Handout
Lineare Algebra 2

Dirk Lorenz

Inhaltsverzeichnis

1 Ringe und Polynome

Definition 1.1. Ein *Ring* ist eine Menge R versehen mit zwei Verknüpfungen $+, \cdot$, so dass gilt:

- R ist mit der Addition eine *abelsche Gruppe* (d.h. $+$ ist assoziativ, kommutativ, es gibt ein neutralen Element 0 und zu jedem $a \in R$ existiert ein inverses Elemente $-a$ mit $a + (-a) = 0$).
- Die Multiplikation $\cdot$ ist assoziativ, d.h. für alle $a, b, c \in R$ gilt $(ab)c = a(bc)$.
- Es gelten die Distributivgesetze

$$a(b + c) = ab + ac, \qquad (a + b)c = ac + bc.$$

Ist die Multiplikation auch kommutativ, so heißt R *kommutativer Ring*.

In einem Ring kann es ein Einselement geben, d.h. eine 1 mit $a1 = 1a = a$, es gibt aber auch Ringe ohne Einselement.

Schnell sieht man, dass in jedem Ring $0a = a0 = 0$ gilt.

Folgt aus $0a = (0 + 0)a = 0a + 0a$.

Einen Ring, in dem aus $ab = 0$ immer $a = 0$ oder $b = 0$ folgt, nennen wir *nullteilerfrei*.

Beispiel.

- Natürlich sind $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ ebenfalls Ringe, da sie sogar Körper sind. Die ganzen Zahlen $\mathbb{Z}$ bilden ebenfalls einen Ring, sogar einen kommutativen Ring mit Einselement.
- Ist I ein reelles Intervall, so ist

$$R = \{f : I \to \mathbb{R}\}$$

versehen mit der üblichen Addition und Multiplikation von Funktionen ein kommutativer Ring. Das Einselement ist die konstante Funktion $f(x) \equiv 1$. Beachte: Es gibt Funktionen, die nicht Null sind, und kein multiplikativ Inverses haben, z.B. jede Funktion mit einer Nullstelle (die nicht konstant Null ist).

- Die Restklassen $\mathbb{Z}_m$ bilden für jedes $m = 1, 2, \ldots$ einen Ring (wenn Addition und Multiplikation wie in der großen Übung 2 in Linearer Algebra 1 per Modulo-Rechnung definiert sind).
- $2\mathbb{Z} = \{2k \,:\, k \in \mathbb{Z}\}$ ist ein Ring ohne Einselement.
- $K^{n \times n}$ ist ein nicht-kommutativer Ring mit Einselement (E_n) er ist nicht nullteilerfrei, da $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) = 0$.

$\triangle$

Definition 1.2. Ist R ein Ring und $R' \subset R$ eine Teilmenge, so heißt R' *Unterring* von R, wenn R' bezüglich Addition und Multiplikation abgeschlossen ist.

Sind R und S zwei Ringe mit Verknüpfungen $+, \cdot$ bzw. $\oplus, \odot$, so heißt $\varphi : R \to S$ ein *Ring-Homomorphismus*, wenn $\varphi(a+b) = \varphi(a) \oplus \varphi(b)$ und $\varphi(a \cdot b) = \varphi(a) \odot \varphi(b)$ gelten.

Man könnte auch anders Definieren: $R' \subset R$ ist ein Unterring von R, wenn R' mit den Einschränkungen der Verknüpfungen auf R selbst ein Ring ist.

Zum Beispiel ist die Menge $m\mathbb{Z} = \{mn \mid n \in \mathbb{Z}\}$ ein Unterring von $\mathbb{Z}$ und die Abbildung $\mathbb{Z} \to \mathbb{Z}_m$, $a \mapsto a + m\mathbb{Z}$ ist ein Ring-Homomorphismus.

Ein wichtiges Beispiel für Ringe sind Polynomringe. Diese basieren auf einem Körper: Ist K ein Körper, und sind $a_0, \ldots, a_n \in K$ so ist der Ausdruck

$$f(t) = a_0 + a_1 t + \cdots + a_n t^n$$

ein *Polynom* mit Koeffizienten in K. Oftmals schreiben wir $f = a_0 + a_1 t + \cdots + a_n t^n$ statt $f(t)$ für Polynome. Das Polynom, bei dem alle Koeffizienten $a_k = 0$ sind, heißt *Nullpolynom*.

Wir könnten Polynome als Abbildungen auffassen - das tun wir jedoch *nicht*, unter anderem, da wir nicht festlegen wollen, was wir für die *Unbestimmte t* einsetzen. Sicherlich können wir Element aus K einsetzen, jedoch gehen auch andere Objekte die wie zum Beispiel quadratische Matrizen mit Elementen in K.

Die Menge aller Polynome mit Koeffizienten in K bezeichnen wir mit $K[t]$. Wir statten $K[t]$ mit der üblichen Addition aus: Seien $f = a_0 + a_1 t + \cdots + a_n t^n$ und $g = b_0 + b_1 t + \cdots + b_n t^n$ zwei Polynome. Dann ist die Summe definiert als

$$f + g = (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n.$$

Beachte, dass diese Defintion auch funktioniert, wenn der höchste Exponent in f und g nicht gleich ist, da wir einfach durch entsprechende Null-Koeffizienten ergänzen können.

Die Multiplikation von Polynomen erhält man, indem man die Ausdrücke nach den Rechenregeln ausmultipliziert, also

$$\begin{aligned}
f \cdot g &= (a_0 + a_1 t + \cdots + a_n t^n)(b_0 + b_1 t + \cdots + b_m t^m) \\
&= a_0(b_0 + b_1 t + \cdots + b_m t^m) + \cdots + a_n t^n (b_0 + b_1 t + \cdots + b_m t^m) \\
&= a_0 b_0 + a_0 b_1 t + \cdots + a_0 b_m t^m \\
&\qquad + a_1 b_0 t + a_1 b_1 t^2 + \cdots + a_1 b_m t^{m+1} \\
&\qquad + \cdots \\
&\qquad + a_n b_0 t^n + a_n b_1 t^{n+1} + \cdots + a_n b_m t^{n+m} \\
&= a_0 b_0 \\
&\qquad + (a_1 b_0 + a_0 b_1) t \\
&\qquad + (a_2 b_0 + a_1 b_1 + a_0 b_2) t^2 \\
&\qquad + \cdots \\
&\qquad + a_n b_m t^{n+m} \\
&= c_0 + c_1 t + \cdots + c_{n+m} t^{n+m}
\end{aligned}$$

mit

$$c_k = \sum_{i+j=k} a_i b_j.$$

Beispiel.

Für $f = t^3 - 3t^2 - t$, $g = t^2 + 1 \in \mathbb{R}[t]$ ist

$$\begin{aligned} f \cdot g = (t^3 - 3t^2 - t)(t^2 + 1) &= t^5 - 3t^4 - t^3 + t^3 - 3t^2 - t \\ &= t^5 - 3t^4 - 3t^2 - t. \end{aligned}$$

Für $f = t^2 + 1$, $g = t^3 + t + 1 \in \mathbb{Z}_2[t]$ ist

$$\begin{aligned} f \cdot g &= (t^2 + 1)(t^3 + t + 1) \\ &= t^5 + t^3 + t^3 + t + t^2 + 1 \\ &= t^5 + (1+1)t^3 + t^2 + t + 1 \\ &= t^5 + t^2 + t + 1. \end{aligned}$$

$\triangle$

Ist $f \cdot g = h$, so sind f und g *Teiler* von h.
Man prüft schnell nach:

Satz 1.3. *Ist K ein Körper, so ist $K[t]$ versehen mit Addition und Multiplikation ein Ring.*

Definition 1.4. Der *Grad* (engl. degree) eines Polynoms f ist

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

Wir nennen ein Polynom vom Grad n *normiert*, wenn $a_n = 1$.

Der Grad ist grob „der höchste Exponent". Beachte den Sonderfall $f = 0$. Polynome von Grad 0 sind von der Form $f = a_0$ mit $a_0 \neq 0$.

Wie schon in einer Randbemerkung erwähnt, können wir zu einem Polynom $f \in K[t]$ eine Funktion $\tilde{f} : K \to K$ definieren, im dem wir $\tilde{f}(t) = f(t) = a_0 + a_1 t + \cdots + a_n t^n$ setzen. Dass diese, auf den ersten Blick pedantisch erscheinende, Unterscheidung wichtig sein kann, sieht man an folgendem Beispiel:

Beispiel.

Wir betrachten den Körper $\mathbb{Z}_2$ mit zwei Elementen und die beiden Polynome

$$f = 0 \text{ (das Nullpolynom)}, \quad g = t + t^2.$$

Dann gilt für die entsprechenden Funktionen von $\mathbb{Z}_2$ nach $\mathbb{Z}_2$:

$$\begin{aligned} \tilde{f}(0) &= 0, \\ \tilde{f}(1) &= 0, \\ \tilde{g}(0) &= 0 - 0^2 = 0, \\ \tilde{g}(1) &= 1 + 1^2 = 1 + 1 = 0, \end{aligned}$$

d.h. die beiden Funktionen $\tilde{f}$ und $\tilde{g}$ sind gleich (da die Abbildungsvorschriften gleich sind), aber die Polynome sind nicht gleich (es gilt z.B. $\deg f = -\infty$, aber $\deg g = 2$). $\triangle$

Mit anderen Worten: Das obige Beispiel zeigt, dass die Abbildung

$$\tilde{\ }: K[t] \to \mathrm{Abb}(K, K),$$

welche einem Polynom $f \in K[t]$ die Funktion $\tilde{f} : K \to K$, definiert durch $\tilde{f}(x) = f(x)$, zuordnet, im Allgemeinen nicht injektiv ist.

Sind A, B zwei Mengen, so ist $\mathrm{Abb}(A, B)$ die Menge der Abbildungen von A nach B.

Lemma 1.5. *Ist K ein Körper, dann gilt für $f, g \in K[t]$*

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

wobei wir die Konventionen $n + (-\infty) = (-\infty) + n = (-\infty) + (-\infty) = -\infty$ verwenden.

Hier sieht man, dass die Konvention $\deg(0) = -\infty$ sinnvoll und praktisch ist.

Beweis.
Ist $f = 0$ oder $g = 0$, so steht auf beiden Seiten der Gleichung $-\infty$.

Es seien $\deg(f) = n > -\infty$ und $\deg(g) = m > -\infty$, d.h. für die Koeffizienten der jeweiligen höchsten Potenzen gilt $a_n \neq 0$, bzw. $b_m \neq 0$. Dann hat die höchste Potenz in fg den Exponenten $m + n$ und der Koeffizient ist $a_n b_m \neq 0$. □

2 Polynomdivision

Ein wesentlicher Unterschied eines Ringes zu einem Körper besteht darin, dass es in Ringen im Allgemeinen keine multiplikativ Inversen gibt, d.h. man kann in Ringen nicht dividieren. Ein prominentes Beispiel ist der Ring $\mathbb{Z}$ (im Gegensatz zum Körper $\mathbb{Q}$). Für die ganzen Zahlen $\mathbb{Z}$ kennen wir allerdings die *Division mit Rest*, d.h. zu zwei ganzen Zahlen $m, n \in \mathbb{Z}$ mit $m \neq 0$ können wir immer n mit Rest durch m dividieren. Diese Division mit Rest ist dadurch gekennzeichnet, dass

$$n = a \cdot m + b$$

mit $a \in \mathbb{Z}$ und $0 \leq b < |m|$ gilt (und a und b sind hierdurch eindeutig bestimmt). Eine entsprechende Division mit Rest gibt es nicht in jedem Ring, allerdings immer in Polynomringen.

Zum Beispiel ist für $n = 15$, $m = 6$

$$15 = 2 \cdot 6 + 3,$$

für $n = 10$, $m = -3$

$$10 = (-3) \cdot (-3) + 1,$$

und für $n = -10$, $m = -3$

$$-10 = 4 \cdot (-3) + 2.$$

Die algebraische Struktur, in der die Division mit Rest erklärt ist, heißt *euklidscher Ring*.

Satz 2.1. *Es seien $f, g \in K[t]$ und $g \neq 0$. Dann gibt es eindeutig bestimmte $q, r \in K[t]$ mit*

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

Hier steht q für „Quotient“ und r für „Rest“.

Beweis.
Zeigen wir zuerst die Eindeutigkeit: Seien $q, q', r, r' \in K[t]$ Polynome mit

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

Daraus bekommen wir

$$0 = (q - q')g + r - r'$$

und daraus folgt

$$(q - q')g = r' - r.$$

Wir bilden auf beiden Seiten den Grad und mit Lemma 1.5 folgern wir

$$\deg(r' - r) = \deg(q - q') + \deg(g). \tag{$*$}$$

Wäre $q \neq q'$, so wäre $\deg(q - q') \geq 0$ und aus der obigen Gleichung würde

$$\deg(r' - r) \geq \deg(g)$$

folgen. Das ist aber ein Widerspruch zu $\deg(r), \deg(r') < \deg(g)$. Daher ist $q = q'$ und aus $(*)$ folgt

$$\deg(r' - r) = \deg(0) + \deg(g) = -\infty + \deg(g) = -\infty$$

und damit $r = r'$.

Die Existenz zeigen wir, indem wir ein Verfahren zur Berechnung von q und r angeben: Wir schreiben

$$f = a_n t^n + \cdots + a_1 t + a_0, \quad g = b_m t^m + \cdots + b_1 t + b_0$$

mit $a_n, b_m \neq 0$. Ist $n < m$ so haben wir mit

$$f = 0 \cdot g + f, \quad \text{mit} \quad \deg(f) < \deg(g)$$

die gesuchten $q = 0$ und $r = f$ gefunden

Sei also nun $n \geq m$.

Schritt 1: Wir dividieren die Terme höchster Ordnung und erhalten

$$q_1 = \tfrac{a_n}{b_m} t^{n-m}$$

als höchsten Term des gesuchten q.

Schritt 2: Wir setzen

$$f_1 = f - q_1 \cdot g.$$

Da der höchste Term von $q_1 \cdot g$ genau $\frac{a_n}{b_m} t^{n-m} \cdot b_m t^m = a_n t^n$ ist, gilt

$$\deg(f_1) < \deg(f).$$

Schritt 3: Wir prüfen, ob $\deg(f_1) < m$ gilt. Ist dies so, so sind wir fertig, wenn wir $q = q_1$ und $r = f_1$ setzen. Andernfalls beginnen wir wieder bei Schritt 1 mit f_1 statt f (d.h. wir bekommen den nächsten Term q_2 von q durch Division der höchsten Terme und setzen $f_2 = f_1 - q_2 \cdot g$ und $\deg(f_2) < \deg(f_1)$).

Da die Grade der f_i bei jedem Durchgang um mindestens eins abnehmen, gibt es einen Index $k \leq n - m + 1$, so dass erstmalig

$$f_k = f_{k-1} - q_k \cdot g, \quad \text{und} \deg(f_k) < m = \deg(g).$$

An dieser Stelle brechen wir das Verfahren ab, und bekommen durch Einsetzen

$$\begin{aligned} f &= q_1 \cdot g + f_1 \\ &= q_1 \cdot g + q_2 \cdot g + f_2 \\ &= q_1 \cdot g + q_2 \cdot g + q_3 \cdot g + f_3 \\ &= \cdots = q_1 \cdot g + q_2 \cdot g + q_3 \cdot g + \cdots + q_k \cdot g + f_k \\ &= (q_1 + q_2 + \cdots + q_k) \cdot g + f_k \end{aligned}$$

und wir haben die Lösung gefunden, nämlich

$$q = q_1 + \cdots + q_k, \quad r = f_k.$$

□

Das Verfahren, das wir für den Existenzbeweis benutzt haben, ist genau das gleiche, das man in der Schule zum schriftlichen dividieren von ganzen Zahlen lernt.

Beispiel.

Wir betrachten $K = \mathbb{R}$, $f = 3t^3 + 2t + 1$ und $g = t^2 - 4t$. Wir rechnen nach dem Schema des Beweises, d.h. genau so wie bei der schriflichen Division von ganzen Zahlen. Dabei berücksichtigen wir den „fehlenden" Term t^2:

$$\begin{array}{l}
\big(\ \ 3t^3 \qquad\qquad + 2t + 1\,\big) : \big(t^2 - 4t\big) = 3t + 12 + \dfrac{50t+1}{t^2-4t} \\
\underline{-\,3t^3 + 12t^2} \\
\qquad\quad 12t^2 \ \ + 2t \\
\qquad \underline{-\,12t^2 + 48t} \\
\qquad\qquad\quad\ 50t
\end{array}$$

Das heißt, wir bekommen

$$f = q \cdot g + r \quad \text{mit} \quad q = 3t + 12, \quad r = 50t + 1.$$

Ein zweites Beispiel: $f = t^6 - 3t^5 + t^2 - 3t$, $g = t - 3$:

$$\begin{array}{l}
\big(\ \ t^6 - 3t^5 + t^2 - 3t\,\big) : \big(t - 3\big) = t^5 + t \\
\underline{-\,t^6 + 3t^5} \\
\qquad\qquad\ \ t^2 - 3t \\
\qquad\quad\ \ \underline{-\,t^2 + 3t} \\
\qquad\qquad\qquad\ \ 0
\end{array}$$

In diesem Fall geht die Division ohne Rest auf, d.h. wir haben

$$q = t^5 + t, \quad r = 0.$$

Es gilt also $f = q \cdot g$. Dies hilft uns, die Nullstellen von f zu finden: Da 3 eine Nullstelle von g ist, ist 3 auch eine Nullstelle von f. Die weiteren Nullstellen von f sind genau die Nullstellen von q, also von $q = t^5 + t = t(t^4 + 1)$. Hier sehen wir das q nur eine reelle Nullstelle (nämlich 0) hat. Insgesamt haben wir f faktorisiert als

$$f = t(t^4 + 1)(t - 3),$$

was es uns erlaubt, alle Nullstellen von f abzulesen. △

Kommen wir nun zur genaueren Untersuchung von *Nullstellen* von Polynomen. Dabei heißt $\lambda \in K$ eine Nullstelle von $f \in K[t]$, falls $f(\lambda) = 0$ gilt.

Beispiel.

1. Polynome vom Grad 0 haben keine Nullstellen. Es ist ja $f = a$ mit $a \neq 0$ (für $a = 0$ ist der Grad ja $-\infty$).
2. Polynome von Grad 1, also Polynome der Form $f = at + b$, haben immer genau eine Nullstelle, nämlich $\lambda = -b/a$, denn $f(\lambda) = a\left(-\frac{b}{a}\right) + b = -b + b = 0$.
3. Das Polynom $f = t^2 + 1$ vom Grad 2 hat

- in $\mathbb{R}[t]$ keine Nullstelle, denn für $t \in \mathbb{R}$ gilt stets
$$f(t) = t^2 + 1 \geq 1 > 0.$$
- in $\mathbb{C}[t]$ die beiden Nullstellen $\lambda_{1/2} = \pm i$, denn
$$f(\pm i) = (\pm i)^2 + 1 = -1 + 1 = 0.$$
- in $\mathbb{Z}_2[t]$ genau eine Nullstelle, denn
$$f(0) = 0^2 + 1 = 1, \quad f(1) = 1^2 + 1 = 0.$$

  Merke: Die Anzahl der Nullstellen hängt auch vom zu Grunde liegenden Körper ab.

