin{aligned}
\operatorname{Bild}(F^\vee) &= \{f\colon V \to K \mid \substack{f \text{ linear} \\ f = g \circ F \text{ für ein } g \in W^\vee}\} \\
&\subseteq \{f\colon V \to K \mid f \text{ linear und } \operatorname{Ker}(F) \subseteq \operatorname{Ker}(f)\} \\
&\cong \operatorname{Hom}(V/\operatorname{Ker}(F), K) = (V/\operatorname{Ker}(F))^\vee.
\end{aligned}$$

Wir wissen außerdem, dass $\overline{F}\colon V/\operatorname{Ker}(F) \to W$ injektiv ist, d.h., die Abbildung $\overline{F}\colon V/U \to \operatorname{Bild}(F) \subset W$ ist ein Isomorphismus. Ist darum umgekehrt $\overline{f} \in (V/\operatorname{Ker}(F))^\vee$ gegeben, so existiert ein $g \in W^\vee$ mit $\overline{f} = g \circ F$, d.h.

$$\operatorname{Bild}(F^\vee) \cong (V/\operatorname{Ker}(F))^\vee.$$

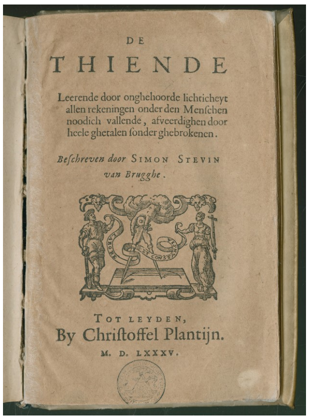
Das klärt den genauen Zusammenhang zwischen $\operatorname{Ker}(F)$ und $\operatorname{Bild}(F^\vee)$.

Die reellen Zahlen als Quotientenvektorraum

Quotienten sind eine sehr gute Möglichkeit, eine exakte Definition der reellen Zahlen zu finden. Auf die Frage, was Sie sich genau unter den reellen Zahlen vorstellen, haben Sie zunächst „alle Zahlen - bis auf die komplexen“ geantwortet, und mir auf Nachfrage dann erklärt, dass es nicht möglich ist, alle reellen Zahlen anzugeben, weil diese unendlich viele Nachkommastellen haben. Wenn Sie diese Zeilen lesen, haben Sie vielleicht mehr Ideen, wie Sie versuchen könnten genauer zu sagen, was reelle Zahlen nun genau sein sollten.

In der linearen Algebra 1 hatten wir einmal versucht, die pragmatische Lösung zu verwenden, dass reelle Zahlen einfach unendliche Dezimalzahlen sind, also formale Ausdrücke der Form

$$n, a_1 a_2 a_3 \ldots \text{ wobei } n \in \mathbb{Z} \text{ und } a_i \in \{0, 1, \ldots, 9\} \text{ für } i \in \mathbb{Z}_{>0}.$$



Dezimalbrüche sind noch nicht so alt, in Europa sind diese 1585 durch *Simon Stevin* eingeführt worden. Sein Buch „De Thiende“ wurde von der Druckerei Plantin Moretus in Antwerpen verlegt, wo Sie es auch anschauen können. Der Buchdruck hat die Schreibweise verbreitet. In der arabischen Welt hatte *al-Uqlidisi* schon um 952 über Dezimalbrüche geschrieben.

Hierbei wissen wir außerdem, dass

$$1 = 0,9999\ldots$$

zum Beispiel weil in $\mathbb{R}$ die Gleichung $3 \cdot \frac{1}{3} = 1$ gelten sollte, oder weil der Abstand

$$1 - 0,\underbrace{9999\ldots9}_{N-\text{Ziffern}} = \frac{1}{10^N}$$

kleiner als jede positive Zahl, also 0 ist. Also müssen wir mit diesem Ansatz, die reellen Zahlen zu konstruieren, sicher Zahlen die auf $\overline{9}$ enden

Es ist eine gute Idee, weitere Argumente für die Gleichheit zu suchen.

$$n, a_1 a_2 a_3 \ldots a_N \overline{9}$$

wobei $a_N \neq 9$ mit
$$n, a_1a_2a_3 \dots (a_N + 1)$$
identifizieren. Wir notieren diese Äquivalenzrelation als $\sim_{0.\bar{9}=1}$. Das liefert

$$\mathbb{R}_{\text{dez}} := \{n, a_1a_2a_3 \cdots \mid n \in \mathbb{Z}, a_i \in \{0, 1, \dots, 9\} \text{ für } i \in \mathbb{Z}_{>0}\} / \sim_{0.\bar{9}=1} .$$

Es ist aber nicht klar wie wir reelle Zahlen miteinander addieren, multiplizieren oder gar dividieren können, denn die Verfahren für ganze Zahlen, oder für endliche Dezimalbrüche fangen wegen möglicher Überträge immer „hinteren Ende“ an $r = n, a_1a_2a_3 \dots$ und $s = n', b_1b_2b_3 \dots$ addieren indem wir die endlichen Dezimalbrüche $n, a_1a_2a_3 \dots a_N$ und $n', b_1b_2b_3 \dots b_N$ addieren (oder multiplizieren), dann passen die Resultate wegen der Überträge nicht immer zusammen.

Um das zu klären sollten wir uns zunächst überlegen, was wir mit dem unendlichen Dezimalbruch meinen. Endliche Dezimalbrüche sind rationale Zahlen:
$$n, a_1a_2a_3 \dots a_N = n + \frac{a_1}{10} + \frac{a_2}{100} + \frac{a_3}{1000} + \cdots + \frac{a_N}{10^N}.$$
Betrachten wir für einen unendlichen Dezimalbruch die Folge
$$(q_N)_{N \in \mathbb{N}_{>0}} \text{ der Zahlen } q_N = n + \sum_{i=1}^{N} \frac{a_i}{10^i}$$
so unterscheiden sich die Zahlen für $M > N$ nur noch um weniger als $1/10^N$, da dieser Unterschied kleiner als jede positive Zahl wird, stellen wir uns die zugehörige reelle Größe als Grenzwert dieser Folge von Zahlen vor.

Das klärt aber noch immer nicht wie wir mit Zahlen dieser Form rechnen können, denn wenn wir $(q_N)_{N \in \mathbb{N}_{>0}}$ und $(p_N)_{N \in \mathbb{N}_{>0}}$ termweise multiplizieren oder addieren, passen die Ergebnisse noch immer nicht schön als Dezimalbrüche zusammen.

An dieser Stelle gilt es nicht in Panik zu verfallen, sondern ruhig nachzudenken, woher das Problem kommt. Das ist eigentlich nur entstanden, weil wir uns für endliche Dezimalbrüche entschieden haben, mindestens für die Wahl der 10 in den Formeln gab es keinen wirklichen Grund, Lebewesen mit mehr oder weniger als 10 Fingern sollten die gleichen reellen Zahlen finden.

Ein Grundprinzip an dieser Stelle ist vielleicht „Wenn uns keine natürliche Wahl einfällt, können wir versuchen alle möglichen Wahlen zu erlauben“. Wir könnten also einfach alle Folgen $(q_m)_{m \in \mathbb{N}_{>0}}$ von rationalen Zahlen erlauben, die sich schließlich (also für genügend großen Index) um weniger als jede positive Zahl unterscheiden, also zum Beispiel für alle N um weniger als $1/10^N$. Formal können wir das so aufschreiben:

$$CF_{\mathbb{Q}} := \left\{ (q_m)_{m\in\mathbb{N}_{>0}} \;\middle|\; \begin{array}{c} q_m \in \mathbb{Q} \text{ für alle } m \\ \text{Für alle } N \in \mathbb{N} \text{ existiert } m_0 \in \mathbb{N} \text{ so dass} \\ |q_m - q_n| < \frac{1}{10^N} \text{ für alle } m, n > m_0 \end{array} \right\}.$$

Die Formulierung „schließlich" ist in der Formel leider etwas sperrig. Schauen Sie sich bitte einmal in Ruhe an, wieso die Formel einfach „unterscheiden sich schließlich um weniger als jede positive Zahl" mathematisch genau fasst. Folgen, die diese Bedingung erfüllen heißen Cauchy-Folgen, daher die Bezeichnung $CF_{\mathbb{Q}}$.

Bemerkung. 1. Die Menge aller Folgen

$$\text{Folgen}_{\mathbb{Q}} := \{(q_m)_{m\in\mathbb{N}_{>0}} \mid q_m \in \mathbb{Q}\}$$

ist ein $\mathbb{Q}$-Vektorraum, denn wir können Folgen wie Vektoren komponentenweise addieren und mit Skalaren multiplizieren.

2. Es ist nicht schwer zu sehen, dass $CF_{\mathbb{Q}} \subseteq \text{Folgen}_{\mathbb{Q}}$ ein Untervektorraum ist (wenn Sie zum Beispiel zwei Folgen $(p_m), (q_m)$ zu $(p_m + q_m)$ addieren und sich p_m, q_m für $m > m_0$ jeweils um weniger als $1/10^{N+1}$ unterscheiden, dann unterscheiden sich die Folgenglieder von $p_m + q_m$ um weniger als $2/10^{n+1}$ und das ist kleiner als $1/10^N$).

Wann sollten zwei Folgen $(p_m), (q_m)$ die gleiche reelle Zahl darstellen? Wenn die Differenz $p_m - q_m$ schließlich kleiner als jede positive Zahl wird. Das liefert uns aber eine Äquivalenzrelation die von einem Unterraum herkommt! Die Teilmengen der Folgen, die 0 darstellen sind mit dieser Idee nämlich

$$NF_{\mathbb{Q}} := \left\{ (q_m)_{m\in\mathbb{N}_{>0}} \;\middle|\; \begin{array}{c} q_m \in \mathbb{Q} \text{ für alle } m \\ \text{Für alle } N \in \mathbb{N} \text{ existiert } m_0 \in \mathbb{N} \text{ so dass} \\ |q_m| < \frac{1}{10^N} \text{ für alle } m > m_0 \end{array} \right\}$$

und das ist ein Untervektorraum von CF (hier steht NF für Nullfolgen). Wir können also den Quotientenraum

$$CF_{\mathbb{Q}}/NF_{\mathbb{Q}} =: \mathbb{R}_{CF}$$

als Definition der reellen Zahlen verwenden, denn die Äquivalenzklassen $[(q_m)]$ sind genau die Folgen, die gegen den gleichen Grenzwert konvergieren sollen. Da wir uns schon überlegt haben, dass der Quotient ein Vektorraum ist, haben wir hier schon einmal eine Addition $+$ und eine Multiplikation mit Elementen von $\mathbb{Q}$. In der Analysis überlegen Sie sich noch, dass auch $\cdot$ wohldefiniert ist, das ist ebenfalls nicht so schwer. Wie bei den rationalen Zahlen und den Vektorraumaxiomen für Quotientenvektorräume, ist es – wenn Sie gezeigt haben, dass $[(q_m)] \cdot [(p_m)] := [(q_m \cdot p_m)]$ wohldefiniert ist – einfach, die Körperaxiome für $\mathbb{R}_{CF}$ zu zeigen, denn die Rechenregeln die für die Folgenglieder gelten, übertragen sich dann automatisch auf die Äquivalenzklassen $[(q_m)]$.

