

Prof. Dr. Sebastian Riedel,  
wiss. Mitarbeit: Dr. Leonie Brinker

**Modul 61811**

**Mathematische Grund-  
lagen von Data Science**

**LESEPROBE**

Fakultät für  
**Mathematik und  
Informatik**

---

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

# Mathematische Grundlagen von Data Science

Sebastian Riedel

Version vom 8. Mai 2024

Wissenschaftliche Mitarbeit: Leonie Brinker

# Inhaltsverzeichnis

<b>0</b>	<b>Notation</b>	<b>9</b>
0.1	Zahlen . . . . .	9
0.2	Vektoren . . . . .	9
0.3	Matrizen . . . . .	11
	<b>Vorwort</b>	<b>12</b>
<b>1</b>	<b>Daten in hochdimensionalen Räumen</b>	<b>13</b>
1.1	Motivation: Image Compression . . . . .	13
1.2	Eigenwerte, Eigenvektoren und Diagonalisierbarkeit . . . . .	16
1.2.1	Eigenwerte und Eigenvektoren . . . . .	17
1.2.2	Diagonalisierbarkeit . . . . .	20
1.3	Singulärwertzerlegung . . . . .	23
1.3.1	Reduzierte Singulärwertzerlegung . . . . .	27
1.3.2	Anwendung: Matrixapproximation . . . . .	28
1.4	Beispielaufgabe: Singulärwertzerlegung und Gram-Schmidt . . . . .	30
<b>2</b>	<b>Hochdim. Wahrscheinlichkeitstheorie</b>	<b>35</b>
2.1	Motivation: Unvollständige Information und „unsichere“ Daten . . . . .	35
2.2	Zufallsvektoren . . . . .	37
2.2.1	Kovarianz und Korrelation . . . . .	39
2.2.2	Unabhängigkeit . . . . .	41
2.2.3	Bedingte Zufallsvariablen und bedingte Dichten . . . . .	43
2.3	Konvergenzarten von Zufallsvariablen . . . . .	44
2.4	Ungleichungen und Konvergenzsätze . . . . .	47
2.5	Die mehrdimensionale Normalverteilung . . . . .	55
2.6	Der bedingte Erwartungswert . . . . .	57
2.7	Motivation, Teil 2: Bayes-Filter und Mondlandung . . . . .	64
<b>3</b>	<b>Optimierung</b>	<b>69</b>
3.1	Motivation: Support Vector Machines . . . . .	70
3.2	Konvexe Mengen und konvexe Funktionen . . . . .	73
3.3	Optimierungsprobleme ohne Nebenbedingungen . . . . .	78
3.3.1	Extremstellen differenzierbarer Funktionen . . . . .	79
3.3.2	Gradientenabstiegsverfahren . . . . .	80
3.3.3	Das Newton-Verfahren . . . . .	85
3.3.4	Das stochastische Gradientenabstiegsverfahren . . . . .	86
3.4	Optimierungsprobleme unter Nebenbedingungen . . . . .	90