4. Das Polynom $f = t^2 - 2$ vom Grad 2 hat in $\mathbb{R}[t]$ und $\mathbb{C}[t]$ die beiden Nullstellen $\lambda_{1/2} = \pm\sqrt{2}$, über $\mathbb{Q}[t]$ jedoch keine Nullstelle.
5. Ist K ein Körper mit endlich vielen Elementen $\{a_1, \ldots, a_n\}$, so hat das Polynom
$$f = (t - a_1) \cdots (t - a_n)$$
den Grad n und auch n Nullstellen (nämlich alle Elemente von K), das Polynom
$$g = (t - a_1) \cdots (t - a_n) + 1$$
hingegen hat auch den Grad n, aber keine Nullstelle.

$\triangle$

Im Allgemeinen ist das Bestimmen von Nullstellen von Polynomen schwierig, auch über vertrauten Körpern wie $\mathbb{R}$ (und über $\mathbb{C}$ etwas einfacher, aber trotzdem schwierig). Es gibt explizite Formeln für die Nullstellen von Polynomen vom Grad 1, 2, 3 und 4, die nur zweite und dritte Wurzeln benutzen. Für den Grad 2 (quadratische Funktionen) ist das die bekannte pq-Formel. Dass diese Formel auch ebenso für komplexe Polynome vom Grad 2 gilt, sollten Sie in den Übungen beweisen. Für den Grad 3 und 4 gibt es die ziemlich komplizierten Formeln von Cardano (aus dem 16. Jahrhundert).

Die (ggf. komplexen) Nullstellen eines reellen Polynoms $x^2 + px + q$ sind
$$-\tfrac{p}{2} \pm \sqrt{\tfrac{p^2}{4} - q}.$$

Dass es für Grad > 4 im Allgemeinen *gar keine* Formeln mehr gibt, die die Nullstellen in elementaren Rechenoperationen und Wurzeln ausdrücken (also auch nicht mit höheren Wurzel-Ausdrücken), besagt der Satz von Abel-Ruffini (aus dem 19. Jahrhundert). Ein genaueres Studium der Bestimmung von Nullstellen von Polynomen führt in das Gebiet der Galois-Theorie, einem Teilgebiet der Gruppentheorie.

3 Abspalten von Linearfaktoren

Haben wir eine Nullstelle eines Polynoms gefunden, so genügt es für die Suche nach weiteren Nullstellen, ein Polynom von geringerem Grad zu betrachten. Dies wird durch die Polynomdivision aus Satz 2.1 möglich:

Lemma 3.1. *Ist $\lambda \in K$ eine Nullstelle von $f \in K[t]$, so gibt es ein eindeutig bestimmtes Polynom g mit $\deg(g) = \deg(f) - 1$ und $f = (t - \lambda) \cdot g$.*

Beweis.
Wir dividieren f nach Satz 2.1 mit Rest durch $(t - \lambda)$. Es gibt also eindeutig bestimmte q, r mit $f = (t - \lambda)q + r$ und $\deg(r) < \deg(t - \lambda) = 1$. Das heißt aber, dass r ein konstantes Polynom ist, also $r = a_0$ für ein $a_0 \in K$.

Da λ eine Nullstelle von f ist, folgt

$$0 = f(\lambda) = (\lambda - \lambda) \cdot g(\lambda) + a_0 = a_0,$$

also ist $r = 0$. Die Formel für den Grad folgt aus

$$\deg(f) = \deg(t - \lambda) + \deg(g) = 1 + \deg(g).$$

□

Das Abspalten von Nullstellen, also die Faktorisierung $f = (t - \lambda) \cdot g$ hat folgendes Lemma zur Folge:

Lemma 3.2. *Es sei K ein Körper und $f \in K[t]$. Ist die Anzahl der Nullstellen von f gleich m und ist f nicht das Nullpolynom, so folgt $m \leq \deg(f)$.*

Beweis.
Wir führen Induktion über den Grad von f aus: Ist $\deg(f) = 0$, so gilt $f = a_0 \neq 0$. Die Anzahl der Nullstellen ist $m = 0$, und damit ist die Behauptung richtig.

Sei nun $\deg(f) = n \geq 1$ und die Aussage gelte für Polynome vom Grad $\leq n - 1$. Hat f keine Nullstelle, dann stimmt die behauptete Ungleichung. Hat f die Nullstelle λ, so können wie diese nach Lemma 3.1 abspalten, d.h. es gibt g mit

$$f = (t - \lambda)g, \quad \text{und} \quad \deg(g) = n - 1.$$

Wir sehen, dass alle Nullstellen von f, die ungleich λ sind, auch Nullstellen von g sind. Schreiben wir l für die Anzahl der Nullstellen von g, so ist nach Induktionsannahme $l \leq n - 1$ und daher $m \leq l + 1 \leq n$. □

Eine weitere Folgerung ist die Tatsache, dass es auf unendlichen Körpern K keinen wesentlichen Unterschied zwischen Polynomen und den zugehörigen Funktionen auf K gibt. Es gilt nämlich:

Korollar 3.3. *Ist K unendlich, dann ist die Abbildung*

$$\tilde{}: K[t] \to \mathrm{Abb}(K, K),$$

welche einem Polynom die entsprechende Funktion von K nach K zuordnet, injektiv.

Seien $f_1, f_2 \in K[t]$ und $g = f_1 - f_2$. Gilt $\tilde{f_1} = \tilde{f_2}$, so ist $\tilde{g} = 0$, also $g(\lambda) = 0$ für alle $\lambda \in K$. Nach Lemma 3.2 geht dies nur für das Nullpolynom, also $g = 0$, und das heißt $f_1 = f_2$.

Ist λ eine Nullstelle von f, so ist $f = (t - \lambda) \cdot g$. Nun kann es sein, dass λ wiederum eine Nullstelle von g ist; in diesem Fall ist λ eine *mehrfache Nullstelle* von f.

Ein einfaches Beispiel: $f = t^2 - 2t + 1$ hat die Nullstelle $\lambda = 1$ und es gilt $f = (t-1)(t-1)$, also $f = (t-1)g$ mit $g = t - 1$.

Definition 3.4. Ist $f \in K[t]$, $f \neq 0$ und $\lambda \in K$, so ist

$$\mathrm{ord}(f, \lambda) := \max\{r \in \mathbb{N} \ : \ f = (t - \lambda)^r g \text{ für ein Polynom } g\}$$

die *Vielfachheit* oder *Ordnung* der Nullstelle.

Beispiel.
Für das Polynom $f = t^4 - 2t^3 + 2t^2 - 2t + 1 \in \mathbb{R}[t]$ gilt

$$f = (t-1)^2(t^2+1),$$

also gilt $\mathrm{ord}(f, 1) = 2$. △

Es gilt

$$\mathrm{ord}(f, \lambda) = 0 \iff f(\lambda) \neq 0.$$

Ist $f = (t - \lambda)^r \cdot g$ und $r = \mathrm{ord}(f, \lambda)$, dann ist (nach Definition der Vielfachheit), λ keine Nullstelle von g. Mit anderen Worten: Die Vielfachheit einer Nullstelle gibt an, wie oft der *Linearfaktor* $(t - \lambda)$ in f enthalten ist.

Habe nun $f \in K[t]$ den Grad n, genau die Nullstellen $\lambda_1, \dots, \lambda_k \in K$ und die Vielfachheiten seien $r_i = \mathrm{ord}(f, \lambda_i)$, $i = 1, \dots, k$. Dann ist

$$f = (t - \lambda_1)^{r_1} \cdots (t - \lambda_k)^{r_k} \cdot g,$$

und das Polynom g hat den Grad $n - r_1 - \cdots - r_k$ und keine weiteren Nullstellen. Die beiden Extremfälle sind $\deg(g) = n$ (also $g = f$, d.h. f selbst hat keine Nullstellen) und $\deg(g) = 0$. Im zweiten Fall gilt

$$f = a_n(t - \lambda_1)^{r_1} \cdots (t - \lambda_k)^{r_k},$$

für ein $a_n \neq 0$. Wir sagen in diesem Fall: *f zerfällt in Linearfaktoren.*

Ein Hauptergebnis über Nullstellen von Polynomen haben wir schon im ersten Semester benutzt:

Satz 3.5 (Fundamentalsatz der Algebra). *Jedes Polynom $f \in \mathbb{C}[t]$ mit $\deg(f) > 0$ hat mindestens eine Nullstelle.*

Auch hier geben wir keinen Beweis, sondern vertrösten auf die Vorlesung „Funktionentheorie".

Eine direkte Folgerung aus diesem Satz ist:

Eine kleine Übersicht über klassische Beweise dieses Satzes findet man im Kapitel „Fundamentalsatz der Algebra" im Buch *Zahlen* von Ebbinghaus et. al.

Korollar 3.6. *Jedes Polynom $f \in \mathbb{C}[t]$ vom Grad $n \geq 1$ zerfällt in Linearfaktoren, d.h. es gibt $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ mit $n = \deg(f)$, so dass*

$$f = a(t - \lambda_1) \cdots (t - \lambda_n).$$

Beachte: Die λ_i dürfen durchaus gleich sein, d.h. die gleiche Nullstelle kann mehrfach auftreten.

Da f nach dem Fundamentalsatz eine Nullstelle λ hat, können wir diese abspalten und bekommen $f = (t - \lambda) \cdot g$ mit $\deg(g) = \deg(f) - 1$. Ist $\deg(g) > 0$ können wir eine weitere Nullstelle abspalten, und so weiter, bis wir n Nullstellen abgespalten haben.

Über $\mathbb{R}$ kann ein solcher Satz nicht richtig sein: Das Polynom $f = t^2 + 1$ hat keine Nullstellen in $\mathbb{R}$ und wir können daher keinen Linearfaktoren abspalten. Es lässt sich trotzdem noch etwas mehr sagen.

Wie üblich betrachten wir $\mathbb{R}$ als Teilmenge von $\mathbb{C}$ und ebenso auch $\mathbb{R}[t]$ als Teilmenge von $\mathbb{C}[t]$. Letzteres heißt nichts anderes, als dass wir Polynome mit reellen Koeffizienten auch als Polynome in $\mathbb{C}[t]$ auffassen können. In diesem Sinn hat das reelle Polynom $f = t^2 + 1$, aufgefasst als Polynom in $\mathbb{C}[t]$ die beiden Nullstellen $\pm i$.

Lemma 3.7. *Ist $f \in \mathbb{R}[t]$ (bzw. $f \in \mathbb{C}[t]$ mit reellen Koeffizienten) und ist $\lambda \in \mathbb{C}$ eine Nullstelle von f, so ist auch $\bar{\lambda}$ eine Nullstelle von f.*

Erinnere: $\bar{z}$ ist die komplex konjugierte Zahl, welche aus z entsteht, indem der Imaginärteil das Vorzeichen wechselt, also für $z = x + iy$ ist $\bar{z} = x - iy$.

Beweis.
Ist z eine Nullstelle von f, so gilt (da für die Koeffizienten $a_i = \bar{a}_i$ gilt)

$$\begin{aligned} f(\bar{z}) &= a_0 + a_1\bar{z} + \cdots + a_n\bar{z}^n \\ &= \overline{a_0} + \overline{a_1}\,\overline{z} + \cdots + \overline{a_n}\,\overline{z^n} \\ &= \overline{a_0 + a_1 z + \cdots + a_n z^n} \\ &= \overline{f(z)} = 0. \end{aligned}$$

D.h. $\bar{z}$ ist auch eine Nullstelle von f. □

Lemma 3.8. *Sei $f \in \mathbb{R}[t]$ und $\lambda \in \mathbb{C}$ eine nicht-reelle Nullstelle von f. Dann gilt*

i) $g := (t - \lambda)(t - \bar{\lambda}) \in \mathbb{R}[t]$ *und*

ii) *es gibt ein $q \in \mathbb{R}[t]$ mit $f = g \cdot q$.*

Beweis.
Für i) ist zu zeigen, dass das Polynom g trotz des Auftretens der komplexen Zahlen nur reelle Koeffzienten hat. Dazu schreiben

wir $\lambda = \alpha + i\beta$ und rechnen

$$\begin{aligned}(t-\lambda)(t-\bar\lambda) &= (t-\alpha-i\beta)(t-\alpha+i\beta)\\ &= t^2 - t\alpha + it\beta - t\alpha + \alpha^2 - i\alpha\beta - it\beta + i\alpha\beta + \beta^2\\ &= t^2 - 2\alpha t + \alpha^2 + \beta^2.\end{aligned}$$

Für ii) dividieren wir f durch g mit Rest, d.h. es gibt zwei Polynome $q, r \in \mathbb{R}[t]$ mit

$$f = g\cdot q + r, \quad \text{mit} \quad \deg(r) < \deg(g) = 2.$$

Die Gleichung der Polynome impliziert, dass auch für komplexe Argumente z gilt $f(z) = g(z)q(z) + r(z)$. Einsetzen von λ und $\bar\lambda$ gibt $r(\lambda) = r(\bar\lambda) = 0$. Da r höchstens den Grad 1 hat, folgt mit Lemma 3.2, dass $r = 0$ gilt. □

Lemma 3.9. *Für $f \in \mathbb{R}[t]$ gilt für jedes paar $\lambda, \bar\lambda$ von konjuguerten Nullstellen*

$$\operatorname{ord}(f,\lambda) = \operatorname{ord}(f,\bar\lambda).$$

Beweis.
Wir zeigen die Behauptung durch Induktion über k mit $k \leq \operatorname{ord}(f,\lambda)$. Dafür zeigen wir, dass wir f schreiben können als $f = g^k f_k$ mit dem Polynom $g = (t-\lambda)(t-\bar\lambda)$ und einem $f_k \in \mathbb{R}[t]$.

Im Fall $k = 0$ ist nichts zu zeigen (dann ist $g^0 = 1$ und $f = f_k$ tut es).

Sei also jetzt $\operatorname{ord}(f,\lambda) \geq k+1$ und die Behauptung für k richtig. Dann ist also $f = g^k f_k$ und es gilt $f_k(\lambda) = 0$. Und nach Lemma 3.8 können wir f_k schreiben als $f_k = g\cdot f_{k+1}$ mit $f_{k+1} \in \mathbb{R}[t]$. Es folgt $f = g^{k+1} f_{k+1}$, wie behauptet. □

Zusammen haben wir folgenden Satz über Nullstellen von reellen Polynomen bewiesen:

Satz 3.10 (Fundamentalsatz für reelle Polynome)**.** *Jedes Polynom $f \in \mathbb{R}[t]$ mit $\deg(f) = n \geq 1$ lässt sich zerlegen als*

$$f = a(t-\lambda_1)\cdots(t-\lambda_r)\cdot g_1\cdots g_m$$

wobei $a, \lambda_1, \dots, \lambda_r \in \mathbb{R}$ und die $g_1, \dots, g_m \in \mathbb{R}[t]$ normierte Polynome mit Grad 2 sind, welche keine reellen Nullstellen haben. Insbesondere ist $n = r + 2m$.

Korollar 3.11. *Jedes Polynom $f \in \mathbb{R}[t]$ mit ungeradem Grad hat mindestens eine reelle Nullstelle.*

Folgt aus der Gleichung $n = r + 2m$: Ist n ungerade, so muss $r \geq 1$ sein.

Diese Aussage haben wir hier rein algebraisch bewiesen. Man kann die Aussage auch mit Methoden der Analysis beweisen: Ist der Grad ungerade, nehmen wir (ohne Einschränkung) an, dass der führende Koeffizient positiv ist. Dann nimmt das Polynom für große positive Argumente positive und für große negative Argumente negative Werte an. Nach dem Zwischenwertsatz hat es also eine Nullstelle.

4 Invariante Fahnen und Trigonalisierbarkeit

Wir erinnern an den Begriff der Diagonalisierbarkeit und an uns bisher bekannte Eigenschaften:

- Eine lineare Abbildung $f : V \to V$ heißt diagonalisierbar, wenn es eine Basis B von V gibt, so dass $M_B^B(f)$ eine Diagonalmatrix ist.
- Entsprechend nennen wir eine Matrix $A \in K^{n\times n}$ diagonalsierbar, wenn es eine invertierbare Matrix $S \in K^{n\times n}$ gibt, so dass $S^{-1}AS$ eine Diagonalmatrix ist.
- Ist eine lineare Abbildung (bzw. Matrix) diagonalisierbar, so muss das charakteristische Polynom in Linearfaktoren zerfallen.
- Ist eine lineare Abbildung (bzw. Matrix) diagonalisierbar, so muss die Dimension der Eigenräume gleich den Vielfachheiten der Eigenwerte sein, d.h. wir haben, dass geometrische und algebraische Vielfachheit übereinstimmen.
- Eine lineare Abbildung (bzw. Matrix) ist genau dann diagonalisierbar, wenn es eine Basis aus Eigenvektoren von f gibt, oder anders gesagt, wenn die Summe der Dimensionen der Eigenräume gleich der Dimension des ganzen Raumes ist.

Das ist die gleiche Aussage wie im ersten Punkt: In S steht dann die Basis B spaltenweise.

Das ist nur eine notwendige Bedingung.

Wir wollen nun untersuchen, was man für nicht-diagonalisierbare Abbildungen noch aussagen kann.

Wir fragen zuerst, wann eine lineare Selbstabbildung durch eine obere Dreicksmatrix dargestellt werden kann:

Definition 4.1. Eine lineare Abbildung $f : V \to V$ heißt *trigonalisierbar*, wenn es eine Basis B von V gibt, so dass $M_B^B(f)$ eine obere Dreiecksmatrix ist.

Ebenso heißt eine Matrix $A \in K^{n\times n}$ trigonalisierbar, wenn es ein invertierbares $S \in K^{n\times n}$ gibt, so dass $S^{-1}AS$ eine obere Dreiecksmatrix ist.

Um Trigonalisierbarkeit besser zu verstehen, untersucht man die Unterräume von V, die von f in sich selbst abgebildet werden.

Definition 4.2. Es sei $f : V \to V$ linear. Wir nennen einen Unterraum $W \subset V$ f-invariant, wenn $f(W) \subset W$.

Entsprechend sprechen wir von A-invarianten Unterräumen, wenn die eine lineare Abbildung durch eine Matrix A gegeben ist.

Definition 4.3. Ein Vektorraum V ist die *direkte Summe* der Unterräume $W_1, \ldots, W_k$, geschrieben

$$V = W_1 \oplus \cdots \oplus W_k,$$

wenn gilt

1. $V = W_1 + \cdots + W_k$ und

Diese Bedingungen sind äquivalent dazu, dass sich jedes $v \in V$ *eindeutig* als Summe $v = w_1 + \cdots + w_k$ mit $w_i \in W_i$ schreiben lässt.
Ist $W \subset V$ und gibt es auf V ein Skalarprodukt, so ist $V = W \oplus W^\perp$.

2. Sind $w_1 \in W_1$, ..., $w_k \in W_k$ mit $w_1 + \dots + w_k = 0$, so folgt $w_1 = \dots = w_k = 0$.

Beispiel.