Um mit der Definition zufrieden zu sein, sollten wir uns aber noch überlegen, dass die Abbildung $\mathbb{R}_{\text{dez}} \xrightarrow{\cong} \mathbb{R}_{CF}$ die einen Dezimalbruch $(n, a_1 a_2 \dots)$ auf die zugehörige Cauchyfolge $(q_m = n, a_1 \dots a_m)_{m \geq 1}$ abbildet, eine bijektive Abbildung definiert, d.h. tatsächlich jede Äquivalenzklasse $[(q_m)] \in \mathbb{R} = CF_{\mathbb{Q}}/NF_{\mathbb{Q}}$ eine Folge enthält, die zu einer unendlichen Dezimalzahl

$$n, a_1 a_2 \dots$$

gehört. Wenn wir das wissen, dann haben wir gezeigt, dass wir unendliche Dezimalbrüche doch addieren und multiplizieren können, indem wir die endliche Brüche addieren (bzw. multiplizieren), das Ergebnis als Cauchyfolge auffassen und dann wieder als Dezimalbruch umrechnen.

Für diejenigen unter Ihnen, die die Analysis besuchen, ist es eine gute Übung, sich zu überlegen, dass jede Äquivalenzklass $[(q_m)] \in \mathbb{R}_{CF}$ einen unendlichen Dezimalbruch enthält, die anderen, können das als Vorbereitung für die Analysis nutzen in der folgenden Aufgabe gebe ich eine kurze Anleitung, wie Sie das versuchen könnten.

Aufgabe 2 (Freiwillige Aufgabe)**.** Gegeben eine Cauchyfolge $(q_m)_{m \in \mathbb{Z}_{>0}} \in CF_{\mathbb{Q}}$. Ziel der Aufgabe ist zu zeigen, dass es einen unendlichen Dezimalbruch $n, a_1 a_2 \dots$ gibt, so dass $[(q_m)] = [(n, a_1 \dots, a_m)] \in \mathbb{R}_{CF}$.

1. Der erste Versuch ist vielleicht, für jeden der Brüche $q_m \in \mathbb{Q}$ durch einen Dezimalbruch $d_m = n + \sum_{i=1}^{m} \frac{a_i}{10^i}$ mit $|q_m - d_m| < \frac{1}{10^m}$ zu ersetzen. Geben Sie ein Beispiel einer Cauchyfolge (q_m) an, für die die diese Folge d_m nicht durch die Folgenglieder eines unendlichen Dezimalbruchs gegeben ist.

2. Finden Sie für die Folge (q_m) zunächst eine Folge $(q'_m) \in CF_{\mathbb{Q}}$ so dass $[(q_m)] = [(q'_m)] \in \mathbb{R}_{CF}$ und für alle $N \in \mathbb{Z}_{>0}$ und $m, k \geq N$ gilt, dass $|q'_m - q'_k| \leq \frac{1}{10^N}$.

3. Finden Sie eine Folge $(q''_m) \in CF_{\mathbb{Q}}$ so dass $[(q'_m)] = [(q''_m)] \in \mathbb{R}_{CF}$ und $q''_1 \leq q''_2 \leq \dots \leq q''_m \leq \dots$ eine monoton steigende Folge ist und für alle $N \in \mathbb{Z}_{>0}$ und $m, k \geq N$ gilt, dass $|q'_m - q'_k| \leq \frac{1}{10^N}$.

4. Finden Sie einen unendlichen Dezimalbruch $n, a_1 a_2 \dots$ so dass $[(q''_m)] = [(n, a_1 \dots, a_m)] \in \mathbb{R}_{CF}$, indem Sie die Konstruktion aus 1. auf die Folge (q''_m) anwenden

Bemerkung. 1. Die Konstruktion der reellen Zahlen als Quotientenvektorraum von Cauchyfolgen nach Nullfolgen hat den weiteren Vorteil, dass der Quotient $\mathbb{R}_{CF} = CF_{\mathbb{Q}}/NF_{\mathbb{Q}}$ die charakterisierende Eigenschaft hat, dass es für jede andere Konstruktion eines Körpers K, der die rationalen Zahlen $\mathbb{Q}$ enthält, in dem jede Cauchyfolge einen Grenzwert besitzt und der Grenzwert für Nullfolgen 0 ist, eine natürliche Abbildung $\mathbb{R}_{CF} \to K$ gibt, die mit $+$ und $\cdot$ verträglich ist.

2. Es gibt andere natürliche Möglichkeiten, einen Abstand zwischen rationalen Zahlen zu definieren, die mit den Rechenoperationen $+, \cdot$ verträglich sind. Wenden wir die obige Konstruktion auf Cauchyfolgen bezüglich dieser Abstände an, bekommen wir andere interessante Zahlbereiche, die in der Zahlentheorie nützlich sind. Zm Beispiel gibt es für eine jede Primzahl p einen Abstand $\|\|_p$ auf den ganzen Zahlen $\mathbb{Z}$, für den ganze Zahlen n, m nah beieinander liegen wenn die Differenz $n - m$ durch eine hohe Potenz von p teilbar ist, formal können wir $\|n - m\|_p := \frac{1}{p^N}$ definieren, wobei N die größte ganze Zahl ist, für die p^N die Zahl $n - m$ teilt.

Anwendung für die Elementargeometrie

Vielleicht haben Sie in der Schule einmal etwas über die Geometrie der Ebene gelernt. Sie wissen sicher, dass durch je zwei verschiedene Punkte der Ebene genau eine Gerade verläuft und sich umgekehrt je zwei Geraden sich entweder in genau einem Punkt schneiden oder parallel sind. Ganz ähnlich gibt es für alle Resultate in der ebenen Geometrie, die aus einer Konstruktion folgern, dass eine Konstruktion eine Reihe von Punkten liefert, die auf einer Geraden liegen, ein zweites Resultat, das besagt, dass eine andere Konstruktion Geraden liefert, die sich in einem Punkt schneiden.

Mit Quotienten und dem Dualraum können wir erklären, warum das so ist. Für jeden Untervektorraum $U \subseteq V$ eines K-Vektorraums V bezeichnen wir die Inklusionsabbildung von U in V mit i und die Quotientenabblidung mit q, d.h. wir bekommen Abbildungen

$$U \overset{i}{\hookrightarrow} V \overset{q}{\twoheadrightarrow} V/U$$

und es gilt $U = \operatorname{Ker}(q)$, insbesondere ist $q \circ i = 0$ nach Konstruktion von V/U die Nullabbildung. Ist U eindimensional, also $\dim U = 1$ so ist $\dim V/U = \dim V - 1$. Für die dualen Abbildungen

$$U^\vee \overset{i^\vee}{\twoheadleftarrow} V^\vee \overset{q^\vee}{\hookleftarrow} (V/U)^\vee$$

ist $q^\vee$ injektiv (wir hatten gesehen, dass die Elemente von $(V/U)^\vee$ genau den Abbildungen in $V^\vee$ entsprechen, die auf U verschwinden), die Abbildung $i^\vee$ ist surjektiv und entspricht der Einschränkung von Abbildungen $V \to K$ auf U. Also gilt wieder $\operatorname{Ker}(i^\vee) = \operatorname{Bild}(q^\vee) \cong (V/U)^\vee$.

Jeder m-dimensionale Unterraum U von V definiert also einen $n - m$-dimensionalen Unterraum

$$U^\top \cong (V/U)^\vee \subseteq V^\vee.$$

Ist umgekehrt $W \subset V^\vee$ ein $n - m$-dimensionaler Unterraum von $V^\vee$ so ist

$$U := \{\mathbf{v} \in V \mid f(\mathbf{v}) = 0 \forall f \in W\}$$

ein m-dimensionaler Unterraum, weil die Bedingung $f(\mathbf{v}) = 0 \forall f \in W$ äquivalent zur Bedingung $f(\mathbf{v}) = 0$ für alle Elemente einer Basis von W ist und die lineare Unabhängigkeit von $f_1, \dots, f_{n-m}$ in $V^\vee$ bedeutet, dass die Gleichungen $f_i(\mathbf{v}) = 0$ linear unabhängig sind.

Also entsprechen m-dimensionale Unterräume von V genau $n-m$-dimensionalen Unterräumen von $V^\vee$, d.h. die Abbildungen

$$\left\{ U \mid \begin{matrix} U \subseteq V \text{ Unterraum} \\ \dim U = m \end{matrix} \right\} \leftrightarrows \left\{ W \mid \begin{matrix} W \subseteq V^\vee \text{Unterraum} \\ \dim W = n-m \end{matrix} \right\}$$

$$U \mapsto U^\top = (V/U)^\vee$$

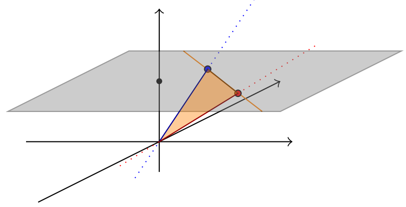
$$\{\mathbf{v} \in V \mid f(\mathbf{v}) = 0 \forall f \in W\} =: W^\top \hookleftarrow W$$

sind zueinander inverse Zuordnungen.

Nach Definition gilt außerdem, dass diese Abbildung den Durchschnitt $\cap$ und die Summe $+$ von Unterräumen vertauscht, d.h. $(U_1 \cap U_2)^\top = U_1^\top + U_2^\top$ und umgekehrt ist $(U_1 + U_2)^\top = U_1^\top \cap U_2^\top$.

Was hat das mit ebener Geometrie zu tun? Eine Gerade in der Ebene muss im Gegensatz zu einem 1-dimensionalen Unterraum nicht notwendig den Ursprung 0 enthalten. Das scheint zunächst ein Problem für die lineare Algebra zu sein, aber es gibt einen wunderbaren Trick, wie wir dieses Problem loswerden können: Betten wir nämlich die Ebene $\mathbb{R}^2$ als Ebene durch den Punkt $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ in den $\mathbb{R}^3$ ein, also

$$E := \left\{ \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \in \mathbb{R}^3 \mid x, y \in \mathbb{R} \right\} \subset \mathbb{R}^3,$$



$\mathbb{R}^3$ mit Ebene der Höhe 1. Punkte in E entsprechen Geraden, Geraden in E entsprechen Ebenen in $\mathbb{R}^3$.

entsprechen mittels den Zuordnungen

$$M \subset E \mapsto \mathrm{Span}(W) \subseteq \mathbb{R}^3 \qquad \text{bzw.}$$
$$U \subseteq \mathbb{R}^3 \mapsto U \cap E$$

Punkte in E genau 1-dimensonalen Unterräumen von $\mathbb{R}^3$, die nicht in der x, y-Ebene liegen und

Geraden in E den 2-dimensionalen Unterräumen von $\mathbb{R}^3$, die nicht gleich der x, y-Ebene sind.

Damit können wir Aussagen über Schnitte von Geraden in E in Aussagen über Schnitte von 2-dimensionalen Unterräumen in $\mathbb{R}^3$ übersetzen, und $P \in L$ entspricht dann $\mathrm{Span}(P) \subset \mathrm{Span}(L)$.

Eine Gerade in $\mathbb{R}^3$ entspricht aber einem $2-$dimensionalen Unterraum im Dualraum $(\mathbb{R}^3)^\vee$ und ein 2-dimensionaler Unterraum $\mathrm{Span}(L)$ entspricht einer Geraden in $(\mathbb{R}^3)^\vee$. Im Dualraum vertauschen sich also die Rollen von Geraden und Ebenen. Da $(\mathbb{R}^3)^\vee \cong \mathbb{R}^3$ ist, erklärt das, wieso es in der ebenen Geometrie zu jeder richtigen Aussage auch eine „duale Aussage“ gilt, in der die Rollen von Punkten und Geraden vertauscht werden.