3.4.1	Dualität . . . . .	94
3.4.2	Der Satz von Karush-Kuhn-Tucker . . . . .	97
<b>4</b>	<b>Geometrie hochdimensionaler Räume</b>	<b>101</b>
4.1	Motivation: Dimensionsreduktion und „bag of words“ . . . . .	101
4.2	Eigenschaften $d$ -dimensionaler Kugeln . . . . .	104
4.2.1	Volumen der $d$ -dimensionalen Kugel . . . . .	106
4.3	Orthogonalität . . . . .	110
4.4	Normalverteilte Zufallsvariablen in hochdimensionalen Räumen .	113
4.4.1	Anwendung: Dimensionsreduktion bei Nächster-Nachbarn-Suche . . . . .	115
<b>5</b>	<b>Statistik I</b>	<b>119</b>
5.1	Motivation: Clustering und Schätzen . . . . .	119
5.2	Das statistische Modell . . . . .	121
5.3	Parameterschätzung . . . . .	124
5.3.1	Gütekriterien für Schätzer . . . . .	125
5.3.2	Das Maximum-Likelihood-Prinzip . . . . .	129
5.3.3	Bayes-Schätzer - Idee und Definition . . . . .	132
5.4	Konfidenzintervalle . . . . .	136
5.4.1	Motivation, Definition und Konstruktion . . . . .	136
5.4.2	Normalapproximation . . . . .	141
5.4.3	Konfidenzintervalle im Binomialmodell . . . . .	144
5.4.4	Ordnungsintervalle . . . . .	145
<b>6</b>	<b>Statistik II</b>	<b>149</b>
6.1	Motivation: ANOVA und Machine Learning . . . . .	149
6.2	Statistische Tests . . . . .	154
6.2.1	Grundlegende Begriffe . . . . .	154
6.2.2	Alternativtests . . . . .	157
6.2.3	Beste einseitige Tests . . . . .	158
6.2.4	Parametertests im Gaußschen Produktmodell . . . . .	164
6.2.5	Chiquadrat-Anpassungstest . . . . .	167
6.2.6	Chiquadrat-Test auf Unabhängigkeit . . . . .	169
<b>7</b>	<b>Stochastische Prozesse, Zeitreihen</b>	<b>175</b>
7.1	Motivation: Google's PageRank Algorithmus . . . . .	176
7.2	Allgemeine stochastische Prozesse und Irrfahrten . . . . .	180
7.3	Markovketten . . . . .	183
7.3.1	Stochastische Matrizen und Übergangswahrscheinlichkeiten	185
7.3.2	Stopzeiten und die starke Markoveigenschaft . . . . .	190
7.3.3	Stationäre Zustände und statistisches Gleichgewicht . . .	196
	<b>Anhang</b>	<b>203</b>
<b>A</b>	<b>Mehrdimensionale Analysis</b>	<b>203</b>
A.1	Der Raum $\mathbb{R}^d$ . . . . .	203
A.1.1	Norm und Skalarprodukt . . . . .	203
A.1.2	Topologie . . . . .	206
A.1.3	Konvergenz . . . . .	210

A.2	Mehrdimensionale Abbildungen . . . . .	211
A.2.1	Stetigkeit . . . . .	211
A.2.2	Extremstellen . . . . .	212
A.2.3	Fixpunkte . . . . .	213
A.3	Mehrdimensionale Differentialrechnung . . . . .	213
A.3.1	Totale Differenzierbarkeit . . . . .	213
A.3.2	Partielle Differenzierbarkeit . . . . .	215
A.3.3	Rechenregeln . . . . .	217
A.3.4	Reellwertige Funktionen . . . . .	219
<b>B</b>	<b>Lineare Algebra</b>	<b>223</b>
B.1	Objekte, LGS und Gaußalgorithmus . . . . .	223
B.1.1	Der Rang einer Matrix . . . . .	227
B.1.2	Lineare Gleichungssysteme . . . . .	232
B.2	Quadratische Matrizen . . . . .	236
B.2.1	Inverse Matrizen . . . . .	236
B.2.2	Die Determinante . . . . .	238
B.2.3	Orthogonale und unitäre Matrizen . . . . .	242
<b>C</b>	<b>Wahrscheinlichkeitstheorie</b>	<b>245</b>
C.1	Wahrscheinlichkeitsräume . . . . .	245
C.1.1	Bedingte Wahrscheinlichkeiten und Unabhängigkeit . . . . .	251
C.2	Reelle Zufallsvariablen . . . . .	254
C.3	Erwartungswert und Varianz . . . . .	260
<b>D</b>	<b>Quantile</b>	<b>265</b>
D.1	Standardnormalverteilung . . . . .	265
D.2	Studentsche $t$ -Verteilung . . . . .	267
D.3	$\chi^2$ -Verteilung . . . . .	268