- Für jedes f sind $\{0\}$ und V f-invariant.
- Ist f diagonalisierbar und $\{v_1, \dots, v_n\}$ eine Basis aus Eigenvektoren, so sind die Räume
$$W_i = \langle \{v_i\} \rangle$$
f-invariant und wir können V zerlegen als
$$V = W_1 \oplus \dots \oplus W_n.$$
- Ein Eigenraum $\operatorname{Eig}(f, \lambda)$ ist ebenfalls f-invariant, da für $v \in \operatorname{Eig}(f, \lambda)$ gilt $f(v) = \lambda v \in \operatorname{Eig}(f, \lambda)$.
- Für diagonalisierbares f gibt es auch folgende Zerlegung in f-invariante Unterräume
$$V = \operatorname{Eig}(f, \lambda_1) \oplus \dots \oplus \operatorname{Eig}(f, \lambda_k)$$
(wenn wir über alle Eigenwerte summieren).
- Ist $A \in K^{n \times n}$ eine obere Dreiecksmatrix, so sind die Unterräume
$$W_r = \langle \{e_1, \dots, e_r\} \rangle$$
für alle $r \in \{1, \dots, n\}$ A-invariant.

Es gilt $x \in W_r$ genau dann, wenn $x_k = 0$ für $k \geq r$. In diesem Fall gilt aber auch für $y = Ax$, dass $y_k = 0$ für $k \geq r$.

△

Lemma 4.4. *Ist $W \subset V$ ein f-invarianter Unterraum, so ist das charakteristische Polynom $\chi_{f|_W}$ von $f|_W$ ein Teiler des charakteristischen Polynoms χ_f.*

Beweis.
Es sei B' eine Basis von W und wir ergänzen diese zu einer Basis B von V. Dann gilt für die Darstellungsmatrix
$$M_B^B(f) = \begin{pmatrix} M_{B'}^{B'}(f|_W) & * \\ 0 & A \end{pmatrix}.$$
(Die Null unten links folgt, da $f(W) \subset W$.)

Nach den Eigenschaften der Determinante folgt, dass $\chi_f = \chi_{f|_W} \cdot \chi_A$. □

Beispiel.

Wir betrachten die lineare Abbildung, die durch die Matrix

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

gegeben ist. Ihr charakteristisches Polynom ist $t^2 - 2$. Dieses Polynom hat in $\mathbb{Q}[t]$ keine Teiler mit Grad ≥ 1 (sonst hätte es eine Nullstelle in $\mathbb{Q}$). Daher gibt es wegen Lemma 4.4 keinen A-invarianten Unterraum von $\mathbb{Q}^2$. △

Definition 4.5. Eine *Fahne* eines n-dimensionalen Vektorraumes V ist eine Folge von Unterräumen V_r, $r = 0, \ldots, n$ mit $\dim(V_r) = r$ und der Eigenschaft

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V.$$

Eine Fahne heißt *f-invariant* (für eine lineare Abbildung $f : V \to V$), wenn jedes V_r f-invariant ist.

Beispiel.

- Im vorigen Beispiel bilden die W_r eine A-invariante Fahne.
- Für eine obere Dreiecksmatrix A (wie im Beispiel auf Seite 16) bilden die Räume $W_r = \langle\{e_1, \ldots, e_r\}\rangle$ eine Fahne.

△

Hat man eine f-invariante Fahne, so folgt aus $f(V_1) \subset V_1$ und $\dim(V_1) = 1$, dass es einen Eigenvektor von f geben muss.

Aus $f(V_2) \subset V_2$ folgt das nicht! Wieso?

Lemma 4.6. *Für eine lineare Abbildung $f : V \to V$ sind äquivalent:*

i) Es gibt eine f-invariante Fahne.

ii) f ist trigonalisierbar.

Beweis.

i) $\Longrightarrow$ ii): Sei $V_0, \ldots, V_n$ eine f-invariante Fahne. Dann ist jedes $v_1 \in V_1$ eine Basis von V_1 und wir ergänzen dieses sukzessive zu einer Basis $B = \{v_1, \ldots, v_n\}$ von V, so dass $\{v_1, \ldots, v_r\}$ jeweils eine Basis von V_r ist. Wegen $f(V_r) \subset V_r$ lässt sich jedes $v \in f(V_r)$ als Linearkombination der $v_1, \ldots, v_r$ schreiben und daher ist $M_B^B(f)$ eine obere Dreiecksmatrix.

ii) $\Longrightarrow$ i): Ist umgekehrt B eine Basis, so dass $M_B^B(f)$ eine obere Dreiecksmatrix ist, so setzen wir wiederum $V_r = \langle\{v_1, \ldots, v_r\}\rangle$. Dann ist $\dim(V_r) = r$ und es gilt $V_1 \subset \cdots \subset V_n = V$. Da $M_B^B(f)$ eine obere Dreicksmatrix ist, muss jedes $v \in f(V_r)$ eine Linearkombination der $v_1, \ldots, v_r$ sein, d.h. es gilt $f(V_r) \subset V_r$. □

Satz 4.7 (Trigonalisierungssatz)**.** *Es sei V ein n-dimensionaler K-Vektorraum und $f : V \to V$ eine lineare Abbildung. Dann ist äquivalent:*

i) f *ist trigonalisierbar.*

ii) Das charakteristische Polynom χ_f *zerfällt über* K *in Linearfaktoren, d.h. es gibt* $\lambda_1, \dots, \lambda_n \in K$ *(nicht notwendigerweise verschieden), so dass*

$$\chi_f = \pm(t - \lambda_1) \cdots (t - \lambda_n).$$

Beweis.
i) $\implies$ **ii):** Ist $A = M_B^B(f)$ eine obere Dreiecksmatrix mit Diagonaleinträgen a_{ii}, so ist das charakteristische Polynom

$$\chi_f = \chi_A = (a_{11} - t) \cdots (a_{nn} - t).$$

ii) $\implies$ **i):** Wir zeigen, dass es eine f-invariante Fahne gibt und machen dazu eine Induktion über $n = \dim(V)$:

Für $n = 1$ ist nichts zu zeigen. Sei also $n \geq 2$. Da f nach Annahme den Eigenwert λ_1 hat, gibt es einen zugehörigen Eigenvektor v_1 und wir ergänzen diesen zu einer Basis $B = \{v_1, w_2, \dots, w_n\}$. Mit $V_1 = \langle\{v_1\}\rangle$ und $W = \langle\{w_2, \dots, w_n\}\rangle$ gilt dann

$$V = V_1 \oplus W$$

und

$$M_B^B(f) = \begin{pmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & & & \\ \vdots & & \tilde{A} & \\ 0 & & & \end{pmatrix}$$

mit $\tilde{A} \in K^{n-1 \times n-1}$ mit den Einträgen $(a_{ij})_{i,j=2,\dots,n}$. Der Raum V_1 ist per Definition f-invariant.

Jetzt definieren wir die linearen Abbildungen $h : W \to V_1$ und $g : W \to W$ durch

$$h(w_j) = a_{1j} v_1, \quad g(w_j) = a_{2j} w_2 + \cdots + a_{nj} w_n.$$

Wir beobachten, dass für $w \in W$ gilt $f(w) = h(w) + g(w)$. Da $\tilde{A}$ die Darstellungsmatrix von g in der Basis $\{w_2, \dots, w_n\}$ von W ist, folgt für die charakteristischen Polynome von f und g, dass

$$\chi_f = (\lambda_1 - t)\chi_g.$$

Nach unserer Annahme folgt also $\chi_g = (\lambda_2 - t) \cdots (\lambda_n - t)$, d.h. das charakteristische Polynom von g zerfällt in Linearfaktoren und wir können die Induktionsvoraussetzung von g anwenden. D.h. es existiert eine g-invariante Fahne in W, d.h. Unterräume

$$\{0\} = W_0 \subset W_1 \subset \cdots \subset W_{n-1} = W.$$

Wir machen daraus eine Fahne im größeren Raum V wie folgt. Den Unterraum V_1 haben wir schon und wir definieren

$$V_r = V_1 + W_{r-1}, \quad r = 2, \dots, n.$$

Es bleibt zu zeigen, dass diese Fahne f-invariant ist. Dazu sei $v \in V_r$, d.h. $v = \mu v_1 + w_r$ mit $\mu \in K$ und $w_r \in W_{r-1}$. Dann ist

$$\begin{aligned} f(v) &= f(\mu v_1 + w_r) = \mu f(v_1) + f(w_r) \\ &= \underbrace{\mu \lambda_1 v_1 + h(w_r)}_{\in V_1} + \underbrace{g(w_r)}_{\in W_{r-1}} \in V_1 + W_{r-1} = V_r. \end{aligned}$$

□

5 Berechnen von Trigonalisierungen

Ein Korollar aus dem Satz 4.7 ist es noch wert, festgehalten zu werden:

Korollar 5.1. *Jede lineare Abbildung eines endlichdimensionalen $\mathbb{C}$-Vektorraumes in sich selbst ist trigonalisierbar.*

Jedes Polynom über $\mathbb{C}$ zerfällt in Linearfaktoren.

Über anderen Körpern als $\mathbb{C}$ zerfällt nicht jedes Polynom in Linearfaktoren, also ist auch nicht jede Matrix trigonalisierbar. Wir illustrieren dies am Beispiel der Matrix

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

Betrachten wir sie zuerst über dem Vektorraum $\mathbb{R}^2$:

Beispiel.

Das charakteristische Polynom von A ist $\chi_A = t^2 + 1$. Diese zerfällt über $\mathbb{R}$ nicht in Linearfaktoren (das Polynom hat keine Nullstelle). Daher ist A nicht trigonalisierbar, d.h. *für jedes invertierbare $S \in \mathbb{R}^{2\times 2}$ ist $S^{-1}AS$ keine obere Dreiecksmatrix.* $\triangle$

Betrachten wir die gleiche Matrix als Matrix in $\mathbb{C}^{2\times 2}$:

Beispiel.

Hier zerfällt das charakteristische Polynom in Linearfaktoren, nämlich

$$\chi_A = t^2 + 1 = (t-i)(t+i).$$

Es gibt also zwei verschiedene Eigenwerte $\lambda_1 = i$, $\lambda_2 = -i$, also auch zwei linear unabhängige Eigenvektoren, z.B.

$$v_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix} \text{ und } v_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Man rechnet schnell nach, dass

$$Av_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} i \\ 1 \end{pmatrix} = i \begin{pmatrix} 1 \\ -i \end{pmatrix} = iv_1$$

und

$$Av_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} -i \\ 1 \end{pmatrix} = -i \begin{pmatrix} 1 \\ i \end{pmatrix} = -iv_2.$$

Insbesondere ist A über $\mathbb{C}$ trigonalisierbar (sogar diagonalisierbar) und mit

$$S = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \quad S^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2}i \\ \frac{1}{2} & -\frac{1}{2}i \end{pmatrix}$$

gilt dann

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

$\triangle$

Schauen und wir uns noch einen dritten Fall an, nämlich den Vektorraum $\mathbb{Z}_2^2$:

Beispiel.

Wir rechnen nun über $\mathbb{Z}_2$, wo wir insbesondere $1+1=0$ und $-1=1$ berücksichtigen müssen. Insbesondere ist

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

In diesem Fall hat $\chi_A(t) = t^2+1$ eine Nullstelle, nämlich $\lambda = 1$, denn es gilt $\chi_A(1) = 1\cdot 1+1 = 1+1 = 0$. Eine weitere Nullstelle gibt es nicht, da $\chi_A(0) = 0+1$. Was heißt das für die Linearfaktoren? Wenn wir t^2+1 über $\mathbb{Z}_2$ durch den Linearfaktor $t-1 = t+1$ dividieren, so bekommen wir $(t^2+1) : (t+1) = t+1$, also gilt über $\mathbb{Z}_2$, dass $t^2+1 = (t+1)^2$ (was man schnell prüft: $(t+1)^2 = t^2+t+t+1 = t^2+(1+1)t+1 = t^2+1$).

Bestimmen wir nun den Eigenraum zum Eigenwert 1. Wir suchen also den Kern von

$$A - E_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Diese Matrix hat Rang 1 und ihr Kern wird aufgespannt von $v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$. Der Eigenraum ist also eindimensional, insbesondere ist die geometrische Vielfachheit des Eigenwertes 1 nur 1 und damit echt kleiner als die algebraische Vielfachheit.

Wir können also schließen, dass A über $\mathbb{Z}_2$ zwar trigonalisierbar ist (da das charakteristische Polynom in Linearfaktoren zerfällt), aber nicht diagonalisierbar.

Um eine Trigonalisierung zu bestimmen, ergänzen wir v_1 zu einer Basis. Wir nehmen $v_2 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ und schreiben v_1, v_2 als Spalten in eine Matrix

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

Diese hat (über $\mathbb{Z}_2$ gerechnet) die inverse Matrix

$$S^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

(was man schnell prüft: $SS^{-1} = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0+1 & 1+1 \\ 0+0 & 1+0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$). Wir bekommen damit eine Trigonalisierung:

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

△

Nun bemerken wir noch, dass die Beweismethode von Satz 4.7

zwar immer eine obere Dreiecksmatrix liefert, es aber verpassen kann, eine noch „bessere" Darstellung zu finden:

Beispiel.
Wir betrachten die Matrix

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

(als lineare Abbildung von $\mathbb{R}^2$ in sich). Wir wählen als $v_1 = e_1$ und haben den zugehörigen Eigenwert $\lambda_1 = 1$. Wir ergänzen nun beliebig zu einer Basis, z.B. mit $w_2 = \left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right)$. Die Darstellungsmatrix von D in der Basis $\{v_1, w_2\}$ ist also $S^{-1}DS$ mit $S = \left(\begin{smallmatrix} v_1 & w_2 \end{smallmatrix}\right)$, also

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

Das ist zwar eine obere Dreiecksmatrix, aber keine Diagonalmatrix mehr.

Hätten wir mit $w_2 = e_2$ ergänzt, so wäre $S = E_2$, und $S^{-1}DS = D$ eine Diagonalmatrix. △

Kommen wir nun dazu, wie man Trigonalisierungen systematisch ausrechnen kann: Hierzu betrachten wir den Fall von Matrizen, d.h. wir haben $A \in K^{n\times n}$ mit charakteristischem Polynom $\chi_A = \pm(t-\lambda_1)\cdots(t-\lambda_n)$ und suchen ein $S \in GL_n(K)$, so dass

$$D = S^{-1}AS$$

eine obere Dreiecksmatrix ist.

Das allgemeine Vorgehen ist wie folgt:

1. Berechne einen Eigenvektor v_1 zum Eigenwert λ_1, ergänze diesen zu Basis und schreibe diese in eine Matrix $S_1 = \begin{pmatrix} v_1 & w_2 & \cdots & w_n \end{pmatrix} \in K^{n\times n}$.

2. Bilde
$$S_1^{-1}AS_1 = \begin{pmatrix} \lambda_1 & r_1 \\ 0 & A_1 \end{pmatrix}$$
mit $A_1 \in K^{n-1\times n-1}$ und $r_1 \in K^{1\times n-1}$.

3. λ_2 ist auch Eigenwert von A_1, berechne einen zugehörigen Eigenvektor $v_2 \in K^{n-1}$, ergänze diesen zu Basis des K^{n-1} und schreibe diese in eine Matrix $\tilde{S}_2 = \begin{pmatrix} v_2 & w_3 & \cdots w_n \end{pmatrix} \in K^{n-1\times n-1}$.

4. Bilde
$$S_2 = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{S}_2 \end{pmatrix}$$
und
$$S_2^{-1}S_1^{-1}AS_1S_2 = S_2^{-1}\begin{pmatrix} \lambda_1 & r_1 \\ 0 & A_1 \end{pmatrix} S_2 = \left(\begin{array}{c|cc} \lambda_1 & \multicolumn{2}{c}{\tilde{r}_1} \\ \hline 0 & \lambda_2 & r_2 \\ 0 & 0 & A_2 \end{array}\right)$$

mit $A_2 \in K^{n-2\times n-2}$ und $r_2 \in K^{1\times n-2}$.

5. Fahre mit A_2 entsprechend fort und setze schließlich $S = S_1 S_2 \cdots S_n$. Wir bekommen
$$S^{-1}AS = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Beispiel.

Es sei $V = \mathbb{R}^3$ und
$$A = \begin{pmatrix} 3 & 4 & 3 \\ -1 & 0 & -1 \\ 1 & 2 & 3 \end{pmatrix}.$$
Das charakteristische Polynom ist
$$\begin{aligned}
\chi_A &= \det\left(\begin{pmatrix} 3-\lambda & 4 & 3 \\ -1 & -\lambda & -1 \\ 1 & 2 & 3-\lambda \end{pmatrix}\right) \\
&= (3-\lambda)\det\left(\begin{pmatrix} -\lambda & -1 \\ 2 & 3-\lambda \end{pmatrix}\right) - 4\det\left(\begin{pmatrix} -1 & -1 \\ 1 & 3-\lambda \end{pmatrix}\right) \\
&\quad + 3\det\left(\begin{pmatrix} -1 & -\lambda \\ 1 & 2 \end{pmatrix}\right) \\
&= (3-\lambda)\Big(-\lambda(3-\lambda)+2\Big) - 4\Big(-(3-\lambda)+1\Big) + 3\Big(-2+\lambda\Big) \\
&= -\lambda^3 + 6\lambda^2 - 12\lambda + 8 \\
&= -(\lambda-2)^3.
\end{aligned}$$
Wir haben also $\lambda_1 = \lambda_2 = \lambda_3 = 2$. Der Rang von
$$A - 2E_3 = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -2 & -1 \\ 1 & 2 & 1 \end{pmatrix}$$
ist 2 und daher ist die Dimension des Eigenraumes $\mathrm{Eig}(A,2)$ nur 1. Die algebraische Vielfachheit des Eigenwertes 2 ist hingegen 3 und wir schließen, dass A nicht diagonalisierbar ist.

Berechnen wir einen Eigenvektor von A zum Eigenwert 2, d.h. einen Vektor im Kern von $A - 2E_3$. Wir bekommen z.B.
$$v_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$
Wir können diesen durch Standard-Basisvektoren zu einer Basis ergänzen. Wir nehmen (willkürlich) die Basis $\{v_1, e_2, e_3\}$. Diese schreiben wir in eine Matrix
$$S_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

mit inverser Matrix

$$S_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Wir berechnen

$$S_1^{-1}AS_1 = \begin{pmatrix} 2 & 4 & 3 \\ 0 & 4 & 2 \\ 0 & -2 & 0 \end{pmatrix}.$$

Jetzt betrachten wir die Matrix unten rechts, d.h. wir setzen

$$A_1 = \begin{pmatrix} 4 & 2 \\ -2 & 0 \end{pmatrix}.$$

Wir wissen, dass diese Matrix wieder 2 als doppelten Eigenwert hat, d.h., sie hat wiederum einen Eigenvektor

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

Wir ergänzen zu einer Basis Basis $\{v_2, e_2\}$ des $\mathbb{R}^2$ und setzen

$$S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Es ergibt sich

$$S = S_1S_2 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$$

und wir bekommen mit

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

die gesuchte Trigonalisierung von A. △

6 Potenzen von linearen Abbildungen

Für eine lineare Selbstabbildung $f \in \mathrm{Hom}(V,V)$ können wir Potenzen definieren: Man setzt $f^0 = \mathrm{id}$ und rekursiv

$$f^{n+1} = f \circ f^n, \quad n = 0,1,2,\dots$$

(also wie gewohnt $f^1 = f$, $f^2 = f \circ f$ usw.).

Das erlaubt es uns auch, lineare Selbstabbildungen eines K-Vektorraums in Polynome in $K[t]$ einzusetzen, d.h. für ein Polynom $p = a_0 + a_1 t + \dots + a_n t^n$ ist

$$p(f) = a_0 \,\mathrm{id} + a_1 f + \dots + a_n f^n \in \mathrm{Hom}(V,V).$$

Liegt das Polynom faktorisiert vor, also z.B. $p = (t-\lambda_1)(t-\lambda_2)$, so ist

$$p(f) = (f - \lambda_1 \,\mathrm{id}) \circ (f - \lambda_2 \,\mathrm{id}).$$

Für quadratische Matrizen A gilt im ersten Fall

$$p(A) = a_0 E + a_1 A + \dots + a_n A^n.$$

und im zweiten Fall

$$p(A) = (A - \lambda_1 E) \cdot (A - \lambda_2 E).$$

Beispiel.