Als Bonus erklärt diese Konstruktion auch, dass wir die Fallunterscheidung dass sich zwei Geraden in einem Punkt schneiden oder parallel sind, loswerden können, wenn wir statt 1-dimensonalen

Unterräumen von $\mathbb{R}^3$, die nicht in der xy–Ebene liegen einfach alle 1-dimensionalen Unterräume als Punkte zulassen würden. Die Menge der 1-dimensionalen Unterräume von $\mathbb{R}^3$:

$$\mathbb{P}^2(\mathbb{R}) := \mathbb{R}^3 \setminus \{0\} / \sim \ \text{ wobei } \begin{array}{c} \mathbf{v} \sim c \cdot \mathbf{v}\, \forall c \neq 0 \in \mathbb{R}, \\ \forall \mathbf{v} \in \mathbb{R}^3 \setminus \{0\} \end{array}$$

heißt projektiver Raum und wird Ihnen hoffentlich noch öfters begegnen.

Als Menge enthält $\mathbb{P}^2(\mathbb{R})$ für jeden Punkt der Ebene $E \subset \mathbb{R}^3$ genau eine Äquivalenzklasse und zusätzlich für jeden 1-dimensionalen Unterraum der xy–Ebene $\mathbb{R}^2$ einen Punkt, diese zusätzlichen Punkte werden meist „Punkte im Unendlichen“ genannt.

Aus der Perspektive der Punkte von Geraden in E ist das eine naheliegende Bezeichnung, aber aus der Perspektive der linearen Algebra ist das etwas merkwürdig, denn die Geraden in $\mathbb{R}^3$ die in der xy–Ebene liegen sind natürlich genauso gut wie die anderen Geraden im $\mathbb{R}^3$, ausgezeichnet werden diese erst durch die Wahl der Koordinaten.

Tensorprodukte – Rechnen mit Buchstaben und Symbolen Teil 2

Die letzte Konstruktion von neuen Vektorräumen aus bekannten ist das Tensorprodukt. Wir haben K^n als Raum von Spaltenvektoren eingeführt, genauso bilden Matrizen $\mathrm{Mat}_{n,m}(K)$ einen Vektorraum, dessen Elemente wir entweder konkret als $n \times m$-Liste von Zahlen, oder als Raum von Abbildungen $A \colon \mathbb{R}^m \to \mathbb{R}^n$ auffassen können. Sie können sich sicher Vorstellen, dass Daten hin und wieder als mehrdimensionale Listen von $n_1 \times \cdots \times n_d$ Zahlen auftauchen. Wie bei linearen Abbildungen, ist es nützlich zwischen der konkreten Beschreibung ($n \times m$-Liste von Zahlen) und einer abstrakten Beschreibung (lineare Abbildung) wechseln zu können, weil eine geschickte Wahl einer Basis aus einer komplizierten, eine einfache Matrix machen kann.

Beispiel: Einfache Darstellung linearer Abbildungen

Beispiel 61. Wie auf dem Übungsblatt können wir uns überlegen, dass jede lineare Abbildung $A \colon V \to W$ vom Rang 1 für geeignete $f_1 \in V^\vee, \mathbf{w}_1 \in W$ in der Form

$$A\mathbf{v} = f_1(\mathbf{v}) \cdot \mathbf{w}_1$$

schreiben können. Dieser Spezialfall ist nicht schwer zu sehen: A definiert eine Abbildung

$$\overline{A} \colon V / \operatorname{Ker}(A) \to \operatorname{Bild}(A) = \operatorname{Span}(\mathbf{w}_1) \subset \mathbb{R}^n$$

und nach Konstruktion ist $V / \operatorname{Ker}(A) \to \operatorname{Bild}(A)$ ein Isomorphismus von 1-dimensionalen Vektorräumen. Also ist $A\mathbf{v} = f_1(\mathbf{v})\mathbf{w}_1$, wobei wir mit der Koordinatenabbildung $\mathrm{Koord}_{\mathbf{w}_1} \colon \operatorname{Bild}(A) \to K$ die Abbildung f_1 als $f_1 = \mathrm{Koord}_{\mathbf{w}_1}(A\_\_)$ schreiben können. Insbesondere ist f_1 also linear.

Auf dem Übungsblatt überlegen Sie sich weiter, dass wir jede lineare Abbildung vom Rang r als Linearkombination

$$A = \sum_{i=1}^{r} f_i \cdot \mathbf{w}_i$$

für gewisse $f_i \in V^\vee$ und $\mathbf{w}_i \in \mathrm{Bild}(A)$ schreiben können.

Es lassen sich also alle linearen Abbildungen zwischen endlichdimensionalen Vektorräumen als Linearkombinationen von formalen Produkten aus Elementen aus $V^\vee$ und W schreiben. Für die Ausdrückeder Form $f \cdot \mathbf{w}$ gelten die bilinearen Rechenregeln

$$\begin{aligned}(f_1 + f_2) \cdot \mathbf{w} &= f_1 \cdot \mathbf{w} + f_2 \cdot \mathbf{w}, \\ f \cdot (\mathbf{w}_1 + \mathbf{w}_2) &= f \cdot \mathbf{w}_1 + f \cdot \mathbf{w}_2 \text{ und} \\ (cf) \cdot \mathbf{w} &= c(f \cdot \mathbf{w}) = f \cdot (c\mathbf{w})\end{aligned}$$

für alle $f, f_1, f_2 \in V^\vee, \mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W, c \in K$.

FRAGE: Sind das alle Rechenregeln für diese Ausdrücke? Versuchen Sie einmal sich selbst klar zu machen, was mit dieser Frage genau gemeint ist.

FÜR MICH BEDEUTET DAS: Wenn ich die Menge aller formalen Linearkombinationen von Symbolen der Form $f \cdot \mathbf{w}$ mit $f \in V^\vee, \mathbf{w} \in W$ betrachte und je zwei Ausdrücke als gleich betrachte, die ich mit den Rechenregeln ineinander überführen kann, erhalte ich dann eine neue Beschreibung für $\mathrm{Hom}(V, W)$?

Quotientenräume erlauben uns einerseits, diese Konstruktion formal exakt zu formulieren und andererseits, zu beweisen, dass wir tatsächlich alle Rechenregeln gefunden haben. Das ist ein ganz schönes Beispiel dafür, dass wir einen Überblick darüber bekommen können, welche verschiedenen Formeln die gleichen Abbildungen definieren.

Konstruktion von $V \otimes W$

ALLGEMEINES PROBLEM:

1. Finde eine Konstruktion eines Vektorraums, dessen Elemente formale Linearkombinationen von Elementen einer gegebenen Menge M sind.

2. Finde eine Variante dieser Konstruktion in der die Symbole zusätzlich eine Liste L zuvor gewünschter Rechenregeln erfüllen.

3. Finde eine Methode, das Ergebnis so zu beschrieben, dass wir Basen des konstruierten Raums finden können, um eine Übersicht zu bekommen, wie groß das Ergebnis geworden ist.

LASSEN SIE UNS VORNE BEGINNEN: Wir suchen für die Menge $\{(f, \mathbf{w}) \mid f \in V^\vee, \mathbf{w} \in W\}$ einen Vektorraum dessen Elemente genau die Linearkombinationen der Ausdrücke sind – im Beispiel denken wir an $f \cdot m$, ich schreibe hier $(f, \mathbf{w})$ um zwischen dem Symbol $(f, \mathbf{w})$ und der linearen Abbildung $f \cdot m$ unterscheiden zu können.

Um uns die Schreibarbeit zu erleichtern, ist es praktisch, einfach für jede Menge M den Vektorraum der formalen Linearkombinationen von Elementen von W zu definieren. Lassen Sie uns die Gelegenheit für einen kleinen Umweg nutzen, der auch in anderen Zusammenhängen nützlich ist.

Definition. Ist M eine Menge, und K ein Körper, dann bezeichnen wir mit

$$\bigoplus_{m\in M} K := \{(a_m)_{m\in M} \mid a_m \in K, \text{ nur endliche viele } a_m \neq 0\}$$

die formale direkte Summe. Die Menge $\bigoplus_{m\in M} K$ zusammen mit komponentenweiser Addition $+$ und skalarer Multiplikation $c \cdot (a_m)_{m\in M} := (ca_m)_{m\in M}$ ein K-Vektorraum.

Bemerkung. Eine Verallgemeinerung dieser Definition ist manchmal ebenfalls nützlich: Ist I eine Menge und für jedes $i \in I$ ein Vektorraum V_i gegeben, so bezeichnen wir mit

$$\bigoplus_{i\in I} V_i := \{(\mathbf{v}_i)_{i\in I} \mid \mathbf{v}_i \in V_i, \text{ nur endliche viele } \mathbf{v}_i \neq 0\}$$

die formale direkte Summe. Die Menge $\bigoplus_{i\in I} V_i$ zusammen mit komponentenweiser Addition $+$ und skalarer Multiplikation $c \cdot (\mathbf{v}_i)_{i\in I} := (c\mathbf{v}_i)_{i\in I}$ ein K-Vektorraum.

Genauso wird das direkte Produkt als

$$\prod_{i\in I} V_i\{(\mathbf{v}_i)_{i\in I} \mid \mathbf{v}_i \in V_i\}$$

definiert.

Bei dieser Gelegenheit sollten wir einmal festlegen, was wir bei einem nicht endlich dimensionalem Vektorraum mit einer Basis meinen:

Definition. Sei V ein K-Vektorraum. Eine Teilmenge $(\mathbf{v}_i)_{i\in I} \in V$ heißt

1. *linear unabhängig*, wenn für jede endliche Teilmenge $J \subseteq I$ die Gleichung $\sum_{j\in J} a_j\mathbf{v}_j = 0$ nur die triviale Lösung $a_j = 0$ für alle $j \in J$ hat.

2. *Erzeugendensystem von V* wenn es für jeden $v \in V$ eine endliche Teilmenge $J \subseteq I$ und Elemente $a_j \in K$ gibt, so dass $\mathbf{v} = \sum_{j\in J} a_j\mathbf{v}_j$ eine Linearkombination der $\mathbf{v}_j$ ist.

3. *Basis von V* wenn die Vektoren ein linear unabhängiges Erzeugendensystem von V sind.

Die Definition beinhaltet einen kleinen formalen Trick: Sie hätten „Basis“ wahrscheinlich lieber als „jeder Vektor lässt sich eindeutig als Linearkombination der Basisvektoren schreiben“ formuliert.

Diese Formulierung hat das technische Problem, dass wir nur endliche Linearkombinationen sinnvoll definieren können. Darum müssen wir in dieser Formulierung eine endliche Menge $J \subset M$ auswählen. Dann müssen wir aufpassen was wir mit „eindeutig“ meinen, denn wir können immer $0 \cdot \mathbf{v}_m$ hinzufügen.

Diese lästige Spitzfindigkeit umgeht die Definition recht elegant.

Beispiel 62. Ist I eine Menge und bezeichnen wir für jedes $j \in I$ mit $\mathbf{e}_j \in \bigoplus_{i\in I} K$ den Standardbasisvektor, der an der j-ten Koordinate den Eintrag 1 und an allen Koordinaten $i \neq j$ den Eintrag 0 hat. Dann sind die Vektoren $(\mathbf{e}_j)_{j\in I}$ eine Basis von $\bigoplus_{i\in I} K$.

Notation 63. Im Vektorraum $\bigoplus_{i\in I} K$ kürzen wir den Standardbasisvektor $\mathbf{e}_j$ oft mit $\mathbf{j}$ ab. Damit sind die Elemente von $\bigoplus_{i\in I} K$ dann genau die formalen endlichen Linearkombinationen $\sum_{j\in J} a_j \cdot \mathbf{j}$.