# Lektion 1

## Daten in hochdimensionalen Räumen

In der Data Science stehen wir vor der Herausforderung, eine große Menge von Daten zu verarbeiten und daraus Informationen zu extrahieren. Damit Daten maschinell verarbeitet werden können, müssen sie zuerst zu Zahlenfolgen codiert und gespeichert werden. Häufig sind die Daten auch als Matrizen gegeben. Ein Beispiel ist eine Zeitreihe von Messergebnissen eines Experiments oder einer Simulation. Jede Spalte der Matrix entspricht dann dem Messergebnis zu einem gewissen Zeitpunkt. Der entsprechende Spaltenvektor kann dabei eine sehr hohe Dimension haben. Man denke etwa an Wettermessdaten, die an sehr vielen Orten auf der Erde gleichzeitig erhoben werden. Stark vereinfacht ist dies in Abbildung 1.1 dargestellt. Ein anderes Beispiel ist die Codierung eines Videos, bei dem die Spaltenvektoren die Bildinformationen an einem festen Zeitpunkt beschreiben, wie in Abbildung 1.2. Der Umgang mit solch enorm hochdimensionalen Matrizen stellt eine große Herausforderung dar, was sowohl die Speicherung als auch die Berechnung relevanter Informationen angeht. Glücklicherweise sind die Informationen, die man aus den Daten extrahieren möchte, häufig in niedrigdimensionalen Räumen eingebettet. Deswegen ist es manchmal möglich, die Matrix niedrigdimensional zu approximieren, ohne die relevanten Informationen zu verlieren. Ein gängiges Mittel ist hier die *Singulärwertzerlegung*, die das wohl wichtigste Thema dieses Kapitels ist. Durch sie wird es möglich sein, Informationen, die durch hochdimensionale Matrizen gegeben sind, effektiv zu komprimieren, um die Daten anschließend verarbeiten zu können.

### 1.1 Motivation: Image Compression

Ein großer Teil dieser Kurseinheit handelt von Zerlegungen von (Daten-)Matrizen. In diesem Abschnitt betrachten wir ein Anwendungsbeispiel, bei dem solche Zerlegungen zum Einsatz kommen. Wir benutzen dabei schon einmal einige Begriffe, die in dieser Kurseinheit noch genauer erklärt werden. Diese Stichworte sind kursiv gedruckt. In diesem Abschnitt geht es darum, eine Intuition für die betrachteten Themen zu erhalten, Sie müssen also bei der ersten Lektüre noch nicht alle einzelnen Schritte genau nachvollziehen können.

Eine besondere Stärke der so genannten *Singulärwertzerlegung* ist, dass sie nu-

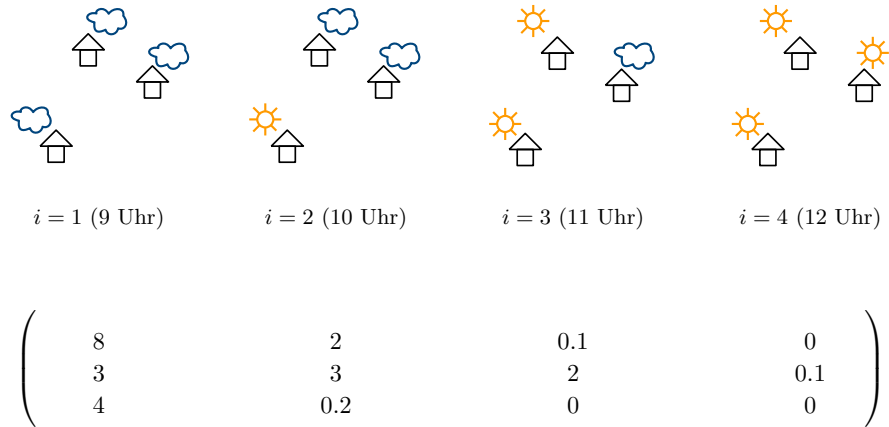


Abbildung 1.1: Bei der Erfassung der Wetterlage könnte zum Beispiel die Niederschlagsmenge (gemessen in  $l/m^2$  in der vergangenen Stunde) in einer Matrix festgehalten werden. Die (kleine) Matrix in diesem Beispiel zeigt, wie ein Gewitter über drei Wetterstationen (Zeilen) in vier Stunden (Spalten) vorbeizieht.