Wir betrachten die Matrix

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

und das Polynom

$$p = (t-1)(t-3) = t^2 - 4t + 3.$$

Wir können nun rechnen

$$\begin{aligned} p(A) &= A^2 - 4A + 3E_3 \\ &= \begin{pmatrix} -7 & -8 & -8 \\ 8 & 9 & 8 \\ 8 & 8 & 9 \end{pmatrix} - 4\begin{pmatrix} -1 & -2 & -2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \\ &= \begin{pmatrix} -7+4+3 & -8+8+0 & -8+8+0 \\ 8-8+0 & 9-12+3 & 8-8+0 \\ 8-8+0 & 8-8+0 & 9-12+3 \end{pmatrix} = 0_{3\times 3}. \end{aligned}$$

Wir können zum Beispiel folgern, dass

$$A^2 = 4A - 3E_3,$$

d.h. A^2 ist eine Linearkombination aus $A^0 = E_3$ und $A^1 = A$. Dies gibt eine lineare Rekursion für die Potenzen von A, denn es gilt für alle n

$$A^n = 4A^{n-1} - 3A^{n-2}, \quad A^0 = E_3,\ A^1 = A.$$

Damit bekommen wir z.B.

$$\begin{aligned} A^3 &= A \cdot A^2 = A(4A - 3E_3) = 4A^2 - 3A \\ &= 4(4A - 3E_3) - 3A = 16A - 12E_3 - 3A = 13A - 12E_3 \end{aligned}$$

und wir bekommen ebenso alle weiteren Potenzen von A als Linearkombination von A und E_3. $\triangle$

Beispiel.
Wir betrachten die Matrix

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

Wie wir schon wissen, ist

$$\chi_A = t^2 + 1.$$

Wir setzen A in χ_A ein und bekommen

$$\chi_A(A) = A^2 + E_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

$\triangle$

In diesem Beispiel liefert A, eingesetzt in sein eigenes charakteristisches Polynom die Nullmatrix.

Potenzen von Matrizen haben viel mit den zugehörigen Eigenwerten und Eigenvektoren zu tun. Wir formulieren die bestimmende Gleichung für einen Eigenraum von f zum Eigenwert λ etwas komplizerter: Für das Polynom $L = t - \lambda$ gilt

$$\mathrm{Eig}(f, \lambda) = \mathrm{Kern}(f - \lambda\,\mathrm{id}) = \mathrm{Kern}(L(f)).$$

Wofür ist diese komplizierte Formulierung nützlich? Betrachten wir ein allgemeines Polynom

$$p = a_r t^r + \cdots + a_1 t + a_0$$

mit $\deg(p) = r$ und setzen hier f ein:

$$p(f) = a_r f^r + \cdots + a_1 f + a_0\,\mathrm{id}_V \in \mathrm{Hom}(V, V).$$

Ist ein Vektor $v \neq 0$ im Kern von $p(f)$, dann gilt

$$a_r f^r(v) + \cdots + a_1 f(v) + a_0 v = 0,$$

d.h. die Vektoren $v, f(v), f^2(v), \ldots, f^r(v)$ sind linear abhängig. Wir definieren den Raum

$$W = \langle \{v, f(v), \ldots, f^{r-1}(v)\} \rangle.$$

Ein $w \in W$ hat die Form $w = k_0 v + k_1 f(v) + \cdots + k_{r-1} f^{r-1}(v) \in W$ und es gilt

$$f(w) = k_0 f(v) + k_1 f^2(v) + \cdots + k_{r-1} f^r(v).$$

Da aber $f^r(v)$ ein Linearkombination der Vektoren $v, f(v), \dots, f^{r-1}(v)$ ist, folgt $f(w) \in W$. Wir haben also gezeigt:

Lemma 6.1. *Es sei* $f \in \mathrm{Hom}(V, V)$ *und* $p = a_r t^r + \dots + a_1 t + a_0$ *ein Polynom mit* $\deg(p) = r$. *Dann ist*

$$W = \langle \{v, f(v), \dots, f^{r-1}(v)\} \rangle$$

für jedes $v \in \mathrm{Kern}(p(f))$ *ein* f*-invarianter Unterraum von* V *mit* $0 \le \dim(W) \le r$ *(und* $\dim(W) \ge 1$ *falls* $v \neq 0$*).*

Beispiel.

Wir betrachten

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

und das Polynom $p = (t-2)^2$ mit $\deg(p) = 2$.

Es ist

$$p(A) = (A - 2E_3)^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

also $\mathrm{Kern}(p(A)) = \langle \{e_1, e_2\} \rangle$.

Zum Beispiel gilt für $v = e_1$

$$Ae_1 = 2e_1.$$

also ist $W = \langle \{e_1, Ae_1\} \rangle = \langle e_1 \rangle$ ein A-invarianter Unterraum. Für $v = e_2$ bekommen wir

$$Ae_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix},$$

also ist $W = \langle \{e_2, Ae_2\} \rangle = \langle \{e_1, e_2\} \rangle$ ebenfalls ein A-invarianter Unterraum. △

Aus dem Lemma folgt direkt:

Korollar 6.2. *Ist* $f \in \mathrm{Hom}(V, V)$ *und* $p = a_r t^r + \dots + a_1 t + a_0$ *ein Polynom mit* $\deg(p) = r$*, so dass* $p(f) = 0$. *Dann ist der Raum* $\langle \{v, f(v), \dots, f^{r-1}(v)\} \rangle$ *für jedes* $v \in V$ *ein* f*-invarinter Unterraum von* V *ist.*

Da $p(f) = 0$ gilt, ist jedes v in $\mathrm{Kern}(p(f))$.

Beispiel.

Ist $f \in \mathrm{Hom}(V, V)$ diagonalisierbar mit den Eigenwerten $\lambda_1, \dots, \lambda_k$, welche wir jetzt als paarweise verschieden annehmen. Wir betrachten das Polynom

$$Q = (t - \lambda_1) \cdots (t - \lambda_k).$$

Da f diagonalisierbar ist, ist

$$V = \mathrm{Eig}(f, \lambda_1) \oplus \cdots \oplus \mathrm{Eig}(f, \lambda_k),$$

d.h. für jedes $v \in V$ gibt es $v_i \in \mathrm{Eig}(f, \lambda_i)$, so dass $v = v_1 + \cdots + v_k$. Es gilt

$$Q(f)(v) = Q(f)(v_1) + \cdots + Q(f)(v_k).$$

Das Polynom Q schreiben wir als $Q = a_0 + a_1 t + \cdots + a_k t^k$ und so sieht man

$$\begin{aligned} Q(f)(v_i) &= a_0 v_i + a_1 f(v_i) + a_2 f^2(v_i) + \cdots + a_k f^k(v_i) \\ &= a_0 v_i + a_1 \lambda_i v_i + a_2 \lambda_i^2 v_i + \cdots + a_k \lambda^k v_i \\ &= Q(\lambda_i) v_i. \end{aligned}$$

Daher folgt wegen $Q(\lambda_i) = 0$ auch

$$Q(f)(v) = 0$$

für alle v, d.h. $Q(f) = 0$ (mit anderen Worten: $Q(f)$ ist die Nullabbildung).

Ist hingegen P ein echter Teiler von Q, so ist ein λ_i keine Nullstelle von P und mit obiger Argumentation bekommt man $P(f)(v_i) \neq 0$, daher ist $P(f) \neq 0$.

Für das charakteristische Polynom $\chi_f = (t - \lambda_1)^{r_1} \cdots (t - \lambda_k)^{r_k}$ gilt ebenso $\chi_f(f) = 0$. △

Wir rechnen das gleiche Beispiel noch einmal mit Matrizen:

Beispiel.

Es sei $A \in K^{n\times n}$ diagonalisierbar, d.h. es existieren eine invertierbare Matrix $S \in K^{n\times n}$ und eine Diagonalmatrix $D \in K^{n\times n}$ mit $A = SDS^{-1}$ (und auf der Diagonalen von D stehen die Eigewerte $\lambda_1, \ldots, \lambda_k$, ggf. mehrfach).

Wir setzen wir A in das Polynom $Q = (t - \lambda_1) \cdot (t - \lambda_k) = a_0 + a_1 t + \cdots + a_k t^k$ ein und benutzen $A^r = SD^r S^{-1}$:

$$\begin{aligned} Q(A) &= a_0 E_n + a_1 A + \cdots + a_k A^k \\ &= a_0 E_n + a_1 SDS^{-1} + \cdots + a_k SD^k S^{-1} \\ &= S(a_0 E_n + a_1 D + \cdots + a_k D^k) S^{-1} \\ &= SQ(D) S^{-1}. \end{aligned}$$

Da D eine Diagonalmatrix ist, sind die Diagonaleinträge von D^r genau die r-ten Potenzen der Diagonaleinträge von D und da jedes Diagonalelement von D eine Nullstelle von Q ist, folgt $Q(D) = 0_{n\times n}$ und damit

$$Q(A) = 0_{n\times n}.$$

△

Dieses Beispiel zeigt, dass es für eine diagonalisierbare Abbildung f stets ein Polynom p mit $\deg(p) \leq n$ gibt (nämlich das charakteristische Polynom), so dass $p(f) = 0$ ist (also die Nullabbildung). Das Beispiel zeigt sogar mehr: Im Fall von diagonalisierbaren Abbildung gibt es ein Polynom q dessen Grad der Anzahl der verschiedenen Eigenwerte entspricht und welches immer noch $q(f) = 0$ erfüllt.

Wie sieht die Situation für nicht-diagonalisierbare f aus? Wir berechnen den allgemeinen Fall in Vektorräumen der Dimension 2 per Hand aus:

Beispiel.
Wir betrachten

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K^{2\times 2}.$$

Das charakteristische Polynom ist

$$\begin{aligned}\chi_A &= (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + ad - bc \\ &= \lambda^2 - \mathrm{Spur}(A)\lambda + \det(A).\end{aligned}$$

Wir setzen nun A in χ_A ein; dazu bemerken wir $A^2 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} a^2+bc & b(a+d) \\ c(a+d) & bc+d^2 \end{smallmatrix}\right)$ und bekommen

$$\begin{aligned}&\chi_A(A) \\ &= A^2 - (a+d)A + (ad-bc)E_2 \\ &= \begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & bc+d^2 \end{pmatrix} - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\ &= \begin{pmatrix} a^2+bc-(a+d)a+ad-bc & b(a+d)-(a+d)b \\ c(a+d)-(a+d)c & bc+d^2-(a+d)d+ad-bc \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

△

Das Beispiel zeigt, dass es für jede Matrix $A \in K^{2\times 2}$ (egal ob diagonalisierbar oder nicht) immer ein immer ein Polynom p vom Grad 2 gibt, so dass $p(A) = 0$ gilt; es geht immer $p = \chi_A$. Würde man die notwendige Ausdauer haben, so könnte man analog zu obigem Beispiel zeigen, dass es für jede 3×3 Matrix ein Polynom p von Grad 3 gibt, so dass $p(A) = 0$ ist (und auch hier würde wieder $p = \chi_A$ gehen). Ein allgemeines Ergebnis hierzu ist das Thema des nächsten Kapitels.

7 Der Satz von Cayley-Hamilton

In diesem Abschnitt beweisen wir den Satz von Cayley-Hamilton, welcher besagt, dass für *jede* lineare Abbildung f gilt, dass $\chi_f(f) = 0$ ist. In Matrix-Schreibweise: Zu *jeder* $n \times n$-Matrix gibt es ein Polynom vom Grad n, nämlich das charakteristische Polynom χ_A, so dass $\chi_A(A) = 0_{n \times n}$.

Bisher wissen wir dieses Ergebnis nur für diagonalisierbare Abbildungen und der Ansatz lässt sich nicht ohne weiteres auf anderen Fälle übertragen.

Ist n die Dimension von V, so ist die Dimension von $\mathrm{Hom}(V, V)$ gleich n^2. Daher sind die Bilder von $\mathrm{id}, f, f^2, \ldots, f^{n^2}$ unter Φ_f linear abhängig in $\mathrm{Hom}(V, V)$. Es gibt also eine nicht-triviale Linearkombination dieser Bilder, welche Null ergibt. Mit anderen Worten: Zu jedem $f \in \mathrm{Hom}(V, V)$ gibt es ein Polynom p_f vom Grad höchstens n^2, welches nicht das Nullpolynom ist, so dass $p_f(f) = 0$ ist.

$\mathrm{Hom}(V, V)$ ist isomoporh zu $K^{n \times n}$.

Für nicht-diagonalisierbare f bekommen wir also so nur die obere Schranke n^2 an den Polynomgrad für p, so dass $p(f) = 0$ ist.

Es geht, mit mehr Aufwand, aber besser. Es gilt nämlich immer die Schranke n und als Polynom geht immer das charakteristische Polynom der Abbildung, wie der folgende Satz zeigt. Wir formulieren ihn hier vorerst nur für den Fall $K = \mathbb{C}$.

Satz 7.1 (Cayley-Hamilton). *Es sei V ein n-dimensionaler $\mathbb{C}$-Vektorraum, $f \in \mathrm{Hom}(V, V)$. Dann gilt für das charakteristische Polynom χ_f von f immer*

$$\chi_f(f) = 0 \in \mathrm{Hom}(V, V).$$

Ebenso gilt für jede Matrix $A \in K^{n \times n}$, immer, dass $\chi_A(A) = 0 \in K^{n \times n}$.
Achtung, dies hier ist **kein** Beweis: $\chi_A(A) = \det(A - AE_n) = \det(A - A) = 0$.
Wieso nicht? $\chi_A(A)$ ist eine Matrix. $\det(A - \lambda E_n)$ ist aber immer eine Zahl, selbst, wenn man formal $\lambda = A$ einsetzt. Schon die erste Gleichung ist also falsch.

Beweis.
Da jedes komplexe Polynom in Linearfaktoren zerfällt, ist f trigonalisierbar, d.h. es gibt eine f-invariante Fahne

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V.$$

Es sei $B = \{v_1, \ldots, v_n\}$ eine Basis von V, so dass $V_i = \langle \{v_1, \ldots, v_i\} \rangle$, d.h. es gilt

$$M_B^B(f) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Dann sind $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ die Eigenwerte von f und das charakteristische Polynom ist

$$\chi_f = (\lambda_1 - t) \cdots (\lambda_n - t).$$

Wir definieren die Polynome $p_k = (\lambda_1 - t) \cdots (\lambda_k - t)$ (d.h. $p_n = \chi_f$) und die zugehörigen linearen Abbildungen $G_k = p_k(f) = (\lambda_1 \mathrm{id} - f) \circ \cdots \circ (\lambda_k \mathrm{id} - f)$ (d.h. es gilt $G_n = \chi_f(f)$).

Jetzt zeigen wir induktiv, dass G_k auf dem Raum V_k die Nullabbildung ist: Ist $k = 1$, so ist $G_1 = (\lambda_1 \mathrm{id} - f)$. Da V_1 von einem Eigenvektor v_1 zum Eigenwert λ_1 aufgespannt wird, gilt für jedes $v \in V_1$, dass $G_1(v) = \lambda_1 v - f(v) = \lambda_1 v - \lambda_1 v = 0$.

Sei nun $k \geq 2$ und sei $v \in V_k$. Dann können wir v zerlegen als $v = w + \mu v_k$ mit $w \in V_{k-1}$. Da V_{k-1} f-invariant ist, gilt $f(w) \in V_{k-1}$ und also auch

$$\lambda_k w - f(w) \in V_{k-1}.$$

Aus der Matrix-Darstellung $M_B^B(f)$ sehen wir, dass $f(v_k) = \alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1} + \lambda_k v_k$, es folgt also, dass

$$\lambda_k v_k - f(v_k) = \alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1} \in V_{k-1}.$$

Wir benutzen die Induktionsvorraussetzung $G_{k-1}|_{V_{k-1}} = 0$ und bekommen

$$\begin{aligned}
G_k(v) &= (G_{k-1} \circ (\lambda_k \operatorname{id} - f))(v) \\
&= (G_{k-1} \circ (\lambda_k \operatorname{id} - f))(w + \mu v_k) \\
&= G_{k-1}((\lambda_k \operatorname{id} - f)(w) + \mu(\lambda_k \operatorname{id} - f)(v_k)) \\
&= G_{k-1}(\underbrace{\lambda_k w - f(w)}_{\in V_{k-1}}) + G_{k-1}(\underbrace{\lambda_k v_k - f(v_k)}_{\in V_{k-1}}) = 0.
\end{aligned}$$

□

Dieser Beweis des Satzes von Cayley-Hamilton funkioniert auch für allgemeine Körper, sobald die Abbildung f (bzw. die Matrix A) trigonalisierbar ist. Über anderen Körpern als $\mathbb{C}$ gilt das nicht für jede Matrix. Wir werden später in der Vorlesung noch zeigen, dass der Satz trotzdem in allen Körpern gilt.

Der Satz von Cayley-Hamilton hat einige Folgerungen, die ggf. überraschend sind, z.B.:

Lemma 7.2. *Ist $A \in K^{n \times n}$ invertierbar, so gibt es ein Polynom p vom Grad $n-1$, so dass*

$$A^{-1} = p(A),$$

mit anderen Worten: A^{-1} ist eine Linearkombination von $E, A, A^2, \ldots, A^{n-1}$.

Der *Beweis* ist eine Übungsaufgabe.

Mit Hilfe des Satzes von Cayley-Hamilton können wir noch zeigen, dass sich reelle Matrizen immer „fast trigonalisieren“ lassen. Dazu ein kurzes Lemma:

Lemma 7.3. *Ist V ein $\mathbb{R}$-Vektorraum, mit $\dim(V) \geq 1$, so hat V einen f-invarianten Unterraum W mit $1 \leq \dim(W) \leq 2$.*

Dann ist $f|_W \in \operatorname{Hom}(W, W)$ und ist $\dim(W) = 2$, so gibt es eine Basis B von W so dass

$$M_B^B(f|_W) = \begin{pmatrix} 0 & -c \\ 1 & -b \end{pmatrix}.$$

Beweis.
Nach dem Fundamentalsatz für reelle Polynome (Satz 3.10) gilt für das charakteristische Polynom

$$\chi_f = \pm(t-\lambda_1)\cdot(t-\lambda_r)q_1(t)\cdots q_m(t)$$

mit quadratischen Polynomen q_k, welche keinen reellen Nullstellen haben.

Ist $r \geq 1$, so gibt es einen Eigenvektor v_1 zum Eigenwert λ_1 und $\langle\{v_1\}\rangle$ ist ein eindimensionaler f-invarianter Unterraum.