Damit haben wir Schritt 1. unserer Konstruktion erklärt: Wir haben jetzt einen Vektorraum, dessen Vektoren genau die endlichen formalen Linearkombinationen von Elementen einer gegebenen Menge sind, d.h. die Symbole der Menge M sind eine Basis dieses Vektorraums.

In unserem Beispiel, eine neue Beschreibung von $\mathrm{Hom}(V, W)$ als formale Summen zu finden würden wir also

$$\bigoplus_{(f,\mathbf{w})\in V^\vee \times W} K$$

betrachten, was geometrisch ein unheimlich großer Vektorraum ist, algebraisch ist der Vektorraum aber harmlos, es sind einfach endliche formale Linearkombinationen, also genau die Ausdrücke mit denen wir im Beispiel rechnen.

In Schritt 2. wollen wir auf dieser Menge die Rechenregeln

$$\begin{aligned}
(f_1 + f_2) \cdot \mathbf{w} &\overset{!}{=} f_1 \cdot \mathbf{w} + f_2 \cdot \mathbf{w}, \\
f \cdot (\mathbf{w}_1 + \mathbf{w}_2) &\overset{!}{=} f \cdot \mathbf{w}_1 + f \cdot \mathbf{w}_2 \text{ und} \\
(cf) \cdot \mathbf{w} &\overset{!}{=} c(f \cdot \mathbf{w}) = f \cdot (c\mathbf{w})
\end{aligned}$$

für alle $f, f_1, f_2 \in V^\vee, \mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W, c \in K$ einführen, d.h. wir hätten gerne, dass

Merken Sie sich den Trick:
Das Umschreiben von $A = B$ zu $A - B = 0$ erlaubt uns die Äquivalenzrelation als Quotienten nach einem Unterraum zu schreiben.

$$\begin{aligned}
(f_1 + f_2) \cdot \mathbf{w} - f_1 \cdot \mathbf{w} - f_2 \cdot \mathbf{w} &\overset{!}{=} 0 \\
f \cdot (\mathbf{w}_1 + \mathbf{w}_2) - f \cdot \mathbf{w}_1 - f \cdot \mathbf{w}_2 &\overset{!}{=} 0 \text{ und} \\
(cf) \cdot \mathbf{w} - c(f \cdot \mathbf{w}) &\overset{!}{=} 0 \overset{!}{=} f \cdot (c\mathbf{w}) - c(f \cdot \mathbf{w}).
\end{aligned}$$

Diese Gleichungen können wir in einem Quotientenraum erzwingen: Sei

$$\mathrm{Bilinrel} \subseteq \bigoplus_{(f,\mathbf{w})\in V^\vee \times W} K$$

Der Unterraum, der von den Elementen der Form

$$\begin{gathered}
\big((f_1 + f_2) \cdot \mathbf{w} - f_1 \cdot \mathbf{w} - f_2 \cdot \mathbf{w}\big), \\
\big(f \cdot (\mathbf{w}_1 + \mathbf{w}_2) - f \cdot \mathbf{w}_1 - f \cdot \mathbf{w}_2\big), \\
\big((cf) \cdot \mathbf{w} - c(f \cdot \mathbf{w})\big) \text{ und } \big(f \cdot (c\mathbf{w}) - c(f \cdot \mathbf{w})\big)
\end{gathered}$$

aufgespannt wird. Im Quotienten

$$V^\vee \otimes W := \bigoplus_{(f,\mathbf{w})\in V^\vee \times W} K / \mathrm{Bilinrel}$$

sind dann die Symbole $f \otimes \mathbf{w} := [(f, \mathbf{w})]$ ein Erzeugendensystem und nach Konstruktion gelten für diese Symbole die gewünschten Rechenregeln.

DAS IST EINE GUTE ART über Quotienten V/U nachzudenken, die Elemente sind durch die Elemente von V gegeben, wir führen aber die neue Rechenregel ein, dass wir $\mathbf{v}$ durch $\mathbf{v}'$ ersetzen dürfen, wenn die Differenz in U liegt.

BEVOR WIR uns überlegen, dass diese Konstruktion nun tatsächlich einen Vektorraum liefert, der zum Raum der linearen Abbildungen $\mathrm{Hom}(V, W)$ isomorph ist, lassen Sie uns die Konstruktion noch einmal allgemein formulieren:

Definition (Tensorprodukt von Vektorräumen)**.** Sind V, W zwei K-Vektorräume, so ist das Tensorprodukt $V \otimes W$ (oder $V \otimes_K W$) definiert als

Wir schreiben $V \otimes_K W$, wenn es nicht klar ist, welchen Körper wir gerade verwenden, sonst lassen wir das K weg.

$$V \otimes W := \bigoplus_{(\mathbf{v},\mathbf{w}) \in V \times W} K/\text{Bilinrel}$$

wobei Bilinrel der Unterraum von $\bigoplus_{(\mathbf{v},\mathbf{w}) \in V \times W} K$ ist, der von den Elementen der Form

$$(\mathbf{v}_1 + \mathbf{v}_2), \mathbf{w}) - (\mathbf{v}_1, \mathbf{w}) - (\mathbf{v}_2, \mathbf{w}),$$
$$(\mathbf{v}, (\mathbf{w}_1 + \mathbf{w}_2)) - (\mathbf{v}, \mathbf{w}_1) - (\mathbf{v}, \mathbf{w}_2),$$
$$((c\mathbf{v}), \mathbf{w}) - c(\mathbf{v}, \mathbf{w}) \text{ und } (\mathbf{v}, (c\mathbf{w})) - c(\mathbf{v}, \mathbf{w})$$

aufgespannt wird. Wir schreiben

$$\mathbf{v} \otimes \mathbf{w} := [(\mathbf{v}, \mathbf{w})]$$

für die Äquivalenzklasse des Elementes $(\mathbf{v}, \mathbf{w})$.

Fazit. Die Elemente von $V \otimes W$ sind also formale Linearkombinationen von Ausdrücken der Form $\mathbf{v} \otimes \mathbf{w}$, ein typisches Element von $V \otimes W$ ist also $\mathbf{v}_1 \otimes \mathbf{w}_1 + \cdots + \mathbf{v}_r \otimes \mathbf{w}_r$. Und für diese Ausdrücke gelten die Rechenregeln

$$\mathbf{v} \otimes \mathbf{w}_1 + \mathbf{v} \otimes \mathbf{w}_2 = \mathbf{v} \otimes (\mathbf{w}_1 + \mathbf{w}_2)$$
$$\mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w} = (\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w}$$
$$c \cdot (\mathbf{v} \otimes \mathbf{w}) = (c\mathbf{v}) \otimes \mathbf{w} = \mathbf{v} \otimes (c\mathbf{w}),$$

insbesondere können wir skalare Vielfache von einer Seite von $\otimes$ auf die andere schieben. Bei Ausdrücken der Form $(\mathbf{v}_1 \otimes \mathbf{w}_1) + (\mathbf{v}_2 \otimes \mathbf{w}_2)$ fällt uns zunächst nur eine Vereinfachung ein wenn entweder die $\mathbf{v}_i$'s oder die $\mathbf{w}_i$'s linear abhängig sind.

Tut die Definition, was Sie soll?

Als allererstes sollten wir uns überlegen, dass der Quotient aus einem sehr großen, einen überschaubaren Vektorraum macht:

Lemma 64. *Sind V, W zwei K-Vektorräume, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis von V und $\mathbf{w}_1, \ldots, \mathbf{w}_m$ eine Basis von W, dann sind die Vektoren*

$$(\mathbf{v}_i \otimes \mathbf{w}_j)_{i=1\ldots n, j=1,\ldots,m}$$

ein Erzeugendensystem des Tensorproduktes $V \otimes_K W$.

Beweis. Es genügt zu zeigen, dass wir alle Elemente des Erzeugendensystems $\mathbf{v} \otimes \mathbf{w}$ als Linearkombination schreiben können, das ist aber nach Definition klar, denn wir können $\mathbf{v}, \mathbf{w}$ als Linearkombinationen $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i, \mathbf{w} = \sum_{j=1}^m b_j \mathbf{w}_j$ der Basen von V und W schreiben und in $V \otimes W$ gilt

$$\begin{aligned}\mathbf{v} \otimes \mathbf{w} &= (\sum_{i=1}^n a_i \mathbf{v}_i) \otimes (\sum_{j=1}^m b_j \mathbf{w}_j) \\ &= \sum_{i=1}^n a_i (\mathbf{v}_i \otimes (\sum_{j=1}^m b_j \mathbf{w}_j)) \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j (\mathbf{v}_i \otimes \mathbf{w}_j).\end{aligned}$$

□

Interessanter ist die Frage, ob die Elemente $(\mathbf{v}_i \otimes \mathbf{w}_j)_{i=1\ldots n, j=1,\ldots,m}$ linear unabhängig sind. Das sieht auf den ersten Blick schwierig aus, weil wir keinen Überblick über die Elemente aus Bilinrel haben. Die Rettung hierfür ist, die charakterisierende Eigenschaft des Quotienten zu verwenden:

Behauptung 65 (Dimension von $V \otimes W$)**.** *Sind V, W zwei K-Vektorräume, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis von V und $\mathbf{w}_1, \ldots, \mathbf{w}_m$ eine Basis von W, dann sind die Vektoren*

$$(\mathbf{v}_i \otimes \mathbf{w}_j)_{i=1\ldots n, j=1,\ldots,m}$$

eine Basis des Tensorproduktes $V \otimes_K W$, insbesondere ist

$$\dim V \otimes W = \dim V \cdot \dim W.$$

Beweis. Um zu zeigen, dass die Vektoren $(\mathbf{v}_i \otimes \mathbf{w}_j)_{i=1\ldots n, j=1,\ldots,m}$ linear unabhängig sind, konstruieren wir für jedes Paar von Indices i, j eine lineare Abbildung

$$d_{i,j} \colon V \otimes W \to K$$

so dass $d_{i,j}(\mathbf{v}_i \otimes \mathbf{w}_j) = 1$ und $d_{i,j}(\mathbf{v}_k \otimes \mathbf{w}_\ell) = 0$ wenn $(i, j) \neq (k, \ell)$.

Das genügt, denn dann ist für jede Linearkombination

$$d_{i,j}(\sum_{k,\ell} a_{k,\ell} \mathbf{v}_k \otimes \mathbf{w}_\ell) = a_{i,j}$$

und also hat die Gleichung $\sum_{k,l} a_{k,l} \mathbf{v}_k \otimes \mathbf{w}_l = 0$ nur die triviale Lösung $a_{i,j} = 0$ für alle i, j.

Um $d_{i,j}$ zu konstruieren, nehmen wir die dualen Basen $\check{\mathbf{v}}_i, \check{\mathbf{w}}_j$ der Basen von V und W. Dann ist für alle i, j die Abbildung

$$D_{i,j} \colon \bigoplus_{(\mathbf{v},\mathbf{w}) \in V \times W} K \to K$$

die auf den Basisvektoren durch $D_{i,j}((\mathbf{v}, \mathbf{w})) := \check{\mathbf{v}}_i(\mathbf{v}) \cdot \check{w}_j(\mathbf{w})$ gegeben ist eine Abbildung, die auf allen Elementen von Bilinrel verschwindet. Also ist die Abbildung

$$\begin{aligned} d_{i,j} \colon V \otimes W &\to K \\ \mathbf{v} \otimes \mathbf{w} &\mapsto D_{i,j}(\mathbf{v}, \mathbf{w}) = \check{\mathbf{v}}_i(\mathbf{v}) \cdot \check{\mathbf{w}}_j(\mathbf{w}) \end{aligned}$$

eine wohldefinierte lineare Abbildung, die wie gewünscht $\mathbf{v}_i \otimes \mathbf{w}_j$ nach 1, aber alle anderen Elemente des Erzeugendensystems auf 0 abbildet. □

Damit können wir nun zeigen, dass wir tatsächlich alle Relationen der Abbildungen $f \cdot \mathbf{w} \in \mathrm{Hom}(V, W)$ gefunden haben.