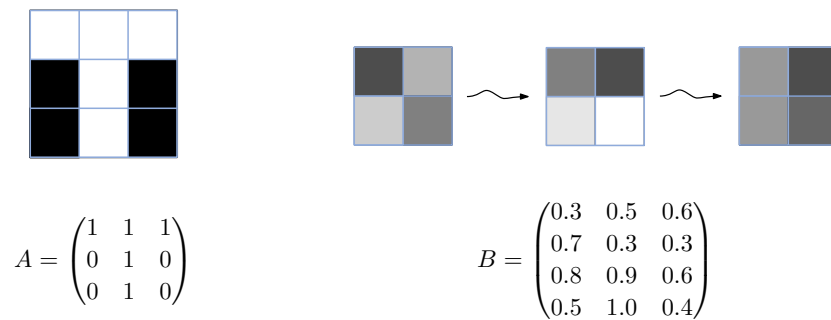


Abbildung 1.2: Die Matrix  $A$  codiert hier die Weiß- und Schwarzwerte eines Bildes mit neun Pixeln. Die Matrix  $B$  beschreibt vier Pixel eines Videos (jede Spalte steht für einen Frame, d.h. ein Einzelbild zu einem festen Zeitpunkt). Für Bildmaterial mit höherer Auflösung wird diese Art der Matrizen schnell sehr groß.