Ist $r = 0$, so wählen wir $v \neq 0$ beliebig und wissen nach Cayley-Hamilton $\chi_f(f) = 0$, d.h. es gilt

$$0 = \chi_f(f)(v) = (q_1(f)\circ\cdots\circ q_m(f))(v).$$

Es gibt also ein $i \in \{1,\dots,m\}$, so dass

$$w = (q_{i+1}(f)\circ\cdots\circ q_m(f))(v) \neq 0, \quad \text{und} \quad q_i(f)(w) = 0$$

(wobei wir $w = v$ nehmen, wenn $i = m$ ist, also wenn $q_m(f)(v) = 0$ ist). Jetzt zeigen wir, dass der Raum $W = \langle\{w, f(w)\}\rangle$ dann f-invariant ist: Wir bezeichnen $q_i = t^2 + b_i t + c_i$ und folgern aus $q_i(f)(w) = 0$, dass

$$f(f(w)) + b_i f(w) + c_i w = 0 \qquad (*)$$

gilt. Hieraus folgt $f(W) \subset W$, denn: Ist $v \in W$, so können wir schreiben $v = \mu_1 w + \mu_2 f(w)$. Also gilt

$$f(v) = \mu_1 f(w) + \mu_2 f(f(w)) = \mu_1 f(w) + \mu_2(-b_i f(w) - c_i w) \in W.$$

Da $f|_W$ keinen Eigenwert hat, ist $B = \{e_1, e_2\}$ mit $e_1 = w$ und $e_2 = f(w)$ eine Basis von W. Es gilt $f(e_1) = f(w) = e_2$ und aus (*) folgt $f(e_2) = f(f(w)) = -b_i f(w) - c_i w = -b_i e_2 - c_i e_1$, d.h. es gilt

$$M_B^B(f|_W) = \begin{pmatrix} 0 & -c_i \\ 1 & -b_i \end{pmatrix}$$

□

Satz 7.4. *Ist V ein $\mathbb{R}$-Vektorraum und $f \in \operatorname{Hom}(V,V)$, so gibt es eine Basis B von V, sodass*

$$M_B^B(f) = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & * & \\ & & \lambda_r & & & \\ & & & B_1 & & \\ & 0 & & & \ddots & \\ & & & & & B_m \end{pmatrix}$$

mit $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ und $B_i = \left(\begin{smallmatrix} 0 & -c_i \\ 1 & -b_i \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$.

Der Beweis geht i.W. analog zum Beweis des Trigonalisierungssatzes (Satz 4.7) und wir lassen die Details hier aus.

Beispiel.

Wir betrachten die reelle Matrix

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

Das charakteristische Polynom ist

$$\chi_A(\lambda) = \det\left(\begin{pmatrix} -\lambda & -1 & 0 \\ 0 & -\lambda & -1 \\ 1 & 0 & -\lambda \end{pmatrix}\right) = -\lambda^3 + 1.$$

Ein Nullstelle (und damit einen Eigenwert) sehen wir schnell, nämlich $\lambda_1 = 1$. Polynomdivision liefert

$$\chi_A(\lambda) = (\lambda - 1)(-\lambda^2 - \lambda - 1).$$

Das Polynom $-\lambda^2 - \lambda - 1$ hat keine reelle Nullstelle (die beiden komplexen Nullstellen sind $\frac{-1\pm i\sqrt{3}}{2}$). Damit ist die Matrix nicht trigonalisierbar, aber wir können die „reelle“ fast obere Dreicksform aus dem vorigen Satz bilden. Dazu berechnen wir einen Eigenvektor zum Eigenwert 1, z.B.

$$v_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Wir ergänzen v_1 zu eines Basis und schreiben diese in

$$S_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Es ergibt sich

$$S_1^{-1}AS_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Der Block $A_2 = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ hat keine reellen Eigenwerte mehr, daher ist $\{w, A_2 w\}$ für jedes $w \neq 0$ eine Basis des $\mathbb{R}^2$. Wir wählen eine solche, nämlich $\{e_1, A_2 e_1\}$ und bilden

$$S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}, \quad S = S_1 S_2 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Wie erwartet, bekommen wir

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

△

8 Ideale und das Minimalpolynom

Definition 8.1. Eine Teilmenge $\mathcal{I}$ eines kommutativen Ringes R heißt *Ideal*, wenn gilt

1. $p, q \in \mathcal{I} \implies p - q \in \mathcal{I}$
2. $p \in \mathcal{I},\ q \in R \implies q \cdot p \in \mathcal{I}$.

In einem nicht-kommutativen Ring definiert dies ein *Linksideal*. Für ein Rechtsideal würde als zweiter Punkt $p \cdot q \in \mathcal{I}$ gefordert und man sagt *Ideal*, wenn beides gilt.

Beispiel.

1. Wir betrachte den Ring $\mathbb{Z}$. Die Mengen $\mathbb{Z}_m = \{nm \ : \ n \in \mathbb{Z}\}$ sind für jedes $m \in \mathbb{Z}$ Ideale in $\mathbb{Z}$ und man kann einfach zeigen, dass alle Ideale von $\mathbb{Z}$ diese Form haben.

2. Wir betrachten den Polynom-Ring $K[t]$. Für eine endliche Teilmenge $M = \{\lambda_1, \dots, \lambda_r\} \subset K$ definieren wir

$$\mathcal{I}_M = \{p \in K[t] \ : \ \forall k = 1, \dots, r : p(\lambda_k) = 0\},$$

also die Menge aller Polynome, die alle Elemente von M als Nullstellen haben.

Dies $\mathcal{I}_M$ ist ein Ideal: Sind $p, q \in \mathcal{I}_M$, dann gilt für alle λ_k, dass $p(\lambda_k) - q(\lambda_k) = 0 - 0 = 0$ und für $p \in \mathcal{I}_M$ und $q \in K[t]$ gilt $p(\lambda_k)q(\lambda_k) = 0 \cdot q(\lambda_k) = 0$.

3. Wir betrachten einen Vektorraum V, ein $f \in \mathrm{Hom}(V, V)$ und dazu die Menge

$$\mathcal{I}_f = \{p \in K[t] \ : \ p(f) = 0 \in \mathrm{Hom}(V, V)\} \subset K[t].$$

Auch diese Menge ist ein Ideal in $K[t]$, was man genau wie im ersten Beispiel einsieht.

$\triangle$

Die Ideale in Polynomringen haben eine spezielle Struktur:

Satz 8.2. *Es sei $\mathcal{I} \subset K[t]$ ein Ideal. Ist $\mathcal{I} \neq \{0\}$, dann gibt es genau ein Polynom $m \in K[t]$ mit den Eigenschaften:*

i) m ist normiert.

ii) Für jedes $p \in \mathcal{I}$ gibt es ein $q \in K[t]$, so dass $p = qm$.

Erinnere: Ein Polynom heißt normiert, wenn der Koeffizient des Terms mit der höchsten Potenz 1 ist.

Im allgemeinen Fall heißen solche Ideale *Hauptideale*. Die Mengen $\mathbb{Z}_m$ in $\mathbb{Z}$ sind ebenfalls Hauptideale.

Beweis.
Es sei

$$d := \min\{r \in \mathbb{N} \ : \ \text{es gibt ein } p \in \mathcal{I} \text{ mit } p \neq 0 \text{ und } r = \deg(p)\}.$$

Dann gibt es ein normiertes $m \in \mathcal{I}$ mit $\deg(m) = d$. Sei nun $p \in \mathcal{I}$. Dann ist der Grad von p größer oder gleich dem Grad von m und wir dividieren p mit Rest durch m und bekommen

$$p = q \cdot m + r,$$

wobei entweder $r = 0$ oder $0 \leq \deg(r) < d$ gilt. Wäre $r \neq 0$, so wäre auch $r = p - q \cdot m \in \mathcal{I}$ und gleichzeitig $\deg(r) < \deg(m) = d$, was ein Widerspruch zur Minimalität von d ist.

Zur Eindeutigkeit: Sei $\tilde{m}$ ein weiteres solches Polynom. Dann ergibt Division von m durch $\tilde{m}$

$$\tilde{m} = q \cdot m$$

(und der Rest ist Null, wie schon eben). Da m und $\tilde{m}$ beide normiert sind und gleichen Grad haben, ist $q = 1$ und $\tilde{m} = m$. □

Definition 8.3. Das Polynom m aus Satz 8.2 heißt *Minimalpolynom von $\mathcal{I}$*. Im Fall

$$\mathcal{I}_f = \{p \in K[t] \ : \ p(f) = 0 \in \operatorname{Hom}(V, V)\} \subset K[t]$$

heißt das Polynom m *Minimalpolynom von f*.

Mit anderen Worten:

> Das Minimalpolynom einer linearen Abbildung f (bzw. einer Matrix A) ist das normierte Polynom m_f (bzw. m_A) mit kleinsten Grad, so dass $m_f(f) = 0$ (bzw. $m_A(A) = 0$).

Der Satz von Cayley-Hamilton sagt, dass das Minimalpolynom ein Teiler des charakteristischen Polynoms ist. Im Allgemeinen unterscheiden sich Minimalpolynom und charakteristisches Polynom, und zwar nicht nur im führenden Koeffizienten, sondern ggf. auch im Grad:

Beispiel.

Es sei

$$A = \begin{pmatrix} 0 & 1 & & & & & \\ & \ddots & \ddots & & & & \\ & & \ddots & 1 & & & \\ & & & \ddots & 0 & & \\ & & & & \ddots & \ddots & \\ & & & & & \ddots & 0 \\ & & & & & & 0 \end{pmatrix} \in K^{n \times n},$$

wobei $d - 1$ Einsen auf der ersten oberen Nebendiagonalen stehen sollen. Da $A - \lambda E_n$ eine obere Dreiecksmatix ist, ist das charakteristische Polynom

$$\chi_A = (-1)^n t^n.$$

Dann bemerkt man,

$$\begin{aligned} Ae_1 &= 0 \\ Ae_2 &= e_1 \\ &\vdots \\ Ae_d &= e_{d-1} \\ Ae_i &= 0, \quad d < i \le n. \end{aligned}$$

Wir schließen, dass für die ersten l Potenzen, genauer für $1 \le l < d$, gilt

$$A^l e_d = e_{d-l} \quad \text{und} \quad A^l e_d = 0 \text{ falls } l \ge d.$$

Es folgt $A^d = 0$ und $A^{d-1} \neq 0$ (da $A^{d-1}e_d = e_1$). Also ist das Minimalpolynom $m_A = t^d$. △

Satz 8.4. *Es sei K ein Unterkörper von $\mathbb{C}$, V ein n-dimensionaler K-Vektorraum, $f \in \operatorname{Hom}(V, V)$. Dann gilt*

i) m_f teilt χ_f.

ii) χ_f teilt m_f^n.

Beweis.
Im Fall $f = 0$ gilt $\chi_f = m_f = 0$ und wir betrachten den Fall $f \neq 0$.

i) hatten wir schon gesehen (folgt direkt aus Cayley-Hamilton).

Für ii) argumentieren wir für $n \times n$ Matrizen. Die Polynome m_A und χ_A zerfallen über $\mathbb{C}$ in Linearfaktoren, insbesondere ist

$$\chi_A = (-1)^n (t - \lambda_1)^{r_1} \cdots (t - \lambda_k)^{r_k}$$

wobei wir $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ paarweise verschieden annehmen. Wir wissen aus i), dass das Minimalpolynom die Form

$$m_A = (t - \lambda_1)^{s_1} \cdots (t - \lambda_k)^{s_k}$$

mit $0 \le s_i \le r_i$ hat. Wir zeigen nun, dass tatsächlich $s_i \ge 1$ für $i = 1, \ldots, k$ gilt. Angenommen es gelte $s_i = 0$. Es sei $v_i \in \mathbb{C}^n$ ein Eigenvektor zum Eigenwert λ_i. Wir wenden $m_A(A)$ darauf an. Wegen $m_A(A)v_i = m_A(\lambda_i)v_i$ (vgl. Beispiel nach Korollar 6.2) folgt $m_A(A)(v_i) \neq 0$, was im Widerspruch zu $m_A(A) = 0$ steht. Jetzt definieren wir das Polynom $q = (t - \lambda_1)^{ns_1 - r_1} \cdots (t - \lambda_k)^{ns_k - r_k}$. Dies ist a priori ein komplexes Polynom. Da aber

$$m_A^n = (-1)^n q \cdot \chi_A$$

gilt, und da das Minimalpolynom und das charakteristische Polynom per Definition Koeffizienten in K haben, muss auch $q \in K[t]$ liegen. □

Dass das Minimalpolynom ein Teiler des charakteristischen Polynoms ist, bedeutet folgendes: Es sei $f \in \operatorname{Hom}(V, V)$ mit einem n-dimensionalen Vektorraumraum V und wir nehmen an, dass das charakteristische Polynom in Linearfaktoren zerfällt. Dann lässt sich das charakteristische Polynom schreiben als

$$\chi_f(t) = (-1)^n (t-\lambda_1)^{a(f,\lambda_1)} \cdots (t-\lambda_r)^{a(f,\lambda_r)},$$

wobei $a(f,\lambda)$ die algebraische Vielfachheit des Eigenwertes λ angibt und die Werte $\lambda_1, \ldots, \lambda_k$ alle verschieden sind. Das Minimalpolynom muss dann von der Form

$$m_f(t) = (t-\lambda_1)^{a_1} \cdots (t-\lambda_r)^{a_r}$$

mit $1 \leq a_k \leq a(f,\lambda_k)$.

Die obere Schranke ergibt sich, da m_f ein Teiler von χ_f ist, und die untere Schranke kommt daher, dass χ_f ein Teiler von m_f^n ist.

Beispiel.

Wir betrachten

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

Als charakteristisches Polynom bekommt man

$$\chi_A = (\lambda-2)^2(\lambda-1)^2.$$

Die möglichen Minimalpolynome sind neben χ_A selbst noch

$$(\lambda-2)(\lambda-1), \quad (\lambda-2)(\lambda-1)^2 \quad \text{und} \quad (\lambda-2)^2(\lambda-1).$$

Wir berechnen

$$(A-2E_4)(A-E_4) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$(A-2E_4)(A-E_4)^2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

und

$$(A-2E_4)^2(A-E_4) = 0_{4\times 4}$$

und schließen, dass das Minimalpolynom $m_A = (\lambda-2)^2(\lambda-1)$ ist. △

Satz 8.5. *Für ein $f \in \mathrm{Hom}(V,V)$ auf einem endlichdimensionalen K-Vektorraum gilt: Ist f diagonalisierbar, so zerfällt m_f in* ***paarweise verschiedene Linearfaktoren****, d.h. es gilt*

$$m_f(t) = (t-\lambda_1)\cdots(t-\lambda_r),$$

wobei die λ_k paarweise verschieden sind.

Beweis.
Es sei f diagonalisierbar und $\lambda_1, \dots, \lambda_r$ die paarweise verschiedenen Eigenwerte. Wir setzen $p = (t - \lambda_1) \cdots (t - \lambda_r)$ und zeigen, dass $p = m_f$ gilt.

Auf jeden Fall ist p ein Teiler von m_f, da alle λ_k auch Nullstellen von m_f sind. Da f diagonalisierbar ist, können wir V zerlegen als

$$V = \operatorname{Eig}(f, \lambda_1) \oplus \cdots \oplus \operatorname{Eig}(f, \lambda_r).$$

Ein $v \in V$ können wir also schreiben als $v = v_1 + \cdots + v_r$ mit $v_k \in \operatorname{Eig}(f, \lambda_k)$. Dann gilt $p(f)(v) = p(f)(v_1) + \cdots + p(f)(v_k) = 0$. Also gilt $p(f) = 0$, und da das Minimalpolynom das Polynom mit dieser Eigenschaft mit kleinsten Grad ist, folgt $p = m_f$. $\square$

Dass auch die Rückrichtung gilt, werden wir später zeigen.

9 Haupträume und Hauptvektoren

Definition 9.1. Sei V ein endlichdimensionaler K-Vektorraum und $f \in \operatorname{Hom}(V,V)$, dessen Minimalpolynom m_f in Linearfaktoren zerfällt. Es sei λ ein Eigenwert von f und $a = \operatorname{ord}(m_f, \lambda)$ die Ordnung der Nullstelle. Dann heißt der Raum

$$\operatorname{Haupt}(f,\lambda) = \operatorname{Kern}((f - \lambda \operatorname{id})^a)$$

der *Hauptraum* von f zum Eigenwert λ.

Ein Vektor $v \in \operatorname{Haupt}(f,\lambda)$ ist ein *Hauptvektor* von f zum Eigenwert λ der Ordnung k, wenn

$$(f - \lambda \operatorname{id})^k v = 0 \quad \text{und} \quad (f - \lambda \operatorname{id})^{k-1} v \neq 0.$$

Insbesondere sind Hauptvektoren nicht 0.

Wir bemerken sofort

$$\operatorname{Eig}(f,\lambda) \subset \operatorname{Haupt}(f,\lambda).$$

Untersuchen wir die Struktur von Haupträumen genauer und beginnen mit einem einfachen Beispiel:

Beispiel.

Wir betrachten die reelle Matrix

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

Das charakteristische Polynom ist $\chi_A = (t-2)^3(t-3)$ und die Eigenwerte sind 2 und 3. Der Eigenraum zum Eigenwert 2 ist

$$\operatorname{Eig}(A,2) = \operatorname{Kern}(A - 2E_4) = \operatorname{Kern}\left(\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right) = \langle \{e_1, e_3\} \rangle.$$

Was ist der entsprechende Hauptraum? Bestimmen wir zuerst das Minimalpolynom: Das Minimalpolynom von A muss mindestens die Terme $(t-3)$ und $(t-2)$ enthalten. Wir berechnen

$$\begin{aligned}(A - 2E_4)(A - 3E_4) &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \neq 0.\end{aligned}$$

Aber da $(A - 2E_4)^2(A - 3E_4) = 0$ gilt, folgt $m_A = (t-2)^2(t-3)$.

Wir lesen ab $\operatorname{ord}(m_A, 2) = 2$ und bekommen den Hauptraum

$$\begin{aligned}\operatorname{Haupt}(A,2) = \operatorname{Kern}((A-2E_4)^2) &= \operatorname{Kern}\left(\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right) \\ &= \langle e_1, e_2, e_3 \rangle.\end{aligned}$$

Wir bemerken, dass $\operatorname{ord}(A,2) = 2 < 3 = \dim(\operatorname{Haupt}(A,2))$. $\triangle$

Zu jedem Eigenvektor $v_1 \in \operatorname{Eig}(f,\lambda)$ gibt es eine *Kette von Hauptvektoren* (auch *Hauptvektorkette*), wie folgt: Man definiert rekursiv $v_2, v_3, \ldots$ durch

$$(f - \lambda\,\mathrm{id})v_k = v_{k-1}, \quad k = 2, \ldots,$$

so lange diese Gleichungssysteme lösbar sind.

Im Allgemeinen sind die v_{k-1} nicht eindeutig durch die v_k bestimmt.

Lemma 9.2. *Für die Vektoren $v_1, \ldots, v_j$ in einer Kette von Hauptvektoren gilt*

$$(f - \lambda\,\mathrm{id})^k v_l = \begin{cases} v_{l-k}, & \text{falls } 0 \le k < l \\ 0, & \text{falls } k \ge l. \end{cases}$$

Außerdem sind die Vektoren $v_1, \ldots, v_j$ linear unabhängig und es gilt $j \le \operatorname{ord}(m_f, \lambda)$.

Beweis.
Die erste Behauptung folgt rekursiv aus der Definition $(f - \lambda\,\mathrm{id})v_k = v_{k-1}$ und $(f - \lambda\,\mathrm{id})v_1 = 0$.

Sei nun $v = \alpha_1 v_1 + \cdots + \alpha_j v_j = 0$ und wir wollen zeigen, dass $\alpha_1 = \cdots = \alpha_j = 0$ gilt.