Folgerung 66 ($\mathrm{Hom}(V, W) \cong V^\vee \otimes W$)**.** *Sind V, W zwei endlich dimensionale K-Vektorräume, so ist die Abbildung*

$$\begin{aligned} V^\vee \otimes W &\to \mathrm{Hom}(V, W) \\ f \otimes \mathbf{w} &\mapsto f(\_\_) \cdot \mathbf{w} \end{aligned}$$

ein wohldefinierter Isomorphismus.

Beweis. Die Abbildung ist wie zuvor wohldefiniert, da die Elemente aus Bilinrel nach Konstruktion auf die 0-Abbildung abgebildet werden. Die Abbildung ist surjektiv, weil Sie gezeigt haben, dass wir jede lineare Abbildung als Linearkombination von Abbildungen der Form $f(\_\_) \cdot \mathbf{w}$ schreiben können. Da die Vektorräume die gleiche Dimension haben, ist die Abbildung daher ein Isomorphismus. □

FAZIT: Damit haben wir nun gezeigt, dass in der Tat zwei Ausdrücke der Form $\sum_{i=1}^n f_i \cdot \mathbf{w}_i$ und $\sum_{j=1}^m f'_j \mathbf{w}'_j$ mit $\mathbf{w}_i, \mathbf{w}'_j \in W$, $f_i, f'_j \in V^\vee$ genau dann die gleiche lineare Abbildung definieren, wenn wir die Ausdrücke mittels der bilinearen Relationen ineinander umformen können.

Die Aussage, dass der Rang einer linearen Abbildung $A\colon V \to W$ genau die kleinste Zahl r ist, für die wir $A = \sum_{i=1}^r f_i \cdot \mathbf{w}_i$ schreiben können, die Sie auf Blatt 12 gezeigt haben lässt sich damit auch so formulieren, dass der Rang der Abbildung die minimale Anzahl von Tensoren $f \otimes \mathbf{w} \in V^\vee \otimes W$ ist, die nötig sind, um $A \in V^\vee \otimes W \cong \mathrm{Hom}(V, W)$ als Summe von Tensoren zu schreiben. Auf diesen „Tensorrang“ kommen wir noch zurück.

DAS ARGUMENT FÜR DIE DIMENSION des Tensorproduktes, erklärt uns auch eine allgemeinere Aussage, genannt die universelle Eigenschaft des Tensorproduktes. Hierbei schreiben wir für K-Vektorräume V, W, T

$$\mathrm{Bilin}(V \times W, T) := \{B\colon V \times W \to T \mid B \text{ bilinear}\}.$$

Für Bilinearformen hatten wir diese Definition für $T = K$ verwendet. Das Tensorprodukt selbst können wir als bilineare Abbildung:

$$\begin{aligned} \otimes\colon V \times W &\to V \otimes W \\ (\mathbf{v}, \mathbf{w}) &\mapsto \mathbf{v} \otimes \mathbf{w} \end{aligned}$$

auffassen, denn wir hatten per Definition für die bilinearen Relationen $(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} = \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}$ usw. gesorgt. Die Abbildung $\otimes$ ist also bilinear (aber nicht linear, nicht surjektiv und nicht injektiv).

Definition. Die Elemente von $V \otimes W$ heißen *Tensoren*. Die Elemnte, die von der Form $\mathbf{v} \otimes \mathbf{w}$ sind, also die Elemente des Bildes von $\otimes\colon V \times W \to V \otimes W$ heißen *einfache Tensoren*.

Beispiel 67. Die einfachen Tensoren in $V^\vee \otimes W \cong \mathrm{Hom}(V, W)$, also die Elemente der Form $f \otimes \mathbf{w}$ entsprechen genau den Abbildungen vom Rang 1.

Satz 68 (Universelle Eigenschaft des Tensorproduktes)**.** *Sind V, W zwei K-Vektorräume, dann sind für alle K-Vektorräume T die Abbildungen*

$$\mathrm{Bilin}(V \times W, T) \leftrightarrows \mathrm{Hom}(V \otimes W, T)$$
$$B \mapsto f_B(\sum_{i=1}^r \mathbf{v}_i \otimes \mathbf{w}_i) := \sum_{i=1}^r B(\mathbf{v}_i, \mathbf{w}_i)$$
$$B_F(\mathbf{v}, \mathbf{w}) := F(\mathbf{v} \otimes \mathbf{w}) \leftarrow\!\shortmid F$$

zueinander inverse Isomorphismen von Vektorräumen, d.h. bilineare Abbildungen auf $V \times W$ *entsprechen* linearen Abbildungen auf $V \otimes W$.

Sie sollten an diesem Satz gerne einmal innehalten. Mit etwas Glück, können Sie die Aussage jetzt lesen und verstehen, vielleicht brauchen Sie dafür etwas Zeit, aber Sie wissen jetzt, dass Sie sich dafür einfach an die Definitionen erinnern müssen und dann sehen, dass die Formeln die in Worten formuliete Aussage präzisieren.

Wenn das so ist, können Sie ein wenig stolz sein, denn das ist nicht so einfach. Der erste Blick auf Formeln wie diese von jemandem, der nicht Mathematik studiert hat ist von Thomas Mann im Roman Königliche Hoheit so beschrieben: *Was er sah, war sinnverwirrend. In einer krausen, kindlich dick aufgetragenen Schrift, die Imma Spoelmanns besondere Federhaltung erkennen ließ, bedeckte ein phantastischer Hokuspokus, ein Hexensabbat verschränkter Runen die Seiten. Griechische Schriftzeichen waren mit lateinischen und mit Ziffern in verschiedener Höhe verkoppelt, mit Kreuzen und Strichen durchsetzt, ober- und unterhalb wagrechter Linien bruchartig aufgereiht, durch andere Linien zeltartig überdacht, durch Doppelstrichelchen gleichgewertet, durch runde Klammern zu großen Formelmassen vereinigt.*

Die Passage ist sicherlich autobiografisch beeinflusst, denn Katharina Mann (geb. Pringsheim) hat wie die Figur der Imma aus dem Roman Mathematik und Physik studiert.

Beweis. Das Argument aus dem vorigen Resultat liefert wieder, dass die Abbildung wohldefiniert ist, denn die Abbildung

$$F_B\colon \oplus_{(\mathbf{v},\mathbf{w}) \in V \times W} \to K$$

die auf den Basisvektoren durch $(\mathbf{v}, \mathbf{w}) \mapsto B(\mathbf{v}, \mathbf{w})$ gegeben ist, nimmt – weil B bilinear ist – auf allen Elementen von Bilinrel den Wert 0 an. Darum ist $f_B(\mathbf{v} \otimes \mathbf{w}) := F_B(\mathbf{v}, \mathbf{w}) = B(\mathbf{v}, \mathbf{w})$ wohldefiniert und linear.

Ist umgekehrt $F\colon V \otimes W \to T$ linear, so ist die Abbildung $B_F(\mathbf{v}, \mathbf{w}) := F(\mathbf{v} \otimes \mathbf{w})$ bilinear, weil $\otimes\colon V \times W \to V \otimes W$ bilinear ist und F linear. Die Konstruktionen

$$B \mapsto F_B \text{ und } F \mapsto B_F$$

sind nach Definition zueinander inverse Abbildungen. $\square$

Anwendungen und Varianten des Tensorproduktes

Im Beweis des Spektralsatzes für selbstadjungierte Abbildungen auf einem euklidischen Vekotrraum $V, \langle , \rangle$ war es einerseits praktisch, die Aussage für allgemeine Vektorräume, nicht nur den $\mathbb{R}^n$ zu formulieren, um im Induktionsargument die Aussage für Unterräume verwenden zu können. Anderseits war es auch praktisch, zwischendurch von $\mathbb{R}^n$ zu $\mathbb{C}^n$ zu wechseln. Es wäre daher eleganter, ohne Wahl einer Basis aus einem $\mathbb{R}$-Vektorraum einen $\mathbb{C}$-Vektorraum der gleichen Dimension zu konstruieren.

Das Tensorprodukt liefert uns eine solche Methode: Ist V ein $\mathbb{R}$-Vektorraum, so können wir

$$V_{\mathbb{C}} := \mathbb{C} \otimes_{\mathbb{R}} V$$

definieren und auf $V_{\mathbb{C}}$ können wir durch $z \cdot (a \otimes \mathbf{v}) := (za) \otimes \mathbf{v}$ eine Multiplikation mit komplexen Zahlen definieren. Ist $V = \mathbb{R}^n$, und $\mathbf{e}_1, \ldots, \mathbf{e}_n$ die Standardbasis von $\mathbb{R}^n$, so sind die Elemente von $\mathbb{C} \otimes \mathbb{R}^n$ Summen von Tensoren der Form $z_1 \otimes \mathbf{e}_1, \ldots, z_n \otimes \mathbf{e}_n$ mit $z_i \in \mathbb{C}$, also ist dieser Raum tatsächlich eine Beschreibung von $\mathbb{C}^n$. Die Konstruktion ist also eine direkte Verallgemeinerung des Übergangs von $\mathbb{R}^n$ zu $\mathbb{C}^n$.

Wir können also mit dem Tensorprodukt die Skalare in K-Vektorräumen vergrößern, wenn ein größerer Körper als der, mit dem wir begonnen haben benötigt wird.

Mehrfache Tensorprodukte geben uns nun doch eine sinnvolle Möglichkeit, Vektoren miteinander zu multiplizieren, das Ergebnis liegt dann aber in einem neuen Vektorraum.

Bemerkung (Assoziativität von $\otimes$)**.** Für K-Vektorräume $V_1, \ldots, V_n$ ist

$$V_1 \otimes V_2 \otimes \ldots \otimes V_n := \bigoplus_{(\mathbf{v}_1,\ldots,\mathbf{v}_n) \in V_1 \times \cdots \times V_n} K / \text{Multlinrel}$$

wobei Multlinrel der Unterraum von $\bigoplus_{(\mathbf{v}_1,\ldots,\mathbf{v}_n) \in V_1 \times \cdots \times V_n} K$ ist, der von den Elementen der Form

$$(\ldots, \mathbf{v}_i + \mathbf{v}_i', \ldots,) - (\ldots, \mathbf{v}_i, \ldots) - (\ldots, \mathbf{v}_i', \ldots) \text{ und}$$

$$(\ldots, (c\mathbf{v}_i), \ldots) - c(\ldots, \mathbf{v}_i, \ldots)$$

aufgespannt wird. Wir schreiben wieder $\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n := [(\mathbf{v}_1, \ldots, \mathbf{v}_n)]$ für die entsprechende Äquivalenzklasse. Das gleiche Argument wie zuvor zeigt, dass für alle K-Vektorräume T lineare Abbildungen $V_1 \otimes \ldots \otimes V_n \to T$ multilinearen Abbildungen von $V_1 \times \cdots \times V_n \to T$ entsprechen.

Sie können sich auch überlegen, dass

$$(\ldots(V_1 \otimes V_2) \otimes V_3) \otimes \ldots) \otimes V_n \xrightarrow{\cong} V_1 \otimes \ldots \otimes V_n$$

isomorph sind. Daher können wir Klammern bei mehrfachen Tensorprodukten wie beim Rechnen in Körpern, weglassen.