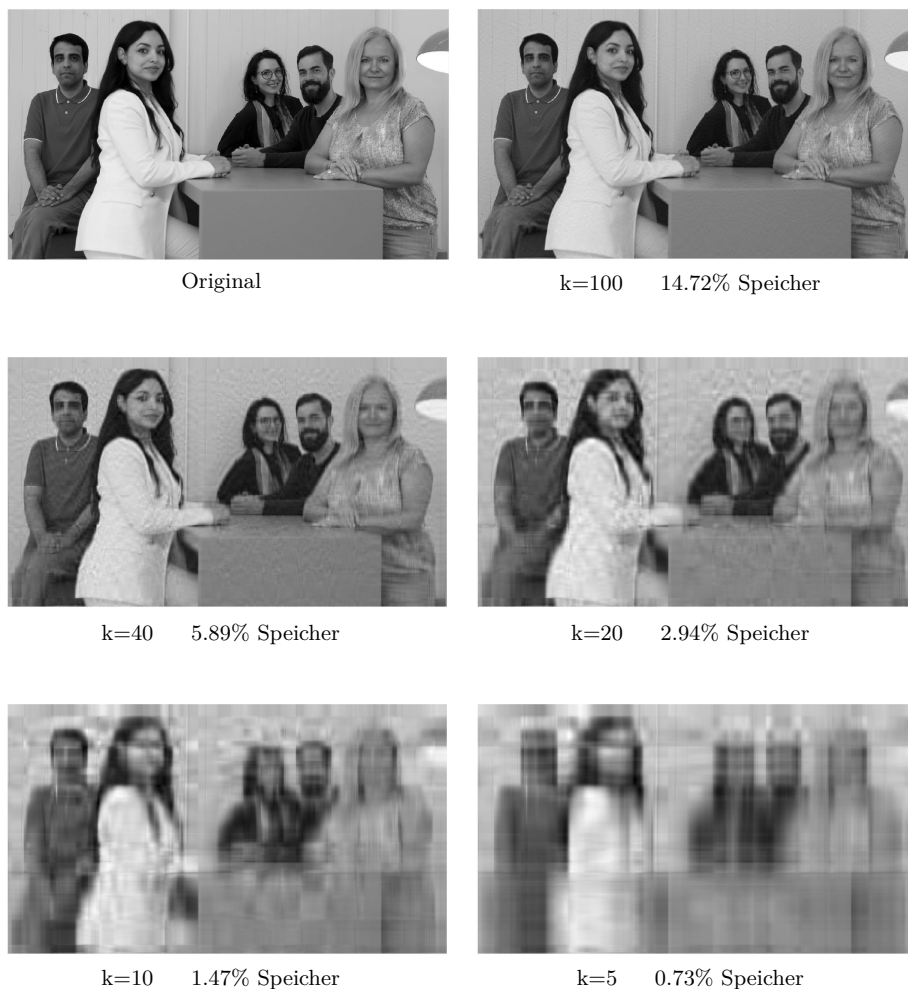


Abbildung 1.3: Singulärwertzerlegung bei einem Schwarzweißbild der Arbeitsgruppe Angewandte Stochastik aus dem Jahr 2023.

merisch stabil ist<sup>1</sup> und garantiert existiert. Insbesondere bietet sie die Möglichkeit einer Approximation durch *niedrigrangige* Matrizen, welche in einem gewissen Sinn optimal ist. Im Kontext der Hauptkomponentenanalyse oder Principal Component Analysis (PCA) bedeutet dies, dass eine große Anzahl statistischer Informationen durch eine geringe Anzahl möglichst aussagekräftiger Komponenten ausgedrückt wird.

Wir stellen uns nun vor, die  $n \times m$ -Matrix  $A$  enthalte die Graustufenwerte eines Bildes – zum Beispiel die des Originalbildes in Abbildung 1.3. Im Beispiel gilt  $n = 1065$  und  $m = 1877$ , also müssen insgesamt  $nm = 1999005$  Werte

<sup>1</sup>Numerisch stabil bedeutet, grob gesagt, dass das Verfahren nahezu unempfindlich ist gegenüber kleinen Störungen. Dies ist wichtig, da wir i.A. nicht davon ausgehen können, dass uns die Daten fehlerfrei gegeben sind. Details und Beispiele hierzu kann man in [KL80] nachlesen.

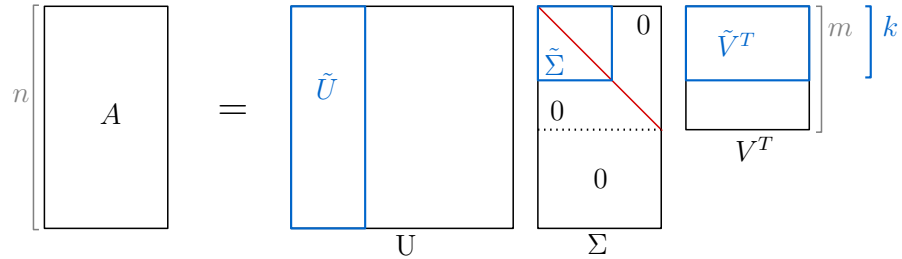


Abbildung 1.4: Die Idee der Singulärwertzerlegung bei der PCA ist, dass (in einem gewissen Sinn)  $A(k) = \tilde{U}\tilde{\Sigma}\tilde{V}^T \approx A$  gilt und dass die „abgeschnittenen“ Matrizen sehr viel weniger Speicherplatz beanspruchen.