Wenden wir $(f - \lambda\,\mathrm{id})^{j-1}$ auf v an, so bekommen wir

$$\begin{aligned} 0 &= (f - \lambda\,\mathrm{id})^{j-1}(\alpha_1 v_1 + \cdots + \alpha_j v_j) \\ &= \alpha_1 \underbrace{(f - \lambda\,\mathrm{id})^{j-1}(v_1)}_{=0} + \cdots + \alpha_j \underbrace{(f - \lambda\,\mathrm{id})^{j-1}(v_j)}_{=v_1} \\ &= \alpha_j v_1, \end{aligned}$$

woraus $\alpha_j = 0$ folgt. Dann folgt durch Anwenden von $(f - \lambda\,\mathrm{id})^{j-2}$ auf v, dass $\alpha_{j-1} = 0$ gilt und auf diese Weise schließen wir, dass alle Koeffizienten Null sind.

Die Kette kann höchstens die Länge $a = \operatorname{ord}(m_f, \lambda)$ haben, da v_k ein Hauptvektor der Ordnung k ist, und die höchstmögliche Ordnung von Hauptvektoren per Definition höchstens a ist. $\square$

Wir führen das Beispiel weiter:

Beispiel.
Wir berechnen die Kette von Hauptvektoren von A zum Eigenvektor e_1. Dazu lösen wir $(A - 2E_4)v_2 = e_1$ und bekommen $v_2 = e_2$. Das Gleichungssystem $(A - 2E_4)v_3 = e_2$ ist nicht lösbar, also bricht die Hauptraumkette $v_1 = e_1, v_2 = e_2$ hier ab.

Die Hauptraumkette zum Eigenvektor e_3 bricht sofort ab, denn $(A - 2E_4)v = e_3$ hat keine Lösung.

Beide Hauptraumketten bilden zusammen eine Basis des Hauptraumes: $\mathrm{Haupt}(A, 2) = \langle\{e_1, e_2, e_3\}\rangle$. $\triangle$

Beispiel.
Wir betrachten die Matrix

$$J = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \in K^{n\times n}.$$

Das charakteristische Polynom ist $\chi_J = (\lambda - t)^n$, λ ist der einzige Eigenwert und hat die algebraische Vielfachheit n. Der Eigenraum ist

$$\mathrm{Eig}(J, \lambda) = \langle\{e_1\}\rangle,$$

also ist die geometrische Vielfachheit 1.

Es gilt

$$(J - \lambda E_n) = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}.$$

Man beobachtet $(J - \lambda E_n)^k \neq 0$ für $k < n$ aber $(J - \lambda E_n)^n = 0$. Daher ist das Minimalpolynom m_f gleich dem charakteristischen Polynom. Es gilt also $\mathrm{ord}(J, \lambda) = n$. Der Hauptraum ist also

$$\mathrm{Haupt}(J, \lambda) = \mathrm{Kern}((J - \lambda E_n)^n) = \mathrm{Kern}(0) = K^n.$$

Die zugehörige Kette von Hauptvektoren ergibt sich wie folgt: Wir nehmen als v_1 den Eigenvektor e_1 und bekommen aus

$$(J - \lambda E_e)v_2 = e_1,$$

dass $v_2 = e_2$. Ebenso ergibt sich aus $(J - \lambda E_n)v_3 = e_2$, dass $v_3 = e_3$ und ebenso folgt $v_k = e_k$ für $k = 1, \ldots, n$. $\triangle$

In diesem Beispiel bekamen wir die Standardbasis als Kette von Hauptvektoren für die Matrix J. Dies lag an der speziellen Struktur der Matrix A. Der folgende Satz zeigt in gewissem Sinn eine Umkehrung davon:

Satz 9.3. *Es sei V ein n-dimensionaler K-Vektorraum und $f \in \mathrm{Hom}(V, V)$, so dass $m_f = (t - \lambda)^n = \pm\chi_f$. Weiterhin sei v_1 ein Eigenvektor zum Eigenwert λ und die zugehörige Kette von Hauptvektoren $v_2, \ldots, v_n$ habe die Länge n. Dann gilt für die Basis $B = \{v_1, \ldots, v_n\}$*

Dass die Hauptvektorkette die Länge n hat, ist eine zusätzliche Annahme. Im Allgemeinen können Hauptvektorketten in diesem Fall jede Länge zwischen 1 und n haben. Können Sie zu jedem $1 \leq k \leq n$ ein Beispiel konstruieren?

$$M_B^B(f) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}.$$

Beweis.
Die Kette von Hauptvektoren erfüllt $f(v_1) = \lambda v_1$ und für $k = 2, \ldots, n$ gilt $(f - \lambda\,\mathrm{id})v_k = v_{k-1}$, also

$$f(v_k) = \lambda v_k + v_{k-1}.$$

Wir können also die Basisdarstellung direkt ablesen. □

Satz 9.4. *Es sei* $\{v_1, \ldots, v_k\}$ *eine Basis von* $\mathrm{Eig}(f, \lambda)$ *und weiterhin seien*

$$\begin{aligned} &v_1, v_{1,2}, \ldots, v_{1,j_1} \\ &v_2, v_{2,2}, \ldots, v_{2,j_2} \\ &\quad\vdots \qquad \vdots \\ &v_k, v_{k,2}, \ldots, v_{k,j_k} \end{aligned}$$

die entsprechenden Hauptvektorketten. Dann bilden all diese Vektoren zusammen eine Basis B von $\mathrm{Haupt}(f, \lambda)$. *Eine solche Basis nennen wir* ***Hauptraumkettenbasis*** *von* $\mathrm{Haupt}(f, \lambda)$.

Beweis.
Dass die Vektoren in B linear unabhängig sind, kann man zeigen in dem man alle Vektoren in B zu Null linearkombiniert und dann genügend hohe Potenzen von $(f - \lambda\,\mathrm{id})$ darauf anwendet, um zu zeigen, dass alle Koeffizienten Null sind.

Zeigen wir jetzt, dass die Vektoren in B ein Erzeugendensystem für $\mathrm{Haupt}(f, \lambda)$ bilden: Es sei $w \in \mathrm{Haupt}(f, \lambda)$ und r die Ordnung des Hauptvektors, d.h. es gilt $r \leq \mathrm{ord}(m_f, \lambda)$, $(f - \lambda\,\mathrm{id})^r w = 0$ und $(f - \lambda)^{r-1} w \neq 0$. Dann lässt sich eine Hauptvektorkette „von hinten aus" aufbauen: Wir setzen $w_r = w$, $w_{r-1} = (f - \lambda\,\mathrm{id})w_r$ usw. bis $w_1 = (f - \lambda\,\mathrm{id})^{r-1} w_r$. Wegen $0 = (f - \lambda\,\mathrm{id})^r w_r = (f - \lambda\,\mathrm{id})w_1$ ist w_1 ein Eigenvektor. Da w_1 damit eine Linearkombination von $v_1, \ldots, v_k$ ist, ist w im Spann von B. □

Beispiel.
Betrachten wir die Matrix

$$A = \begin{pmatrix} \lambda & 1 & \cdots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \in K^{5\times 5},$$

welche wieder das charakteristische Polynom $\chi_A = (\lambda - t)^n$ und λ als einzigen Eigenwert mit algebraischer Vielfachheit n hat. Auch hier ist das Minimalpolynom $m_A = \chi_A$ und der Hauptraum ist wieder $\operatorname{Haupt}(J,\lambda) = K^n$. Eine Hauptvektorkette ergibt sich jetzt wie folgt: $v_1 = e_1$,

$$(A - \lambda E_n)v_2 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} v_2 = e_1 \implies v_2 = e_2,$$

$$(A - \lambda E_n)v_3 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} v_3 = e_2 \implies v_3 = -e_2 + e_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$(A - \lambda E_n)v_4 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} v_4 = v_3 \implies v_4 = \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{pmatrix},$$

$$(A - \lambda E_n)v_5 = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} v_5 = v_4 \implies v_5 = \begin{pmatrix} 0 \\ -1 \\ 3 \\ -3 \\ 1 \end{pmatrix}.$$

Die Hauptvektorkette $v_1, \dots, v_5$ bildet eine Basis des K^5. Stellen wir A in dieser Basis dar, so ergibt sich mit

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad S^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

die besonders einfache Darstellung

$$S^{-1}AS = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix},$$

welche die Form der Matrix J aus dem vorigen Beispiel hat. $\triangle$

10 Die Jordansche Normalform

Nun wollen wir zeigen, dass V komplett in Haupträume zerfällt, wenn das Minimalpolynom komplett in Linearfaktoren zerfällt. Dazu ein vorbereitender Satz:

Satz 10.1. *Es seien $\lambda_1, \dots, \lambda_k \in K$ paarweise verschieden, $a_1, \dots, a_k \geq 1$ und $p = (t - \lambda_1)^{a_1} \cdots (t - \lambda_k)^{a_k} \in K[t]$. Dann gilt für $f \in \operatorname{Hom}(V, V)$*

$$\operatorname{Kern}(p(f)) = \operatorname{Kern}((f - \lambda_1 \operatorname{id})^{a_1}) \oplus \cdots \oplus \operatorname{Kern}((f - \lambda_k \operatorname{id})^{a_k}).$$

Beweis.
Wir machen eine vollständige Induktion nach k.

Für $k = 1$ sind beide Seiten gleich. Sei also $k \geq 2$. Wir setzen

$$p_1 = (t - \lambda_1)^{a_1}, \quad p_2 = \tfrac{p}{p_1} = (t - \lambda_2)^{a_2} \cdots (t - \lambda_k)^{a_k}$$

und zeigen $\operatorname{Kern}(p(f)) = \operatorname{Kern}(p_1(f)) \oplus \operatorname{Kern}(p_2(f))$ (woraus die Behauptung folgt).

Da $p_1(f)$ und $p_2(f)$ kommutieren, gilt $p(f) = p_2(f) \circ p_1(f) = p_1(f) \circ p_2(f)$ und wir sehen $\operatorname{Kern}(p_1(f)) \subset \operatorname{Kern}(p(f))$ und auch $\operatorname{Kern}(p_2(f)) \subset \operatorname{Kern}(p(f))$.

Zeigen wir jetzt $\operatorname{Kern}(p(f)) \subset \operatorname{Kern}(p_1(f)) + \operatorname{Kern}(p_2(f))$. Da p_1 und p_2 teilerfremde Polynome sind, existieren Polynome $h_1, h_2 \in K[t]$, so dass $p_1h_1 + p_2h_2 = 1$ gilt (vgl. Euklidschen Algorithmus große Übung 1). Es gilt also

$$\operatorname{id} = p_1(f) \circ h_1(f) + p_2(f) \circ h_2(f).$$

Für jedes $v \in V$ gilt also

$$v = \underbrace{p_1(f) \circ h_1(f)(v)}_{=v_2} + \underbrace{p_2(f) \circ h_2(f)(v)}_{v_1}.$$

Ist $v \in \operatorname{Kern}(p(f))$, so folgt

$$\begin{aligned} p_2(f)(v_2) &= p_2(f) \circ p_1(f) \circ h_1(f)(v) = p(f) \circ h_1(f)(v) \\ &= h_1(f) \circ p(f)(v) = 0 \end{aligned}$$

und analog auch $p_1(f)(v_1) = 0$, also $v_1 \in \operatorname{Kern}(p_1(f))$ und $v_2 \in \operatorname{Kern}(p_2(f))$. Es folgt $v \in \operatorname{Kern}(p_1(f)) + \operatorname{Kern}(p_2(f))$.

Schließlich zeigen wir, dass die Summe direkt ist (also, dass $\operatorname{Kern}(p_1(f)) \cap \operatorname{Kern}(p_2(f)) = \{0\}$): Sei also $v \in \operatorname{Kern}(p_1(f)) \cap \operatorname{Kern}(p_2(f))$. Dann ist

$$v = h_1(f) \circ p_1(f)(v) + h_2(f) \circ p_2(f)(v) = 0$$

und der Beweis ist fertig. □

Korollar 10.2 (Hauptraumzerlegung)**.** *Ist V ein endlichdimensionaler K-Vektorraum und $f \in \operatorname{Hom}(V, V)$, so dass das Minimalpolynom*

m_f in Linearfaktoren zerfällt, also $m_f = (t-\lambda_1)^{a_1} \cdots (t-\lambda_k)^{a_k}$ (mit paarweise verschiedenen λ_i), so gilt

$$V = \mathrm{Haupt}(f, \lambda_1) \oplus \cdots \oplus \mathrm{Haupt}(f, \lambda_k).$$

Wir wenden den vorigen Satz auf $p = m_f$ und wegen $m_f(f) = 0$ folgt $\mathrm{Kern}(m_f(f)) = V$. Außerdem ist in diesem Fall $\mathrm{Kern}((f - \lambda_i \,\mathrm{id})^{a_1}) = \mathrm{Haupt}(f, \lambda_i)$.

Damit können wir nun leicht die Umkehrung in Satz 8.5 zeigen:

Korollar 10.3. *Zerfällt das Minimalpolynom m_f in* ***paarweise verschiedene*** *Linearfaktoren, so ist f diagonalisierbar.*

Korollar 10.2 gilt mit $a_1 = \cdots = a_k = 1$ und daher sind alle Haupträume Eigenräume.

Definition 10.4. Eine Matrix der Form

$$\begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix} \in K^{n \times n}$$

heißt *Jordan-Block* der Größe n zum Eigenwert λ.

Eine Block-Diagonalmatrix deren Diagonalblöcke alle Jordan-Blöcke sind, heißt *Jordan-Matrix*.

Nun können wir die Jordansche Normalform angeben.

Satz 10.5 (Jordansche Normalform). *Es sei $f \in \mathrm{Hom}(V, V)$ so dass das Minimalpolynom m_f in Linearfaktoren zerfällt. Dann gibt es eine Basis B von V, so dass die Matrixdarstellung $M_B^B(f)$ eine Jordan-Matrix ist, d.h. es gilt*

$$M_B^B(f) = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_l \end{pmatrix}$$

mit Jordan-Blöcken $J_1, \ldots, J_l$.

Beweis.
Das Minimalpolynom sei $m_f = (t-\lambda_1)^{a_1} \cdots (t-\lambda_k)^{a_k}$. Nach Korollar 10.2 ist $V = \mathrm{Haupt}(f, \lambda_1) \oplus \cdots \oplus \mathrm{Haupt}(f, \lambda_k)$. In jedem Hauptraum wählen wir eine Hauptraumkettenbasis nach Satz 9.4. Jede Hauptraumkette liefert nach Satz 9.3 einen Jordan-Block in der Matrix-Darstellung. Alle diese Hauptraumkettenbasen zusammen liefern nach Satz 9.4 und Satz 10.1 eine Basis von V mit der geforderten Eigenschaft. □

Vorgehen zum Berechnen der Jordanschen Normalform einer Abbildung f:

Schritt 1: Faktorisiere das charakteristische Polynom χ_f soweit in Linearfaktoren wie möglich. Zerfällt es nicht vollständig in Linearfaktoren, so lässt sich die Abbildung/Matrix nicht in Jordansche Normalform bringen.

Schritt 2: Zerfällt das charakteristische Polynom vollständig in Linearfaktoren, lese die Eigenwerte $\lambda_1, \dots, \lambda_k$ ab und Berechne Basen der Eigenräume $\operatorname{Eig}(f, \lambda_1)$, ..., $\operatorname{Eig}(f, \lambda_k)$.

Schritt 3: Berechne zu jedem Basisvektor in jedem Eigenraum die zugehörige Hauptvektorkette.

Schritt 4: Alle die Vektoren dieser Hauptvektorketten bilden die gesuchte Basis.

Bemerkung. Die Jordansche Normalform ist nicht eindeutig, sondern nur bis auf Reihenfolge der Jordan-Blöcke.

Bei der Wahl der Basis hat man noch mehr Freiheiten: Man kann beliebige Basen der Eigenräume wählen und außerdem hat man bei der Berechunung der Hauptvektorketten die Freiheit bei der Wahl des Hauptvektors der nächsthöheren Ordnung. ■

Beispiel.

Wir betrachten die Matrix

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

Mit etwas Aufwand lässt sich (unter Ausnutzung der Null-Blöcke) das charakteristische Polynom berechnen:

$$\chi_A = -(t-2)^3(t+1)^4.$$

Wir haben also die beiden Eigenwerte $\lambda_1 = 2$ und $\lambda_2 = -1$. Berechnen wir Basen der Eigenräume. Es gilt

$$\operatorname{Kern}(A - 2E_7) = \operatorname{Kern} \begin{pmatrix} -3 & 1 & 1 & -3 & -3 & -4 & 0 \\ 1 & -4 & -1 & 3 & 4 & 4 & 0 \\ -1 & 1 & -2 & -6 & -8 & -8 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 & 1 & 4 & 0 \\ 0 & 0 & 0 & -3 & -1 & -4 & 0 \\ 0 & 0 & 0 & 3 & 1 & 4 & 0 \end{pmatrix}.$$

Der Rang dieser Matrix ist 5, d.h. es gibt zwei linear unabhängige

Eigenvektoren und wir nehmen

$$v_1 = \begin{pmatrix} -1 \\ 1 \\ -2 \\ 1 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Jetzt berechnen wir die Hauptvektorketten: Für v_1 bekommen wir als eine mögliche Lösung von $(A - 2E_7)v_{1,2} = v_1$ den Vektor

$$v_{1,2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Die Hauptvektorkette zu v_2 bricht sofort ab, da $(A - 2E_7)w = v_2$ keine Lösung hat.

Der Eigenraum zu $\lambda_2 = -1$ ist

$$\operatorname{Kern}(A + E_7) = \operatorname{Kern} \begin{pmatrix} 0 & 1 & 1 & -3 & -3 & -4 & 0 \\ 1 & -1 & -1 & 3 & 4 & 4 & 0 \\ -1 & 1 & 1 & -6 & -8 & -8 & 0 \\ 0 & 0 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 & 4 & 4 & 0 \\ 0 & 0 & 0 & -3 & -1 & -1 & 0 \\ 0 & 0 & 0 & 3 & 1 & 4 & 3 \end{pmatrix}.$$

Wieder ist der Rang 5 und wir bekommen die beiden linear unabhängigen Eigenvektoren

$$v_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 1 \\ -1 \\ 2 \\ 0 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

Berechnen wir die Hauptraumketten: Zu v_3 bekommen wir aus $(A + E_7)v_{3,2} = v_3$ zum Beispiel

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

und aus $(A + E_7)v_{3,3} = v_{3,2}$ auch noch

$$v_{3,3} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

wonach die Hauptvektorkette abbricht. Die letzte Hauptvektorkette zu v_4 müssen wir nicht mehr ausrechnen, da wir mit

$$\{v_1, v_{1,2}, v_2, v_3, v_{3,2}, v_{3,3}, v_4\}$$

schon eine Basis des $\mathbb{R}^7$ gefunden haben. Wir setzen

$$\begin{aligned} S &= \begin{pmatrix} v_1 & v_{1,2} & v_2 & v_3 & v_{3,2} & v_{3,3} & v_4 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & 0 & -1 \\ -2 & 0 & 0 & -1 & 1 & 1 & 2 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

und man kann nachrechnen

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

was die gesuchte Jordansche Normalform ist. △

11 Die adjunkte Matrix und die Cramersche Regel

Im Fall von 2×2 Matrizen gibt es eine explizite Formel für die inverse Matrix, nämlich

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

In gewissen Sinn lässt sich das auf größere Matrizen verallgemeinern. Der Schlüssel dazu sind Minoren (welche wir schon aus der Laplace-Entwicklung der Determinante kennen):

Definition 11.1. Ist $A \in K^{n\times n}$, so bezeichnen wir mit $A_{ij} \in K^{n\times n}$ die Matrix, die aus A entsteht, indem man den Eintrag a_{ij} durch 1 ersetzt, und alle anderen Einträge in der i-ten Zeile und j-ten Spalte durch 0 ersetzt, also

$$A_{ij} = \begin{pmatrix} a_{11} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix}.$$

Weiterhin sei

$$\tilde{a}_{ij} = \det(A_{ij})$$

der i, j-te *Kofaktor* von A, die Matrix $C = (\tilde{a}_{ij})$ heißt *Kofaktormatrix* zu A und

$$A^{\#} = C^T = (\tilde{a}_{ji}) = (\det(A_{ji}))$$

heißt *adjunkte Matrix* oder *komplementäre Matrix* zu A.