Beispiel 69. Ein Elementen von $(\mathbb{R}^2)^{\otimes 3}$ wäre zum Beispiel:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 2 \\ 3 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 7 \\ 0 \end{pmatrix}.$$

Die bilinearen Rechenregeln erlauben uns zum Beispiel die Rechnung

$$\begin{aligned}
&\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}2\\7\end{pmatrix}\otimes\begin{pmatrix}2\\2\end{pmatrix}\\
&=\begin{pmatrix}1\\0\end{pmatrix}\otimes(\begin{pmatrix}2\\0\end{pmatrix}+\begin{pmatrix}0\\7\end{pmatrix})\otimes\begin{pmatrix}2\\2\end{pmatrix}\\
&=\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}2\\0\end{pmatrix}\otimes\begin{pmatrix}2\\2\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\7\end{pmatrix}\otimes\begin{pmatrix}2\\2\end{pmatrix}\\
&=\begin{pmatrix}1\\0\end{pmatrix}\otimes(2\cdot\begin{pmatrix}1\\0\end{pmatrix})\otimes(2\begin{pmatrix}1\\1\end{pmatrix})+\begin{pmatrix}1\\0\end{pmatrix}\otimes(7\cdot\begin{pmatrix}0\\1\end{pmatrix})\otimes\begin{pmatrix}2\\2\end{pmatrix}\\
&=2\cdot 2(\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix})+\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}14\\14\end{pmatrix}\\
&=\begin{pmatrix}4\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}14\\14\end{pmatrix}.
\end{aligned}$$

Definition. Für ein Element t aus einem Tensorprodukt $V_1 \otimes \ldots \otimes V_n$ ist der *Tensorrang* von t die kleinste Zahl r, so dass sich t als Summe

$$t = \sum_{i=1}^{r} \mathbf{v}_{1,i} \otimes \ldots \otimes \mathbf{v}_{n,i}$$

schreiben lässt.

Beispiel 70 (Matrixmultiplikation ist ein 3-facher Tensor.)**.** Die Matrizenmultiplikation ist eine Abbildung

$$\begin{aligned}
\circ\colon \mathrm{Mat}_{n,m} \times \mathrm{Mat}_{m,\ell} &\to \mathrm{Mat}_{n,\ell}\\
A, B &\mapsto A \circ B.
\end{aligned}$$

Diese Abbildung ist bilinear weil die Distributivgesetze $(A_1 + A_2) \circ B = A_1 \circ B + A_2 \circ B$ und $A \circ (B_1 + B_2) = A \circ B_1 + A \circ B_2$ für die Komposition von Abbildungen gelten. Diese bilineare Abbildung entspricht also einer linearen Abbildung

$$\circ\colon \mathrm{Mat}_{n,m} \otimes \mathrm{Mat}_{m,\ell} \to \mathrm{Mat}_{n,\ell}$$

und weil $\mathrm{Hom}(V, W) \cong V^\vee \otimes W$ ist damit einem Tensor

$$\mathrm{Mult} \in \mathrm{Mat}_{n,m}^\vee \otimes \mathrm{Mat}_{m,\ell}^\vee \otimes \mathrm{Mat}_{n,\ell}\,.$$

Bemerkung. 1. Wir haben heimlich verwendet, dass

$$(V \otimes W)^\vee \cong V^\vee \otimes W^\vee.$$

Das gilt, weil die Abbildung $(f \otimes g) \mapsto \big((v \otimes w) \mapsto f(\mathbf{v})g(\mathbf{w})\big)$ ein wohldefinierter Isomorphismus ist. Es ist eine gute Übung, zu versuchen, sich das genau zu überlegen, damit können Sie sich noch besser an die Argumente mit $\otimes$ und $^\vee$ gewöhnen.

2. Das ist alles einfacher, als es aussieht: Das Tensorprodukt $V^\vee \otimes W$ war gemacht um lineare Abbildungen als Summen von Ausdrücken der Form $f \cdot \mathbf{w}$ zu schreiben, der Tensor Mult schreibt $\circ$ als Summe von Ausdrücken der Form $f \cdot g \cdot C$, wobei C eine $n \times \ell$-Matrix ist und f, g jeweils lineare Formeln in den Einträgen der $n \times m$ bzw. $m \times \ell$-Matrizen sind.

GANZ KONKRET gilt zum Beispiel gilt für 2×2-Matrizen

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \circ \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix} =: \begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix}.$$

Wenn wir die Standardbasis für 2×2-Matrizen entsprechend der Einträge a, b, c, d als $\mathbf{e}_a, \mathbf{e}_b, \mathbf{e}_c, \mathbf{e}_d$ notieren und im Tensorprodukt $\mathrm{Mat}_{n,m}^\vee \otimes \mathrm{Mat}_{m,\ell}^\vee \otimes \mathrm{Mat}_{n,\ell}$ die 2. Komponente mit $'$ und die dritte mit $''$ bezeichnen entspricht Mult also dem Tensor

$$\begin{aligned} &\mathbf{e}_a^\vee \otimes \mathbf{e}_{a'}^\vee \otimes \mathbf{e}_{a''} + \mathbf{e}_b^\vee \otimes \mathbf{e}_{b'}^\vee \otimes \mathbf{e}_{a''} \\ +&\mathbf{e}_a^\vee \otimes \mathbf{e}_{b'}^\vee \otimes \mathbf{e}_{b''} + \mathbf{e}_b^\vee \otimes \mathbf{e}_{b'}^\vee \otimes \mathbf{e}_{b''} \\ +&\mathbf{e}_c^\vee \otimes \mathbf{e}_{a'}^\vee \otimes \mathbf{e}_{c''} + \mathbf{e}_d^\vee \otimes \mathbf{e}_{c'}^\vee \otimes \mathbf{e}_{c''} \\ +&\mathbf{e}_c^\vee \otimes \mathbf{e}_{b'}^\vee \otimes \mathbf{e}_{d''} + \mathbf{e}_d^\vee \otimes \mathbf{e}_{d'}^\vee \otimes \mathbf{e}_{d''}. \end{aligned}$$

Konkret ist $\mathbf{e}_a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ und $\mathbf{e}_a^\vee$ die Abbildung $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a$. Der Tensor $\mathbf{e}_a^\vee \otimes \mathbf{e}_{a'}^\vee \otimes \mathbf{e}_{a''}$ entspricht also „Nehme den Eintrag a der ersten Matrix, den Eintrag a' der zweiten, multipliziere diese und schreibe das Ergebnis oben links in das Ergebnis.“

Dass wir hier eine Formel aus 8 einfachen Tensoren $f \otimes g \otimes \mathbf{w}$ erhalten entspricht der Bemerkung, dass wir bei der Berechnung des Produktes 8 Multiplikationen berechnen müssen.

FÜR SIE UND FÜR COMPUTER ist es multiplizieren viel aufwändiger als addieren. Bei der Matrizenmultiplikation müssen wir leider ziemlich viel multiplizieren. Da in den acht verschiedenen Termen in der Multiplikation immer höchstens eine Koordinate übereinstimmt scheint es außerdem einleuchtend, dass wir bei der Berechnung des Produktes tatsächlich 8 Multiplikationen benötigen, d.h. dass wir keinen kürzeren Ausdruck für den Tensor Mult finden können.

Volker Strassen hat Ende der 60er Jahre versucht, diese Aussage für $n \times n$-Matrizen zu beweisen. Das war schwieriger als gedacht, so dass er sich den Fall der 2×2 Matrien noch einmal genau angeschaut hat und dabei die folgende sehr überraschende Formel für Mult gefunden hat, die mit 7 Tensoren, also mit 7 Multiplikationen auskommt

Bitte rechnen Sie die Formel einmal nach, um zu sehen, ob ich mich vertippt habe.

$$\begin{array}{cccccc} & (\mathbf{e}_a^\vee + \mathbf{e}_d^\vee) & \otimes & (\mathbf{e}_{a'}^\vee + \mathbf{e}_{d'}^\vee) & \otimes & (\mathbf{e}_{a''} + \mathbf{e}_{d''}) \\ + & (\mathbf{e}_c^\vee + \mathbf{e}_d^\vee) & \otimes & \mathbf{e}_{a'}^\vee & \otimes & (\mathbf{e}_{c''} - \mathbf{e}_{d''}) \\ + & \mathbf{e}_a^\vee & \otimes & (\mathbf{e}_{b'}^\vee - \mathbf{e}_{d'}^\vee) & \otimes & (\mathbf{e}_{b''} + \mathbf{e}_{d''}) \\ + & \mathbf{e}_d^\vee & \otimes & (\mathbf{e}_{c'}^\vee - \mathbf{e}_{a'}^\vee) & \otimes & (\mathbf{e}_{a''} + \mathbf{e}_{c''}) \\ + & (\mathbf{e}_a^\vee + \mathbf{e}_b^\vee) & \otimes & \mathbf{e}_{d'}^\vee & \otimes & (\mathbf{e}_{b''} - \mathbf{e}_{a''}) \\ + & (\mathbf{e}_c^\vee - \mathbf{e}_a^\vee) & \otimes & (\mathbf{e}_{a'}^\vee + \mathbf{e}_{b'}^\vee) & \otimes & \mathbf{e}_{d''} \\ + & (\mathbf{e}_b^\vee - \mathbf{e}_d^\vee) & \otimes & (\mathbf{e}_{c'}^\vee + \mathbf{e}_{d'}^\vee) & \otimes & \mathbf{e}_{a''}. \end{array}$$

Da sich das Prinzip für $4 \times 4, 8 \times 8, \ldots$ jeweils induktiv mehrfach verwenden lässt, liefert das ein sehr viel schnelleres Verfahren, Matrizen auf dem Computer zu multiplizieren, das darum sehr viel Verwendung findet.

ES IST MITTLERWEILE BEKANNT, dass der Multiplikationstensor für die Multiplikation von 2×2-Matrizen tatsächlich den Tensorrang 7 hat, d.h., dass es keine Formel für die Multiplikation gibt, die mit weniger als 7 Multiplikationen auskommt. Erstaunlicherweise gibt es dafür aber noch keinen einfachen Beweis.

Es ist noch immer unbekannt, wie viele Tensoren die kürzest mögliche Formel für die Multipikation von 3×3 Matrizen benötigt, d.h. was der Tensorrang des Multiplikationstensors für größere Matrizen ist! Es wäre sehr nützlich, das zu klären.