gespeichert werden. Wie wir im Weiteren sehen werden, kann  $A$  als Produkt  $A = U\Sigma V^T$  geschrieben werden, wobei  $U$  und  $V$  *orthogonale Matrizen* sind und  $\Sigma$  eine Blockmatrix bestehend aus einer *Diagonalmatrix* und der Nullmatrix. Die Diagonalmatrix enthält die so genannten *Singulärwerte*. Dies sind bestimmte Werte, die absteigend nach der Größe sortiert sind und die Originalmatrix charakterisieren (vergleichbar mit Eigenwerten). Im Fall  $n > m$  ist dies in Abbildung 1.4 skizziert (schwarze Kästen). Die Diagonale der Matrix  $\Sigma$ , auf der die Singulärwerte stehen, ist rot markiert. Wie man diese Zerlegung berechnen kann, werden wir in dieser Kurseinheit sehen. Die Idee ist nun, nur die ersten (und größten)  $k < \min\{n, m\}$  Singulärwerte zu betrachten, sowie die ersten  $k$  Spalten der Matrix  $U$  und die ersten  $k$  Zeilen der Matrix  $V^T$ . Die Matrix  $A(k)$ , die aus der Multiplikation der verkleinerten Matrizen  $\tilde{U} \in \mathbb{R}^{n \times k}$ ,  $\tilde{\Sigma} \in \mathbb{R}^{k \times k}$  und  $\tilde{V}^T \in \mathbb{R}^{k \times m}$  entsteht, hat dann höchstens *Rang*  $k$  und kann als Approximation an  $A$  verstanden werden. Die verkleinerten Matrizen sind in Abbildung 1.4 blau markiert. Der Vorteil ist, dass der benötigte Speicherplatz stark reduziert wird. Insbesondere müssen nun nur noch  $nk + k + mk$  (Matrix  $\tilde{U}$  +  $k$  Singulärwerte + Matrix  $\tilde{V}$ ) Werte gespeichert werden. Gleichzeitig bleiben die „relevantesten“ Bildinformationen erhalten: wir werden in Abschnitt 1.3 sehen, dass  $A(k)$  die beste Rang- $k$ -Approximation an  $A$  bezüglich der *Frobeniusnorm* ist. In der Abbildung 1.3 sieht man die zu den Matrizen  $A(k)$  gehörigen, komprimierten Bilder für  $k = 100, 40, 20, 10, 5$ . Abbildung 1.5 zeigt die Singulärwerte der Matrix  $A$ . Im Folgenden leiten wir (unter anderem) her, warum dieses Komprimierungsverfahren mathematisch funktioniert. Am Ende des Abschnitts 1.3 kommen wir noch einmal auf das Beispiel zurück. Ein anderes Beispiel findet man in [BK19].

## 1.2 Eigenwerte, Eigenvektoren und Diagonalisierbarkeit

In dieser Kurseinheit wird es vornehmlich darum gehen, wie große Matrizen effektiv durch niedrigrangige Matrizen approximiert werden können. In diesem

## 1.2. EIGENWERTE, EIGENVEKTOREN UND DIAGONALISIERBARKEIT 17

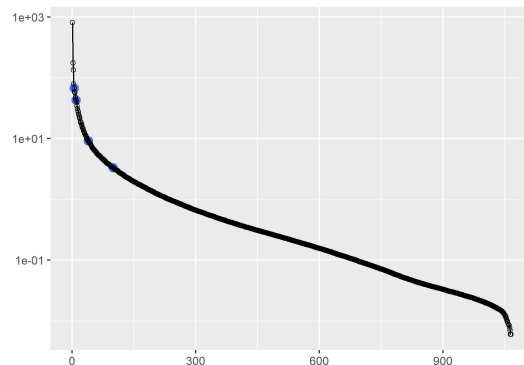


Abbildung 1.5: Die 1065 Singulärwerte des Originalbildes. Die  $y$ -Achse ist hier logarithmisch, die Werte fallen also gerade am Anfang sehr stark ab. Die Stellen  $k = 5, 10, 20, 40, 100$  sind blau markiert.

ersten Teilabschnitt werden wir uns zunächst auf *quadratische* Matrizen konzentrieren, d.h. auf solche, die die selbe Anzahl von Zeilen und Spalten haben.

### 1.2.1 Eigenwerte und Eigenvektoren

Im Folgenden wird es sinnvoll sein, Matrizen zu betrachten, die entweder reelle oder komplexe Einträge haben. Am Ende werden uns in erster Linie reellwertige Matrizen interessieren. Allerdings werden die Eigenwerte, die wir in diesem Teilabschnitt kennenlernen, im Allgemeinen komplexe Zahlen sein, auch wenn es sich um reellwertige Matrizen handelt. Später werden wir aber auch Kriterien kennenlernen, wann die gesuchten Eigenwerte reell sind.