Achtung: Den Begriff *adjunkte* nicht mit *adjungierte* verwechseln!

Im vorigen Semester hatten wir mit A_{ij} die Matrix bezeichnet, welche durch Streichen der i-ten Zeile und j-ten Spalte aus A hervorgeht. Nennen wir diese A'_{ij}, so sehen wir (indem wir die i-te Zeile und j-Spalte nach vorne Permutieren), das $\det(A_{ij}) = (-1)^{i+j}\det(A'_{ij})$.

Lemma 11.2. *Es sei $A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \in K^{n\times n}$. Dann gilt*

$$\det(A_{ij}) = \det\begin{pmatrix} a_1 & \cdots & a_{j-1} & e_i & a_{j+1} & \cdots & a_n \end{pmatrix}.$$

Beweis.
Man kann die Matrix $\begin{pmatrix} a_1 & \cdots & a_{j-1} & e_i & a_{j+1} & \cdots & a_n \end{pmatrix}$ durch Addition von Vielfachen der j-ten Spalte in die Matrix A_{ij} überführen. □

Satz 11.3. *Es sei $A \in K^{n\times n}$ und $A^{\#}$ die adjunkte Matrix. Dann gilt*

$$A^{\#}A = AA^{\#} = \det(A)E_n.$$

Beweis.
Wir berechnen den i, k-ten Eintrag von $A^{\#}A$ und nutzen das vorige

Lemma:

$$
\begin{aligned}
\sum_{j=1}^{n} a_{ij}^{\#} a_{jk} &= \sum_{j=1}^{n} a_{jk} \det(A_{ji}) \\
&= \sum_{j=1}^{n} a_{jk} \det \begin{pmatrix} a_1 & \cdots & a_{i-1} & e_j & a_{i+1} & \cdots & a_n \end{pmatrix} \\
&= \det \begin{pmatrix} a_1 & \cdots & a_{i-1} & \sum_{j=1}^{n} a_{jk} e_j & a_{i+1} & \cdots & a_n \end{pmatrix} \\
&= \det \begin{pmatrix} a_1 & \cdots & a_{i-1} & a_k & a_{i+1} & \cdots & a_n \end{pmatrix}.
\end{aligned}
$$

Diese Determinante ist gleich Null, immer wenn $i \neq k$ (denn dann enthält sie zwei gleiche Spalten) gilt. Für $i = k$ ist sie gleich $\det(A)$. Wir haben also

$$
\sum_{j=1}^{n} a_{ij}^{\#} a_{jk} = \delta_{ik} \det(A),
$$

was die Behauptung zeigt. □

Damit haben wir die exlizite Regel für die inverse Matrix:

Korollar 11.4. *Ist $A \in K^{n \times n}$ invertierbar, so gilt*

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

Beispiel.

Ist $n = 2$, so ist die adjunkte Matrix zu

$$
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
$$

gegeben durch

$$
\begin{aligned}
A^{\#} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{21} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{pmatrix} &= \begin{pmatrix} \det \begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix} & \det \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix} \\ \det \begin{pmatrix} 0 & 1 \\ c & 0 \end{pmatrix} & \det \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.
\end{aligned}
$$

Wir bekommen also die Inverse (im Fall der Invertierbarkeit) als

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

△

Als *Cramersche Regel* bezeichnet man die Anwendung dieser expiziten Formel auf das Lösen von quadratischen Gleichungssystemen:

Beispiel.

Es sei $A \in K^{n\times n}$ invertierbar. Dann hat das Gleichungssystem $Ax = b$ für jedes b die eindeutige Lösung $x = A^{-1}b$. Sind $a_1,\dots,a_n$ die Spalten von A, so hat die j-te Spalte von A^{-1} die Einträge

$$\frac{\det(A_{ji})}{\det(A)} = \frac{\det\begin{pmatrix}a_1 & \dots a_{i-1} & e_j & a_{i+1} & \cdots & a_n\end{pmatrix}}{\det(A)}.$$

Daher ist der i-te Eintrag der Lösung x

$$\begin{aligned}x_i &= \sum_{j=1}^n b_j \frac{\det(A_{ji})}{\det(A)}\\ &= \frac{\det\left(a_1 \quad \dots a_{i-1} \quad \sum_{j=1}^n b_j e_j \quad a_{i+1} \quad \cdots \quad a_n\right)}{\det(A)}\\ &= \frac{\det\begin{pmatrix}a_1 & \dots a_{i-1} & b & a_{i+1} & \cdots & a_n\end{pmatrix}}{\det(A)}.\end{aligned}$$

△

Beispiel.

Es sei

$$A = \begin{pmatrix}3 & -1 & -1\\ 2 & 2 & 3\\ 2 & 4 & 1\end{pmatrix}, \quad b = \begin{pmatrix}1\\0\\0\end{pmatrix}.$$

Wir berechnen (z.B. mit Laplace-Entwicklung nach der ersten Zeile)

$$\det(A) = 3\det\begin{pmatrix}2 & 3\\ 4 & 1\end{pmatrix} + 1\det\begin{pmatrix}2 & 3\\ 2 & 1\end{pmatrix} - 1\det\begin{pmatrix}2 & 2\\ 2 & 4\end{pmatrix} = -38.$$

Nun können wir die Komponenten der Lösung von $Ax = b$ ausrechnen

$$x_1 = -\tfrac{1}{38}\det\begin{pmatrix}1 & -1 & -1\\ 0 & 2 & 3\\ 0 & 4 & 1\end{pmatrix} = -\tfrac{1}{38}\det\begin{pmatrix}2 & 3\\ 4 & 1\end{pmatrix} = \tfrac{10}{38} = \tfrac{5}{19}.$$

$$x_2 = -\tfrac{1}{38}\det\begin{pmatrix}3 & 1 & -1\\ 2 & 0 & 3\\ 2 & 0 & 1\end{pmatrix} = \tfrac{1}{38}\det\begin{pmatrix}2 & 3\\ 2 & 1\end{pmatrix} = -\tfrac{4}{38} = -\tfrac{2}{19}.$$

$$x_3 = -\tfrac{1}{38}\det\begin{pmatrix}3 & -1 & 1\\ 2 & 2 & 0\\ 2 & 4 & 0\end{pmatrix} = -\tfrac{1}{38}\det\begin{pmatrix}2 & 2\\ 2 & 4\end{pmatrix} = -\tfrac{4}{38} = -\tfrac{2}{19}.$$

Was in diesem Fall noch vergleichsweise übersichlich ist, wird bei größeren Gleichungssystemen schnell sehr aufwändig. △

Die Cramersche Regel ist für die praktische Lösung von Gleichungssystem nicht von großer Bedeutung, da es immer andere Verfahren gibt, die besser geeignet sind. Sie kann allerdings in theoretischen Fragestellung hilfreich sein.

Mit Hilfe der adjunkten Matrix können wir nun den Satz von Cayley-Hamilton in voller Allgemeinheit beweisen (unser Beweis aus Abschnitt 7 war nur für den Fall von trigonalisierbaren Matrizen).

Satz 11.5. *Ist V ein endlichdimensionaler K-Vektorraum und $f \in \operatorname{Hom}(V,V)$, so gilt $\chi_f(f) = 0$. Entsprechend gilt für $A \in K^{n\times n}$ immer $\chi_A(A) = 0_{n\times n}$.*

Beweis.
Wir beweisen den Satz für den Fall von Matrizen $A \in K^{n\times n}$ und setzen $B(t) = (A - tE_n)^{\#}$. Nach Satz 11.3 gilt

$$(A - tE_n)B(t) = \det(A - tE_n)E_n = \chi_A(t)E_n.$$

Die Matrix $B(t)$ hat als Einträge Polynome in t (mit Grad kleiner oder gleich $n-1$) und daher gibt es Matrizen $B_0, \dots, B_{n-1} \in K^{n\times n}$, so dass

$$B(t) = \sum_{i=0}^{n-1} t^i B_i.$$

Damit formen wir weiter um

$$\begin{aligned}
\chi_A(t)E_n &= (A - tE_n)B(t)\\
&= (A - tE_n)\sum_{i=0}^{n-1} t^i B_i\\
&= \sum_{i=0}^{n-1} t^i AB_i - \sum_{i=0}^{n-1} t^{i+1}B_i\\
&= AB_0 + \sum_{i=1}^{n-1} t^i(AB_i - B_{i-1}) - t^n B_{n-1}.
\end{aligned}$$

Dies ist eine Gleichheit von Polynomen mit Matrizen als Koeffizienten, und diese können nur gleich sein, wenn alle Koeffizientenmatrizen gleich sind. Koeffizientenvergleich ergibt, wenn wir $\chi_A = (-1)^n t^n + c_{n-1}t^{n-1} + \dots + c_1 t + c_0$ schreiben, die Gleichungen

$$\begin{aligned}
-B_{n-1} &= (-1)^n E_n,\\
AB_i - B_{i-1} &= c_i E_n \ (1 \le i \le n-1),\\
AB_0 &= c_0 E_n.
\end{aligned}$$

Wir multiplizieren diese Gleichungen jeweils von links mit A^i, summieren und bekommen

$$\begin{aligned}
&-A^n B_{n-1} + \sum_{i=1}^{n-1}(A^{i+1}B_i - A^i B_{i-1}) + AB_0\\
&= (-1)^n A^n + c_{n-1}A^{n-1} + \dots + c_1 A + c_0 E_n\\
&\qquad = \chi_A(A).
\end{aligned}$$

Die linke Seite ist eine Teleskopsumme und alle Terme heben sich weg, so dass wir die gewünschte Gleichung $0_{n\times n} = \chi_A(A)$ bekommen. $\square$

12 Projektionen

Es bezeichne $\mathbb{K}$ die reellen oder komplexen Zahlen. Wir betrachten den reellen bzw. komplexen Vektorraum $\mathbb{K}^n$ mit dem Standardskalarprodukt und der induzierten euklidschen bzw. unitären Norm. Für $a \in \mathbb{K}^n$ betrachten wir die Menge

Also $\langle x, y\rangle = x^T y = \sum x_k y_k$ bzw $\langle z, w\rangle = w^H z = \sum z_k \overline{w_k}$.

$$H = \{x \mid \langle a, x\rangle = 0\} = \{a\}^\perp.$$

Ist $a \neq 0$, so ist diese Menge ein $n-1$-dimensionaler Unterraum, also eine Hyperebene. Zu $x \in \mathbb{K}^n$ suchen wir das Element in H, welches den kleinsten Abstand zu x hat. Wir argumentieren geometrisch: Der Vektor a steht senkrecht auf H, daher ist der Punkt in H, der den geringsten Abstand von x hat, von der Form $y = x + ta$ mit $t \in \mathbb{K}$. Wir müssen nur noch t so bestimmen, dass $y \in H$ gilt. Dies liefert die Gleichung

$$\langle a, x + ta\rangle = 0,$$

welche die Lösung

$$t = -\tfrac{\langle x,a\rangle}{\langle a,a\rangle} = -\tfrac{\langle x,a\rangle}{\|a\|^2}$$

hat. Die Abbildung, die x dem nächstgelegenen Punkt auf H zuordnet, ist also

Ist $\mathbb{K} = \mathbb{C}$, so müssen wir Antilinearität in der zweiten Komponente und $\overline{\langle a, x\rangle} = \langle x, a\rangle$ berücksichtigen.

$$P_H x = x - \tfrac{\langle x,a\rangle}{\|a\|^2} a,$$

welche wir auch schreiben können als

$$P_H x = x - \tfrac{1}{\|a\|^2}(a^H x)a = (E_n - \tfrac{1}{\|a\|^2} aa^H)x.$$

Diese Abbildung heißt *Orthogonalprojektion* auf H (und die zugehörige Matrix heißt *Projektionsmatrix*) und hat die Eigenschaft

$$\begin{aligned}
P_H^2 &= P_H P_H \\
&= (E_n - \tfrac{1}{\|a\|^2} aa^H)(E_n - \tfrac{1}{\|a\|^2} aa^H) \\
&= E_n - 2\tfrac{1}{\|a\|^2} aa^H + \tfrac{1}{\|a\|^4} a \underbrace{a^H a}_{=\|a\|^2} a^H \\
&= E_n - 2\tfrac{1}{\|a\|^2} aa^H + \tfrac{1}{\|a\|^2} aa^H \\
&= E_n - \tfrac{1}{\|a\|^2} aa^H = P_H.
\end{aligned}$$

Wir können auch „nicht-orthogonal“ bzw. „schief“ auf H projizieren. Nehmen wir dazu irgendein b, welches nicht in H liegt und bestimmen für ein x ein Skalar t, so dass $x + tb \in H$ gilt, d.h. wir lösen $\langle a, x + tb\rangle = 0$ nach t auf und bekommen

$$t = -\tfrac{\langle x,a\rangle}{\langle b,a\rangle}.$$

Die zugehörige schiefe Projektion ist also

$$Px = x - \tfrac{\langle x,a\rangle}{\langle b,a\rangle} b = (E_n - \tfrac{1}{\langle b,a\rangle} ba^H)x$$

und auch hier gilt $P^2 = P$.

Diese Eigenschaft nimmt man als Grundlage für den abstrakten Begriff der Projektion, auch auf Vektorräumen ohne Skalarprodukt.

Definition 12.1. Eine lineare Abbildung $P \in \mathrm{Hom}(V, V)$ auf einem Vektorraum V heißt *Projektion*, falls $P^2 = P$ gilt.

In den Übungsaufgaben (Blatt 4) wurde schon gezeigt, dass Projektionen die folgenden Eigenschaften haben:

- $P|_{\mathrm{Bild}(P)} = \mathrm{id}_{\mathrm{Bild}(P)}$
- $\mathrm{Bild}(P)$ und $\mathrm{Kern}(P)$ sind komplementäre Unterräume, also $\mathrm{Bild}(P) \cap \mathrm{Kern}(P) = \{0\}$ und $\mathrm{Bild}(P) + \mathrm{Kern}(P) = V$ (bzw. $\mathrm{Bild}(P) \oplus \mathrm{Kern}(P) = V$).

Aus diesen beiden Punkten können wir schon die Eigenräume ablesen: Da P auf seinem Bild die Identität ist, ist

$$\mathrm{Eig}(P, 1) = \mathrm{Bild}(P)$$

und außerdem ist sowieso $\mathrm{Eig}(P, 0) = \mathrm{Kern}(P)$. Da Bild und Kern komplementär sind, folgern wir, dass es keine weiteren Eigenwerte als 1 und 0 gibt und dass P diagonalisierbar ist. Wir könnten dies auch wie folgt einsehen:

Lemma 12.2. *Das Minimalpolynom einer Projektion P ist ein Teiler von $x^2 - x$.*

Beweis.
Es gilt $P^2 = P$, d.h. $P^2 - P = 0$. Also hat das Polynom $p = t^2 - t$ die Eigenschaft $p(P) = 0$. Daher muss m_P ein Teiler von p sein. □

Eine weitere Art, ist das folgende Lemma:

Lemma 12.3. *Für $\lambda \notin \{0, 1\}$ gilt $(P - \lambda\,\mathrm{id})^{-1} = -\frac{1}{\lambda(\lambda-1)}P - \frac{1}{\lambda}\,\mathrm{id}$.*

Beweis.
Wir rechnen einfach nach

$$\begin{aligned}(P - \lambda\,\mathrm{id})\big(-\tfrac{1}{\lambda(\lambda-1)}P - \tfrac{1}{\lambda}\,\mathrm{id}\big) &= -\tfrac{1}{\lambda(\lambda-1)}P^2 - \tfrac{1}{\lambda}P + \tfrac{1}{\lambda-1}P + \mathrm{id}\\ &= \mathrm{id}\,.\end{aligned}$$

□

Ist ein Unterraum U von V gegeben als $U = \langle v_1, \dots, v_k\rangle$ mit linear unabhängigen $v_1, \dots, v_k$ so können wir eine Projektion P mit $\mathrm{Bild}(P) = U$ wie folgt angeben: Wir ergänzen $v_1, \dots, v_k$ zu einer Basis $v_1, \dots, v_n$ und definieren P auf den Basisvektoren als

$$Pv_i = \begin{cases} v_i, & \text{falls } 1 \leq i \leq k \\ 0, & \text{sonst.}\end{cases}$$

In Räumen mit Skalarprodukt können wir auch allgemein von orthogonalen Projektionen reden:

Definition 12.4. Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann heißt eine Projektion P auf V eine *Orthogonalprojektion*, wenn P selbstadjungiert ist, d.h. für alle $x, y \in V$ gilt $\langle Px, y \rangle = \langle x, Py \rangle$ (oder alternativ $P^H = P$ bzw. $P^T = P$).

Satz 12.5. *Es sei P eine Orthogonalprojektion und $U = \mathrm{Bild}(P)$. Dann gilt:*

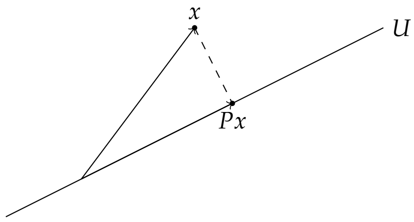
i) Für alle $y \in U$ gilt $x - Px \perp y$

ii) Für alle x gilt $x - Px \perp Px$.

iii) Für alle $y \in U$ gilt

$$\|x - Px\| \leq \|x - y\|.$$

Dieser Satz hat eine einfache graphische Interpretation:



i) Die Verbindungslinie $x - Px$ steht senkrecht auf dem Unterraum U und iii) der Punkt Px ist der Punkt aus U, welcher x am nächsten liegt.

Beweis.

i) Ist $y \in U$, so gilt $y = Py$. Wir rechnen einfach nach

$$\begin{aligned}\langle x - Px, y \rangle &= \langle x - Px, Py \rangle \\ &= \langle P(x - Px), y \rangle \\ &= \langle Px - P^2 x, y \rangle = \langle Px - Px, y \rangle = 0.\end{aligned}$$

ii) Klar, da $Px \in U$.

iii) Für alle y gilt $0 \leq \|Px - y\|^2 = \|Px\| - 2\langle Px, y \rangle + \|y\|^2$. Ist $y \in U = \mathrm{Bild}(P)$, so gilt $y = Py$ und wegen $P^H = P$ folgt

$$0 \leq \|Px\| - 2\langle x, y \rangle + \|y\|^2.$$

Dies formen wir um zu

$$-\|Px\|^2 \leq -2\langle x, y \rangle + \|y\|^2$$

und durch Addition von $\|x\|^2$ auf beiden Seiten bekommen wir

$$\|x\| - \|Px\|^2 \leq \|x - y\|^2.$$

Schließlich bemerken wir noch $\langle x, Px \rangle = \langle x, P^2 x \rangle = \langle Px, Px \rangle = \|Px\|^2$ und folgern

$$\begin{aligned}\|x - Px\|^2 &= \langle x - Px, x - Px \rangle \\ &= \|x\|^2 - \langle x, Px \rangle - \langle Px, x \rangle + \|Px\|^2 \\ &= \|x\|^2 - \|Px\|^2,\end{aligned}$$

woraus die Behauptung folgt.