Das $\otimes$-Produkt als Multiplikation

Wenn wir für jede natürliche Zahl $n \geq 0$

$$V^{\otimes n} := \underbrace{V \otimes \ldots \otimes V}_{n \text{ Faktoren}}$$

setzen und $V^{\otimes 0} := K$ definieren können wir mit Hilfe dieser Konstruktion jeden Vektorraum V auf der direkten Summe

$$T^{\bullet} V := \bigoplus_{i=0}^{\infty} V^{\otimes n}$$

eine Multiplikation erklären, indem wir für einfache Tensoren das Produkt:

$$(\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n) \cdot (\mathbf{v}'_1 \otimes \ldots \otimes \mathbf{v}'_m) := (\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_n \otimes \mathbf{v}'_1 \otimes \ldots \otimes \mathbf{v}'_m)$$

definieren und dies für direkte Summen von Tensoren mit dem Distributivgesetz bilinear fortsetzen, d.h.

$$\begin{aligned}
&(\sum_{j=1}^{r} (\mathbf{v}_{j,1} \otimes \ldots \otimes \mathbf{v}_{j,n_j})) \cdot (\sum_{k=1}^{r'} (\mathbf{v}'_{k,1} \otimes \ldots \otimes \mathbf{v}'_{k,m_k})) \\
&:= \sum_{j=1}^{r} \sum_{k=1}^{r'} (\mathbf{v}_{j,1} \otimes \ldots \otimes \mathbf{v}_{j,n_j} \otimes \mathbf{v}'_{k,1} \otimes \ldots \otimes \mathbf{v}'_{k,m_k}).
\end{aligned}$$

Die Verknüpfungen $+$ und $\cdot = \otimes$ auf der Menge $T^{\bullet} V$ erfüllen also bis auf die Existenz von multiplikativen Inversen und dem Kommutativgesetz für $\otimes$ die Rechenregeln für Körper. Solche Rechenbereiche werden Ring genannt, zum Beispiel sind die ganzen Zahlen $\mathbb{Z}$ oder der Polynomring $k[x]$ *Ringe.* Wir schreiben die Axiome gleich noch formal auf, aber lassen Sie uns erst einmal verstehen, wie wir $T^{\bullet} V$ explizit verstehen können.

Beispiel 71 ($T^{\bullet} V$ ist ein nicht-kommutativer Polynomring)**.** Ist $\mathbf{x}_1, \ldots, \mathbf{x}_d \in V$ eine Basis von V, so wissen wir, dass die Elemente $\{\mathbf{x}_{j_1} \otimes \ldots \otimes \mathbf{x}_{j_n} \mid j_1, \ldots, j_n \in \{1, \ldots, d\}\} \in V^{\otimes n}$ eine Basis von $V^{\otimes n}$, d.h. alle Elemente von $T^{\bullet} V$ lassen sich eindeutig als Linearkombinationen von formalen Produkten der $\mathbf{x}_i$ schreiben:

$$\sum_{n=0}^{N} \sum_{j_1, \ldots, j_n \in \{1, \ldots d\}} a_{j_1, \ldots, j_n} \mathbf{x}_{j_1} \otimes \ldots \otimes \mathbf{x}_{j_n}.$$

Den Ring $T^{\bullet} V$ können wir also nach Wahl einer Basis wie einen Polynomring in den Symbolen $\mathbf{x}_1, \ldots, \mathbf{x}_d$ verstehen, in dem aber $\mathbf{x}_1 \otimes \mathbf{x}_2 \neq \mathbf{x}_2 \otimes \mathbf{x}_1$ ist, d.h.

$T^{\bullet} V$ ist ein nicht-kommutativer Polynomring.

Es ist einerseits ganz schön für nicht kommutative Polynomringe eine konkrete Konstruktion zu haben und andererseits ist das eine Konstruktion eines Polynomrings, die keine Wahl von Koordinaten benötigt. Dass das nützlich ist haben wir bei der Matrix-Multiplikation schon gesehen, denn so können wir exakt formulieren was wir meinen, wenn wir fragen, ob es eine andere Basis von V gibt, bezüglich der ein Polynom möglichst wenig Terme hat.

Zur Erinnerung schreiben wir die Axiome für die Rechenoperationen noch einmal auf, Sie kennen diese aus den Körperaxiomen.

Definition. Eine *Ring* ist eine Menge R zusammen mit Verknüpfungen

$$\begin{aligned} +\colon K \times K &\to K \\ (a,b) &\mapsto a+b \\ \cdot\colon K \times K &\to K \\ (a,b) &\mapsto a \cdot b, \end{aligned}$$

die folgende Eigenschaften erfüllen:

A0 (neutrales Element für +) Es gibt ein Element $0 \in K$ so dass für alle $a \in K$ gilt, dass $a + 0 = a$.

Jeder Zahlbereich sollte eine 0 enthalten.

AKom (Kommutativgesetz) Für alle $a, b \in K$ gilt $a + b = b + a$.

AAss (Assoziativgesetz) Für alle $a, b, c \in K$ gilt

$$(a + b) + c = a + (b + c).$$

Das war der Grund, warum wir in Rechnungen mit $+$ keine Klammern setzen müssen.

AInv (inverse Elemente) Zu jedem Element $a \in K$ existiert ein Element $-a \in K$ so dass $a + (-a) = 0$.

Hier ist $-$ versteckt.

MAss (Assoziativgesetz) Für alle $a, b, c \in K$ gilt $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

Das war der Grund, warum wir in Rechnungen mit $\cdot$ keine Klammern setzen müssen.

Dist (Distributivgesetz) Für alle $a, b, c \in K$ gilt

$$a \cdot (b + c) = a \cdot b + a \cdot c \text{ und}$$

$$(a + b) \cdot c = a \cdot c + b \cdot c.$$

Ein *Ring mit Eins* ist ein Ring, in dem zusätzlich ein 1-Element existiert:

M1 (neutrales Element für $\cdot$) Es gibt ein Element $1 \in K \setminus \{0\}$ so dass für alle $a \in K$ gilt, dass $1 \cdot a = a = a \cdot 1$.

Ein *kommmutativer Ring* ist ein Ring in dem zusätzlich das Kommutativgesetz gilt:

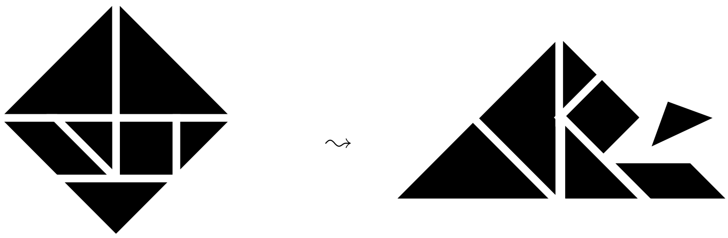
MKom (Kommutativgesetz) Für alle $a, b \in K$ gilt $a \cdot b = b \cdot a$.

Warum wir bei Körpern den Begriff des Schiefkörpers haben falls das Kommutativgesetz nicht gilt, Ringe aber nicht Schiefringe heißen, sondern stattdessen der Begriff „kommutativer Ring“ eingeführt wurde, ist eine Frage, die sich leider nicht mehr klären lässt.

Ausblick: Eine Anwendung eines merkwürdigen Tensorproduktes

Es passiert in der Mathematik recht häufig, dass eine Konstruktion zunächst aus einem konkreten Anlass gefunden wurde und die dann unvermutet in einer so absolut nicht vorgesehenen Variante Verwendung findet. Hierzu zum Abschluss der Vorlesung noch ein Beispiel.

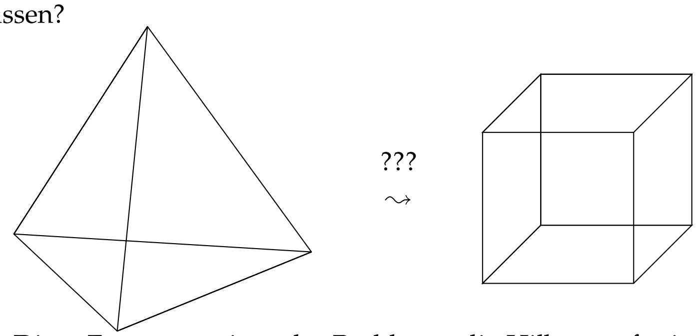
Sie wissen vielleicht, dass sich je zwei Polygone in der Ebene mit gleichem Flächeninhalt so in kleinere Polygone zerteilen lassen, dass sich diese zum jeweils anderen Polygon zusammenlegen lassen.



Aufgabe. Wenn Sie diese Aussage nicht kennen, dann ist das eine schöne Knobelaufgabe: Zeigen Sie, dass sich jedes Polygon mit Flächeninhalt a in Polygone unterteilen lässt, die sich zu einem Rechteck der Kantenlängen 1 und a zusammenlegen lassen.

Das müssen Sie nicht sofort lösen, aber wenn Sie in den Semesterferien einmal irgendwo Wartezeit oder Langeweile haben, denken Sie darüber nach.

Frage: Stimmt das im 3-dimensionalen? Lassen sich je zwei Polyeder mit gleichem Volumen im $\mathbb{R}^3$ so in Polyeder unterteilen, dass sich die Stücke zum jeweils anderen Polyeder zusammensetzen lassen?



Diese Frage war eines der Probleme, die Hilbert auf seiner Liste von 23 offenen Fragen 1900 auf dem ICM in Paris vorgestellt hat. Er vermutete, dass die Antwort nein ist. Die Frage ist sicherlich viel älter, aber es fehlte für sehr lange Zeit eine gute Methode, um das Problem anzugreifen.

Die Antwort ist: „Nein, das ist nicht immer möglich.“ Das wurde von Max Dehn gezeigt und das Prinzip der Lösung können wir mit dem Tensorprodukt verstehen.

Die Grundidee ist vielseitig verwendbar: Da es für geometrische Fragen nicht klar ist, wie wir jemals einen so guten Überblick über alle möglichen Konstruktionen bekommen könnten, dass wir beweisen können, dass keine Konstruktion das gewünschte Ergebnis liefert, sollten wir uns überlegen, ob wir den Polyedern irgendeine überschaubare Liste von Zahlen oder anderen algebraisch formulierten Eigenschaften zuordnen können, die bei Zerlegung gleich bleibt. Die Dehninvariante ist so eine Zuordnung:

$$\begin{aligned} D\colon \{\text{Polyeder}\} &\to \mathbb{R} \otimes_{\mathbb{Q}} \big(\mathbb{R}/(\mathbb{Q}\cdot\pi)\big) \\ P &\mapsto \sum_{e \text{ Kante von P}} \text{Länge}(e) \otimes \text{Innenwinkel}(e). \end{aligned}$$

Im Tensorprodukt fassen wir hier die reellen Zahlen $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum auf, d.h. wir beschließen vorübergehend nur noch die Multiplikation mit rationalen Zahlen zuzulassen. Die Zahl $\pi \in \mathbb{R} \setminus \mathbb{Q}$ definiert dann einen 1-dimensionalen $\mathbb{Q}$-Unterraum, den wir im zweiten Faktor des Tensorproduktes im Quotientenraum als 0 auffassen.

Die Formel kommt aus der Beobachtung, dass beim Zerteilen von Polyedern durch das Zerschneiden von Seitenflächen jeweils 2 neue Kanten e_1, e_2 gleicher Länge ℓ entstehen, deren Innenwinkel α_1, α_2 zusammen $180° = \pi$ ergeben. Im Tensorprodukt $\mathbb{R} \otimes_{\mathbb{Q}} (\mathbb{R}/\mathbb{Q}\cdot\pi)$ ist dann

$$\ell \otimes \alpha_1 + \ell \otimes \alpha_2 = \ell \otimes (\alpha_1 + \alpha_2) = \ell \otimes \pi = \ell \otimes 0 = 0.$$

Das erklärt, warum sich $D(P)$ nicht ändert, wenn wir P zerteilen.

Für den Würfel sehen Sie vielleicht auch gleich, dass $D(\text{Würfel}) = 0$ ist, weil die Innenwinkel der Kanten alle $\pi/2 = 0 \in \mathbb{R}/\mathbb{Q}\pi$ sind.

Für den gleichseitigen Tetraeder können Sie mit dem, was wir in der Linearen Algebra über Skalarprodukte und Winkel gelernt haben cos(Innenwinkel) berechnen. Sie können dann versuchen, sich zu überlegen, warum der entsprechende Winkel nicht in $\mathbb{Q}\pi$ liegen kann und darum $D(\text{Tetraeder}) \neq 0$ gilt. Darum können Sie den Tetraeder nicht so unterteilen, dass sich die Stücke zu einem Würfel zusammensetzen lassen.