**Definition 1.2.1.** Sei  $A \in \mathbb{K}^{n \times n}$ . Gibt es eine Zahl  $\lambda \in \mathbb{K}$  und einen Vektor  $v \in \mathbb{K}^n$ ,  $v \neq 0$ , mit der Eigenschaft

$$Av = \lambda v,$$

so heißt  $\lambda$  *Eigenwert* zum *Eigenvektor*  $v$ . Die Menge

$$\sigma(A) := \{\lambda \in \mathbb{K} : \lambda \text{ ist Eigenwert von } A\}$$

heißt *Spektrum* von  $A$ .

**Übungsaufgabe 1.2.2.** Sei  $\lambda \in \mathbb{K}$  ein Eigenwert einer Matrix  $A \in \mathbb{K}^{n \times n}$ . Man zeige, dass die Menge

$$\{v \in \mathbb{K}^n : v \text{ ist ein Eigenvektor von } A \text{ zum Eigenwert } \lambda\} \cup \{0\}$$

ein Untervektorraum von  $\mathbb{K}^n$  ist.

**Definition 1.2.3.** Sei  $A \in \mathbb{K}^{n \times n}$  eine Matrix und  $\lambda \in \mathbb{K}$  ein Eigenwert von  $A$ . Dann heißt der Unterraum

$$E(\lambda) := \{v \in \mathbb{K}^n : v \text{ ist ein Eigenvektor von } A \text{ zum Eigenwert } \lambda\} \cup \{0\}$$

*Eigenraum* zum Eigenwert  $\lambda$ . Die Dimension von  $E(\lambda)$  heißt *geometrische Vielfachheit* des Eigenwerts  $\lambda$ .

Bevor wir uns der Frage nach der Existenz von Eigenwerten widmen, machen wir folgende Beobachtung:

**Satz 1.2.4.** *Seien  $\lambda_1, \lambda_2 \in \mathbb{K}$  Eigenwerte einer Matrix  $A \in \mathbb{K}^{n \times n}$  mit Eigenvektoren  $v_1, v_2 \in \mathbb{K}^n$  und es gelte  $\lambda_1 \neq \lambda_2$ . Dann sind  $v_1$  und  $v_2$  linear unabhängig.*

*Beweis.* Wir nehmen an,  $v_1$  und  $v_2$  seien linear abhängig, d.h. es existiert eine eindeutige Zahl  $\mu \in \mathbb{K} \setminus \{0\}$  so dass  $\mu v_1 = v_2$  ist. Wenden wir nun  $A$  auf beiden Seiten der Gleichung an, so erhalten wir  $\lambda_1 \mu v_1 = \lambda_2 v_2$ . Da  $\mu$  eindeutig ist, muss dann  $\lambda_1 = \lambda_2$  gelten.  $\square$

Allgemeiner gilt das folgende Resultat:

**Satz 1.2.5.** *Seien  $v_1, \dots, v_k \in \mathbb{K}^n$  Eigenvektoren von  $A \in \mathbb{K}^{n \times n}$  mit Eigenwerten  $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ . Sind  $\lambda_1, \dots, \lambda_k$  paarweise verschieden, so sind  $v_1, \dots, v_k$  linear unabhängig.*

Insbesondere folgt aus dem Satz, dass eine Matrix  $A \in \mathbb{K}^{n \times n}$  nicht mehr als  $n$  paarweise verschiedene Eigenwerte haben kann, da mehr als  $n$  Vektoren in  $\mathbb{K}^n$  notwendigerweise linear abhängig sind.

Wir wollen nun der Frage nachgehen, wann Eigenwerte existieren und wie man diese finden kann.

**Definition 1.2.6.** Sei  $A \in \mathbb{K}^{n \times n}$  und  $I \in \mathbb{K}^{n \times n}$  die Einheitsmatrix. Dann heißt das Polynom

$$p_A(z) := \det(A - zI)$$

in der Variablen  $z \in \mathbb{K}$  *charakteristisches Polynom* von  $A$ . Hierbei bezeichnet  $\det$  die Determinante einer Matrix, siehe Appendix B.2.2.