□

Korollar 12.6. *Ist P die Orthogonalprojektion auf U, so ist $E_n - P$ die Orthogonalprojektion auf $U^\perp$.*

Wegen $P^2 = P$ und $P^H = P$ gilt

$$(E_n - P)^2 = E_n - P - P + P^2 = E_n - P,$$
$$(E_n - P)^H = E_n - P^H = E_n - P.$$

Wegen i) des vorigen Satzes ist $(E_n - P)x \perp U$. Zusammen folgt, dass $E_n - P$ eine Orthonogonalprojektion mit $\mathrm{Bild}(E_n - P) = U^\perp$ ist.

Satz 12.7. *Es sei $A \in \mathbb{K}^{m\times n}$ mit $m \geq n$ und $\mathrm{Rang}(A) = n$. Dann ist $P_{\mathrm{Bild}(A)} = A(A^HA)^{-1}A^H$ die Orthogonalprojektion auf $\mathrm{Bild}(A)$.*

Beweis.
Zuerst zeigen wir, dass $A^HA \in K^{n\times n}$ invertierbar ist. Ist $A^HAx = 0$, so gilt für alle y

$$0 = \langle A^HAx, y\rangle = \langle Ax, Ay\rangle.$$

Das bedeutet aber, dass Ax senkrecht auf dem Bild von A steht. Das ist aber nur möglich, wenn $x = 0$ gilt. Also ist A^HA injektiv, und also auch, da A^HA quadratisch ist, bijektiv.

Mit dem orthogonalen Komplement $\mathrm{Bild}(A)^\perp$ gilt $\mathbb{K}^m = \mathrm{Bild}(A) \oplus \mathrm{Bild}(A)^\perp$. D.h., jedes $y \in \mathbb{K}^m$ lässt sich schreiben als $y = u + v$ mit $u \in \mathrm{Bild}(A)$ und $v \in \mathrm{Bild}(A)^\perp$. Wir zeigen nun, dass für jedes solche y gilt $P_{\mathrm{Bild}(A)}y = u$. Dazu bemerken wir, dass für $v \in \mathrm{Bild}(A)^\perp$ gilt $A^Hv = 0$.

$$\begin{aligned} v \in \mathrm{Bild}(A)^\perp &\iff \forall x \in \mathbb{K}^n : \langle Ax, v\rangle = 0 \\ &\iff \forall x \in \mathbb{K}^n : \langle x, A^Hv\rangle = 0. \end{aligned}$$

Außerdem gibt es für $u \in \mathrm{Bild}(A)$ ein w, so dass $u = Aw$. Es folgt also

$$\begin{aligned} A(A^HA)^{-1}A^H(u+v) &= A(A^HA)^{-1}A^Hu + A(A^HA)^{-1}A^Hv \\ &= A(A^HA)^{-1}A^HAw + 0 \\ &= Aw = u. \end{aligned}$$

Schließlich sehen wir

$$\begin{aligned} P_{\mathrm{Bild}(A)}P_{\mathrm{Bild}(A)} &= A(A^HA)^{-1}A^HA(A^HA)^{-1}A^H \\ &= A(A^HA)^{-1}A^H = P_{\mathrm{Bild}(A)} \end{aligned}$$

und

$$\begin{aligned} P_{\mathrm{Bild}(A)}^H &= (A(A^HA)^{-1}A^H)^H = (A^H)^H((A^HA)^{-1})^HA^H \\ &= A(A^HA)^{-1}A^H = P_{\mathrm{Bild}(A)}. \end{aligned}$$

□

Korollar 12.8. *Ist $A \in K^{m\times n}$ mit $n \geq m$ und $\mathrm{Rang}(A) = m$, so ist $P_{\mathrm{Kern}(A)} = E_m - A^H(AA^H)^{-1}A$ die Orthogonalprojektion auf $\mathrm{Kern}(A)$.*

Es gilt $\mathrm{Kern}(A) = \mathrm{Bild}(A^H)^\perp$. Wir wenden Korollar 12.6 und Satz 12.7 an und bekommen

$$P_{\mathrm{Kern}(A)} = P_{\mathrm{Bild}(A^H)^\perp} = E_m - P_{\mathrm{Bild}(A^H)} = E_m - A^H(AA^H)^{-1}A.$$

13 Die QR-Zerlegung

Die QR-Zerlegung einer Matrix $A \in \mathbb{K}^{m\times n}$ ist die Darstellung als

$$A = QR,$$

wobei $Q \in \mathbb{K}^{m\times m}$ eine orthogonale (bzw. unitäre) Matrix ist und $R \in \mathbb{K}^{m\times n}$ eine obere Dreiecksmatrix ist. Dabei nennen wir eine rechteckige Matrix R eine obere Dreiecksmatrix, wenn die Einträge r_{ij} für $i > j$ Null sind.

Erinnere: Eine Matrix $Q \in \mathbb{R}^{n\times n}$ heißt orthogonal, wenn $Q^T Q = E_n$ und $Q \in \mathbb{C}^{n\times n}$ heißt unitär, wenn $Q^H Q = E_n$.

Bevor wir zeigen, dass jede Matrix eine QR-Zerlegung hat, illustrieren wir, wie man mit Hilfe einer QR-Zerlegung ein quadratisches Gleichungssystem lösen kann. Ist $A = QR$, so ist, wegen $Q^{-1} = Q^T$ (bzw. $= Q^H$), das Gleichungssystem $Ax = b$ äquivalent zu

$$Rx = Q^T b \quad (\text{bzw.} = Q^H b).$$

Auch über- und unterbestimmte Systeme lassen sich mit Hilfe der QR-Zerlegung lösen, wie wir später illustrieren.

Im Folgenden wird nur der komplexe Fall beschrieben. Im reellen Fall muss nur ^{H} durch ^{T} ersetzt werden.

Um zu zeigen, dass es für jede Matrix eine QR-Zerlegung gibt, formulieren wir noch einmal das Orthogonalisierungsverfahren nach Gram-Schmidt. Dazu benutzen wir diesmal Projektionen auf eindimensionale Teilräume: Ist $u \in \mathbb{K}^m$, so ist

$$P_u x = \frac{\langle x, u\rangle}{\|u\|^2} u$$

die orthogonale Projektion auf $\langle\{u\}\rangle$.

Ist u normiert (also $\|u\| = 1$), so gilt $P_u x = \langle x, u\rangle u$.

Offensichtlich ist $\text{Bild}(P_u) = \langle\{u\}\rangle$. Außerdem ist (im komplexen Fall)

$$P_u a = \frac{u^H a}{\|u\|^2} u = \tfrac{1}{\|u\|^2} u u^H a,$$

d.h. P_u wird durch die Matrix $\|u\|^{-2} u u^H$ beschrieben.
Es folgt

$$P_u^2 = \|u\|^{-2} u u^H \|u\|^{-2} u u^H = \|u\|^{-4} u \underbrace{u^H u}_{=\|u\|^2} u^H = P_u$$

und

$$P_u^H = (\|u\|^{-2} u u^H)^H = \|u\|^{-2} u u^H = P_u.$$

Sei $A \in \mathbb{K}^{m\times n}$ mit den Spalten $a_1, \dots, a_n \in \mathbb{K}^m$. Wir konstruieren nun eine Orthonormalbasis von $\text{Bild}(A) = \langle\{a_1, \dots, a_n\}\rangle$: Setze $q_1 = a_1/\|a_1\|$ und für $i = 2, \dots, n$ iteriere

1. Orthogonalisierunng:
$$\tilde{q}_i = a_i - P_{q_1} a_i - \cdots - P_{q_{i-1}} a_i.$$
2. Test auf lineare Unabhängigkeit: Ist $\tilde{q}_i = 0$, breche ab.
3. Normierung:
$$q_i = \tilde{q}_i / \|\tilde{q}_i\|.$$

Wir sammeln ein paar wichtige Aussagen über die Vektoren, die das Gram-Schmidt-Verfahren produziert:

Lemma 13.1. *Sind die Vektoren $a_1, \dots, a_k \in \mathbb{K}^m$ linear unabhängig, so gilt für die Vektoren $q_1, \dots, q_k$ aus dem Gram-Schmidt-Verfahren:*

1. *Die Vektoren $q_1, \dots, q_k$ sind orthonormal.*
2. *Für alle $i = 1, \dots, k$ ist a_i eine Linearkombination aus $q_1, \dots, q_i$.*
3. *Für alle $i = 1, \dots, k$ ist q_i eine Linearkombination aus $a_1, \dots, a_i$.*

Beweis.
Wir zeigen alle drei Aussagen gleichzeitig per Induktion: Für $k = 1$: Nach Konstruktion ist q_1 normiert und damit orthonormal. Ebenso gilt $a_1 = \|a_1\| q_1$ und $q_1 = a_1 / \|a_1\|$ was die Punkte 2 und 3 zeigt.

Gelten die Behauptungen 1, 2 und 3 nun für $i - 1$ mit $i < k$. Wäre $\tilde{q}_i = 0$, so wäre a_i eine Linearkombination aus $q_1, \dots, q_{i-1}$. Da aber jeder dieser Vektoren eine Linearkombination aus $a_1, \dots, a_{i-1}$ ist, wäre auch a_i eine Linearkombination dieser, was der linearen Unabhängigkeit von $a_1, \dots, a_k$ widerspricht.

In Schritt 3 des Verfahrens wird Normierung garantiert. Es bleibt also nur Orthogonalität zu zeigen. Wir zeigen $q_i \perp q_j$ für $j < i$: Da die $q_1, \dots, q_{i-1}$ nach Induktionsannahme orthonormal sind, gilt

$$\begin{aligned}
\langle q_i, q_j \rangle &= q_j^H (a_i - P_{q_1} a_i - \cdots - P_{q_{i-1}} a_i) \\
&= q_j^H (a_i - \langle a_i, q_1 \rangle q_1 - \cdots - \langle a_i, q_{i-1} \rangle q_{i-1}) \\
&= \langle a_i, q_j \rangle - \langle a_i, q_1 \rangle \langle q_1, q_j \rangle - \cdots - \langle a_i, q_{i-1} \rangle \langle q_{i-1}, q_j \rangle \\
&= \langle a_i, q_j \rangle - \langle a_i, q_j \rangle = 0.
\end{aligned}$$

Jetzt zeigen wir, dass a_i eine Linearkombination der $q_1, \dots, q_i$ ist: Aus dem Orthogonalisierungsschritt folgt

$$a_i = \tilde{q}_i + P_{q_1} a_i + \cdots + P_{q_{i-1}} a_i,$$

und da $P_{q_j} a_i$ ein Vielfaches von q_j ist, folgt die Behauptung.

Ebenso sehen wir aus dem Orthogonalisierungsschritt, dass $\tilde{q}_i$ eine Linearkombination aus $a_i, q_1, \dots, q_{i-1}$ ist. Nach Induktionsannahme sind die $q_1, \dots, q_{i-1}$ Linearkombinationen der $a_1, \dots, a_{i-1}$ was die Behauptung zeigt. □

Für die numerische Berechnung der QR-Zerlegung wird nicht das Gram–Schmidt-Verfahren benutzt. Es gibt andere Verfahren, welche auf orthogonalen Spiegelungen (Householder-Reflektionen) basieren, die numerisch stabiler sind.

Korollar 13.2. *Bricht das Gram-Schmidt-Verfahren im k-ten Schritt ab, so sind die Vektoren $a_1, \dots, a_k$ nicht linear unabhängig.*

Der Abbruch liegt an der Bedingung $\tilde{q}_k = 0$, und das bedeutet

$$a_k = P_{q_1} a_k + \cdots + P_{q_{k-1}} a_k.$$

Da jedes q_j Linearkombination der $a_1, \dots, a_{j-1}$ ist, ist auch a_k eine Linearkombination der $a_1, \dots, a_{k-1}$.

Das Gram-Schmidt-Verfahren ist also nicht nur ein Verfahren, um eine Orthonormalbasis für einen Spann zu konstruieren, sondern auch ein Test auf lineare Unabhängigkeit. Außerdem zeigt das Gram-Schmidt-Verfahren die Existenz der QR-Zerlegung und gibt eine Vorschrift, diese zu berechnen.

Satz 13.3. *Es sei $A \in \mathbb{K}^{m\times n}$. Dann existieren ein orthonormales/unitäres $Q \in \mathbb{K}^{m\times m}$ und eine obere Dreicksmatix $R \in \mathbb{K}^{m\times n}$, so dass $A = QR$.*

Beweis.
Es sei $k = \mathrm{Rang}(A)$ (beachte, dass $k \leq \min(m,n)$). Das Gram-Schmidt-Verfahren, angewendet auf die Spalten $a_1,\dots,a_n$ von A (falls es überhaupt welche gibt), liefert orthonormale Vektoren $q_1,\dots,q_k$ mit

$$a_i = \langle a_i, q_1\rangle q_1 + \cdots + \langle a_i, q_{i-1}\rangle q_{i-1} + \|\tilde{q}_i\| q_i,$$

für alle $i = 1,\dots,k$. Wir definieren $\tilde{Q} = \begin{pmatrix} q_1 & \cdots & q_k \end{pmatrix} \in \mathbb{K}^{m\times k}$ und eine Matrix $\tilde{R} \in \mathbb{K}^{k\times k}$ durch

$$R_{ij} = \begin{cases} \langle a_i, q_j\rangle, & i < j \\ \|\tilde{q}_j\|, & i = j \\ 0, & \text{sonst.} \end{cases}$$

Da $\tilde{Q}$ eine Basis des Spanns von A enthält, sind auch die weiteren Spalten $a_{k+1},\dots,a_n$ Linearkombinationen von $q_1,\dots,q_k$. Es gibt also Koeffizienten R_{ij}, so dass auch für $j = k+1,\dots,n$ gilt

$$a_j = R_{1j}q_1 + \cdots + R_{kj}q_k.$$

Wir ergänzen $\tilde{R}$ durch diese Koeffizienten zu $\tilde{\tilde{R}} \in \mathbb{K}^{k\times n}$ und bekommen

$$A = \tilde{Q}\tilde{\tilde{R}}.$$

Dies ist noch nicht die gesuchte QR-Zerlegung, da $\tilde{Q}$ und $\tilde{R}$ noch nicht die richtige Größe haben (es gilt $\tilde{Q} \in \mathbb{K}^{m\times k}$ und $\tilde{\tilde{R}} \in \mathbb{K}^{k\times n}$). Wir ergänzen noch

$$Q = \begin{pmatrix} \tilde{Q} & q_{k+1} & \cdots & q_m \end{pmatrix} \in \mathbb{K}^{m\times m}, \quad R = \begin{pmatrix} \tilde{\tilde{R}} \\ 0 \end{pmatrix} \in \mathbb{K}^{m\times n},$$

wobei $q_{n+1},\dots,q_m$ beliebig sind, so dass $Q \in K^{m\times m}$ orthonormal/unitär ist. □

Betrachten wir ein überbestimmtes Gleichungssystem $Ax = b$, d.h. $A \in \mathbb{K}^{m\times n}$ mit $m > n$. Wir berechnen die QR-Zerlegung von A und bekommen das System $QRx = b$, welches wir umformen zu

$$Rx = Q^H b.$$

Das System hat also die Form

$$\begin{pmatrix} \diagdown \end{pmatrix} x = Q^H b$$

Die fehlende untere Spitze von R kann zustande kommen, wenn A nicht vollen Rang hat.

Nun können wir, analog zum Vorgehen beim Gaußschen Eliminationsverfahren, leicht ablesen, ob es eine Lösung gibt: Hat $Q^H b$ in den letzten Einträgen nur Nullen, existiert eine Lösung, sonst nicht.

Mit Hilfe der QR-Zerlegung lässt sich die Projektion auf das Bild von A ausrechnen. Wir benutzen dazu eine reduzierte Form der QR-Zerlegung: Für ein $A \in \mathbb{K}^{m\times n}$ mit $m \geq n$ ist nämlich

$$A = QR = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} \diagdown \\ \\ \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1.$$

Hat A maximalen Rang (also linear unabhängige Spalten), so ist R_1 invertierbar und mit Satz 12.7 folgt

$$\begin{aligned} P_{\mathrm{Bild}(A)} &= A(A^H A)^{-1} A^H \\ &= Q_1 R_1 (R_1^H Q_1^H Q_1 R_1)^{-1} R_1^H Q_1^H \\ &= Q_1 R_1 (R_1^H R_1)^{-1} R_1^H Q_1^H \\ &= Q_1 R_1 (R_1)^{-1} (R_1^H)^{-1} R_1^H Q_1^H \\ &= Q_1 Q_1^H. \end{aligned}$$

Wir können also (vgl. große Übung 6) auch überbestimmte Gleichungssysteme im kleinste-Quadrate-Sinne lösen: Die Lösung von $Ax = P_{\mathrm{Bild}(A)} b$ ist nämlich (wieder unter Nutzung von $A = Q_1 R_1$) gegeben durch die Lösung von

$$Q_1 R_1 x = P_{\mathrm{Bild}(A)} b = Q_1 Q_1^H b,$$

also

$$x = R_1^{-1} Q_1^H Q_1 Q_1^H b = R_1^{-1} Q_1^H b.$$

Wir können auch direkt, ohne Rückgriff auf Projektionen, zeigen, dass dadurch die kleinste-Quadrate-Lösung gegeben ist:

Lemma 13.4. *Es seien $A \in \mathbb{K}^{m\times n}$ mit $m \geq n$ und $\mathrm{Rang}(A) = n$ mit QR-Zerlegung*

$$A = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix},$$

mit $Q_1 \in \mathbb{K}^{m\times n}$ und $R_1 \in \mathbb{K}^{n\times n}$. Dann ist die Lösung von

$$\min_{x\in\mathbb{K}^n} \|Ax - b\|^2$$

mit $b \in \mathbb{K}^m$ gegeben durch $x = R_1^{-1} Q_1^H b$.

Beweis.
Für eine orthonormale/unitäre Matrix Q gilt

$$\|Qy\|^2 = \langle Qy, Qy\rangle = \langle y, Q^H Qy\rangle = \langle y, y\rangle = \|y\|^2.$$

Also gilt

$$\|Ax - b\|^2 = \|QRx - b\|^2 = \|Rx - Q^H b\|^2.$$

Mit der Aufteilung $Q = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$ sehen wir

$$Rx - Q^H b = \begin{pmatrix} R_1 \\ 0 \end{pmatrix} x - \begin{pmatrix} Q_1^H \\ Q_2^H \end{pmatrix} b = \begin{pmatrix} R_1 x - Q_1^H b \\ -Q_2^H b \end{pmatrix}.$$

Es gilt also

$$\|Ax - b\|^2 = \|R_1 x - Q_1^H b\|^2 + \|Q_2^H b\|^2.$$

In dieser Form können wir das Minimum ablesen: Der erste Term wird minimal für $x = R_1^{-1} Q_1^H b$ (dann ist er Null) und der zweite Term ist unabhängig von x. □