Restekiste

In der Vorlesung haben ich die beiden Varianten $\mathrm{Sym}^n V, \bigwedge^n V$ des Tensorproduktes nicht erklärt. Für den Fall, dass Ihnen später einmal symmetrische oder alternierende Produkte begegnen, belasse ich die kurzen Abschnitte dazu hier im Skript.

Ausblick 1: Symmetrische Produkte und Polynomfunktionen

Mit einer ähnlichen Konstruktion können wir jetzt auch erklären was Polynomfunktionen auf einem Vektorraum sind. Dazu können wir einfach auf dem Tensorprodukt mit einer zusätzlichen Quotientenkonstruktion die Rechenregel:

$$[(\ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots)] = [(\ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots)]$$

erzwingen, d.h. wir fügen zum Unterraum Multlinrel alle Ausdrücke der Form $(\ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots) - (\ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots)$ hinzu:

$$\mathrm{SymBilin} := \mathrm{Span}(\mathrm{Multlinrel}, \{(\ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots) - (\ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots) \mid \mathbf{v}_1, \ldots, \mathbf{v}_n \in V\}).$$

$$\begin{aligned} \mathrm{Sym}^k V &:= \bigoplus_{(\mathbf{v}_1, \ldots, \mathbf{v}_n) \in V_1 \times \cdots \times V_n} K / \mathrm{SymBilin} \\ &= V^{\otimes n} / \mathrm{Span}(\{\ldots \mathbf{v}_i \otimes \ldots \otimes \mathbf{v}_i \otimes \cdots - \ldots \mathbf{v}_j \otimes \ldots \otimes \mathbf{v}_i \otimes \cdots \mid \mathbf{v}_1, \ldots, \mathbf{v}_n \in V\})). \end{aligned}$$

Wir schreiben dann

$$\mathbf{v}_1 \bullet \cdots \bullet \mathbf{v}_k := [\mathbf{v}_1 \otimes \ldots \otimes \mathbf{v}_k]$$

für die Äquivalenzklasse des Elementes $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ in $\mathrm{Sym}^k V$ und nennen $\mathrm{Sym}^k V$ das k-te *symmetrische Produkt* von V.

Das symmetrische Produkt ist also wie das Tensorprodukt der Vektorraum der formalen Linearkombinationen von ausdrücken der Form $\mathbf{v}_1 \bullet \cdots \bullet \mathbf{v}_n$, nur dass für diese Ausdrücke außer den bilinearen Rechenregeln jetzt außerdem gilt, dass $\bullet$ kommutativ ist. Es ist nicht so schwer, sich zu überlegen, dass damit

$$\mathrm{Sym}^\bullet V := \bigoplus_{k=0}^{\infty} \mathrm{Sym}^k V$$

ein kommutativer Ring ist, den wir nach Wahl einer Basis als Polynomring verstehen können.

Einen symmetrische Tensor $p \in \mathrm{Sym}^k V$ können wir damit also als Polynom in den Variablen $\mathbf{x}_1, \dots, \mathbf{x}_n$ verstehen. Noch besser können wir jeden Tensor $p \in \mathrm{Sym}^k(V^\vee)$ als Polynomfunktion auf V auffassen und bekommen damit eine Möglichkeit Polynomfunktionen auf Vektorräumen zu beschreiben, ohne Koordinaten auf V zu wählen. Das gibt uns wieder eine Möglichkeit, die Frage ob ein Polynom in n-Variablen bezüglich anderer Koordinaten weniger Terme hat, präzise zu formulieren.

Ausblick 2: Äußere Produkte in der Analysis

Die Konstruktion von Determinanten hatten wir mit dem Versuch, eine Formel für das Volumen eines Spats zu finden motiviert. Das hatte uns darauf geführt alternierende, multilineare Abildungen

$$D\colon \underbrace{V \times \cdots \times V}_{n-\text{Faktoren}} \to K$$

also multilineare Abbildungen, für die

$$D(\dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots) = 0 \text{ falls } \mathbf{v}_i = \mathbf{v}_j$$

gilt. Daraus folgte, dass

$$D(\dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots) = -D(\dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots) = 0 \text{ falls } i \neq j.$$

Auch solche Abbildungen können wir mit einer Variante des Tensorproduktes beschreiben, indem wir zum Unterraum Multlinrel alle Ausdrücke der Form $(\dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots) - (\dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots)$ hinzufügen:

$$\text{AlternRel} := \text{Span}(\text{Multlinrel}, \{(\dots, \mathbf{v}_i, \dots, \mathbf{v}_i, \dots) \mid \mathbf{v}_1, \dots, \mathbf{v}_n \in V\})$$

$$\begin{aligned} \bigwedge^n V &:= \bigoplus_{(\mathbf{v}_1, \dots, \mathbf{v}_n) \in V_1 \times \cdots \times V_n} K / \text{AlternRel} \\ &= V^{\otimes n} / \text{Span}(\{\dots \mathbf{v}_i \otimes \dots \otimes \mathbf{v}_i \otimes \cdots \mid \mathbf{v}_1, \dots, \mathbf{v}_n \in V\})). \end{aligned}$$

Wir schreiben dann

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n := [\mathbf{v}_1 \otimes \dots \otimes \mathbf{v}_n]$$

für die Äquivalenzklasse des Elementes $(\mathbf{v}_1, \dots, \mathbf{v}_n)$ in $\bigwedge^n V$ und nennen $\bigwedge^n V$ das n-te *äußere Produkt* von V.

Das äussere Produkt ist also wie das Tensorprodukt der Vektorraum der formalen Linearkombinationen von ausdrücken der Form $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n$, nur dass für diese Ausdrücke außer den bilinearen Rechenregeln jetzt außerdem gilt, dass Ausdrücke in denen ein Vektor doppelt vorkommt gleich 0 sind, und damit auch

$$\cdots \wedge \mathbf{v}_i \wedge \cdots \wedge \mathbf{v}_j \wedge \cdots = - \cdots \wedge \mathbf{v}_j \wedge \cdots \wedge \mathbf{v}_i \wedge \dots$$

gilt. Damit können wir uns ähnlich wie für das Tensorprodukt überlegen, wie wir eine Basis des äußeren Produktes bekommen:

Behauptung 72 (Dimension von $\bigwedge^k V$)**.** *Ist V ein K-Vektorraum und $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis von V, dann sind die Vektoren*

$$(\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_k})_{1 \leq i_1 < i_2 < \cdots < i_k \leq n}$$

eine Basis des äußeren Produktes $\bigwedge^k V$, insbesondere ist

$$\dim \bigwedge^k V = \binom{n}{k}$$

die Anzahl der Möglichkeiten k verschiedene Zahlen aus der Menge $\{1, \ldots, n\}$ auszuwählen.

Diese Behauptung möchte ich nicht beweisen, Sie können das aber selbst versuchen. Diese Raum wird Ihnen in der Analysis begegnen, wenn Sie zum Beispiel die Oberfläche eines gekrümmten Objektes im $\mathbb{R}^3$ berechnen möchten. Dafür benötigen Sie einen Oberflächenbegriff. Die Determinante liefert einen guten Volumenbegriff und wir können die Determinante als lineare Abbildung

$$\bigwedge^{\dim V} V \to K$$

auffassen. Darum ist es naheliegend $\bigwedge^2 V$ zu verwenden, um einen Flächenbegriff für 2-dimensionale Teilmengen von V zu erhalten und ganz ähnlich $\bigwedge^k V$ für k-dimensionale Volumen von Teilmengen von V. Was Sie sich für den Moment davon merken sollten, ist dass $\bigwedge^k V$ ein Vektorraum von Symbolen ist, auf dem die multilinearen und die alternierenden Rechenregeln gelten.

Glossar mathematischer Symbole

Mathematische Symbole haben meist eine genauere Bedeutung, als die entsprechenden Wörter, beim Lesen ersetzen wir die Symbole im Kopf durch Wörter. Damit können Sie in Ihren Lösungen prüfen, ob das was Sie aufschreiben lesbar sein könnte.

$=$ „(ist) gleich“

$\neq$ „(ist) nicht gleich“

$\Rightarrow$ „daraus folgt“ oder „impliziert, dass“

$\Leftrightarrow$ „(gilt) genau dann, wenn“

$\{? \mid \text{etwas}\}$ „Die Menge der ? für die etwas gilt“

$\in$ „aus“ oder „Element von“

$\ni$ Ist das umgekehrte „Element von“, d.h. „$x \in M$“ und „$M \ni x$“ bedeuten beide „x ist Element von M“

$\subseteq$ „enthalten in“ oder „Teilmenge von“

$\supseteq$ „enthält“ oder „Obermenge von“

$\cup$ „vereinigt mit“ $\cup$ ist das Symbol für die Vereinigung zweier Mengen (d.h. für die Menge die Entsteht, wenn wir die Elemente der beiden Mengen zu einer Menge zusammenfassen).

$\cap$ „geschnitten mit“ $\cap$ ist das Symbol für den Durchschnitt zweier Mengen (d.h. die Menge, die aus den Elementen besteht, die in beiden Mengen enthalten sind).

$M \setminus N$ „Die Menge M ohne die Elemente aus N“, bzw. „Entferne N aus der Menge M“. Das Symbol „$\setminus$“ ist das Minuszeichen für Mengen, also „entferne“.

$\forall$ „Für alle“

$\#$ „Anzahl“

$\circ$ „verknüpft mit“. Das Symbol „$\circ$“ bezeichnet die Komposition von Abbildungen, d.h. $f \circ g(x) := f(g(x))$.

$\oplus$ „direkte Summme“. Das Symbol bezeichnet die direkte Summe von Untervektorräumen oder abstrakten Vektorräumen, die Gleichung $V = U \oplus W$ bedeutet: Jeder Vektor aus V lässt sich eindeutig als Summe $\mathbf{v} = \mathbf{u} + \mathbf{w}$ eines Elements aus U und eines aus W schreiben.

$\perp$ „senkrecht zu“ oder „orthogonal zu“

$\vee$ „dual“

$\cong$ „isomorph“ oder „natürlich isomorph“

V/U „V modulo U“ – der Quotientenraum

$\otimes$ „tensor“ – das Tensorprodukt

$\to$ „nach“ – der Pfeil der für Abbildungen verwendet wird.

$\mapsto$ „wird abgebildet auf“ – der Pfeil der für die Beschreibung von Abbildungen verwendet wird

$\hookrightarrow$ Pfeil für injektive Abbildungen

$\twoheadrightarrow$ Pfeil für surjektive Abbildungen

Literaturverzeichnis

Gerd Fischer. *Lineare Algebra*, volume 17 of *Grundkurs Mathematik*. Friedr. Vieweg & Sohn, Braunschweig, fifth edition, 1979. ISBN 3-528-17217-7. URL https://link.springer.com/book/10.1007/978-3-8348-9365-9. In collaboration with Richard Schimpl.

Ulrich Görtz. Lineare algebra. Vorlesungsskript, 2020. URL https://math.ug/lecture-notes.html.

David Lay, Steven Lay, and Judi McDonald. *Linear Algebra and Its Applications*. Harlow: Pearson Education, Limited, 2015. ISBN 9781292092232. URL https://elibrary.pearson.de/book/99.150005/9781292092249.

Wolfgang Soergel. Lineare algebra. Vorlesungsskript, 2022. URL http://home.mathematik.uni-freiburg.de/soergel/Skripten/XXLA1.pdf.

Geordie Williamson. Is deep learning a useful tool for the pure mathematician? Preprint, 2023. URL https://arxiv.org/abs/2304.12602.