**Übungsaufgabe 1.2.7.** *Man zeige, dass  $p_A(z)$  tatsächlich ein Polynom vom Grad  $n$  ist.*

**Satz 1.2.8.** *Sei  $A \in \mathbb{K}^{n \times n}$ . Dann sind die folgenden Aussagen äquivalent:*

- (i)  *$A$  besitzt den Eigenwert  $\lambda \in \mathbb{K}$ ,*
- (ii)  *$\lambda$  ist Nullstelle des charakteristischen Polynoms, d.h.  $p_A(\lambda) = 0$ .*

*Beweis.*  $\lambda$  ist ein Eigenwert von  $A$  genau dann wenn es ein  $v \neq 0$  gibt mit  $Av = \lambda v$ . Dies ist äquivalent dazu, dass  $(A - \lambda I)v = 0$  ist. Mit anderen Worten:  $\text{Ker}(A - \lambda I) \neq \{0\}$ . Nach dem Rangsatz (Satz B.1.13) ist das gleichbedeutend mit der Aussage, dass  $\text{rank}(A - \lambda I) < n$  ist. Nach Satz B.2.18 ist dies aber äquivalent zu der Aussage  $p_A(\lambda) = \det(A - \lambda I) = 0$ .  $\square$

Wir haben die Frage nach der Existenz von Eigenwerten also auf die Frage nach der Existenz von Nullstellen von Polynomen verlagert. Der folgende Satz gibt eine Antwort im Fall  $\mathbb{K} = \mathbb{C}$ .

**Theorem 1.2.9** (Fundamentalsatz der Algebra). *Sei  $P(z)$  ein komplexes Polynom vom Grad  $n$ , d.h.  $P$  habe Koeffizienten in  $\mathbb{C}$  und  $z$  sei ebenfalls eine komplexe Zahl. Dann hat  $P$  genau  $n$  Nullstellen  $\lambda_1, \dots, \lambda_n$  und  $P$  kann geschrieben werden als*

$$P(z) = (z - \lambda_1) \cdots (z - \lambda_n).$$

## 1.2. EIGENWERTE, EIGENVEKTOREN UND DIAGONALISIERBARKEIT 19

*Bemerkung 1.2.10.* 1. Man beachte, dass der obige Satz falsch ist für  $\mathbb{K} = \mathbb{R}$ .

Z.B. hat das reelle Polynom  $P(x) = x^2 + 1$  keine reelle Nullstellen.

2. Jedes Polynom mit reellen Koeffizienten kann aufgefasst werden als komplexes Polynom (da jede reelle Zahl auch eine komplexe Zahl ist). Aus dem Fundamentalsatz der Algebra folgt dann, dass ein Polynom mit reellen Koeffizienten komplexe Nullstellen besitzt. Im Fall  $P(z) = z^2 + 1$  sind dies die Nullstellen  $\lambda_1 = i$  und  $\lambda_2 = -i$ .

Theorem 1.2.9 sagt nicht, dass die Nullstellen  $\lambda_1, \dots, \lambda_n$  paarweise verschieden sein müssen, was i.A. auch nicht stimmt. Zum Beispiel ist

$$Q(x) = x^2 - 2x + 1 = (x - 1)(x - 1),$$

das Polynom  $Q$  besitzt also die Nullstellen  $\lambda_1 = \lambda_2 = 1$ . Um dies etwas genauer zu fassen, führen wir die folgenden Begriffe ein:

**Definition 1.2.11.** Sei  $P$  ein komplexes Polynom mit paarweise verschiedenen Nullstellen  $\lambda_1, \dots, \lambda_k$ . Gilt

$$P(z) = (z - \lambda_1)^{i_1} \cdots (z - \lambda_k)^{i_k}$$

mit Zahlen  $i_1, \dots, i_k \in \mathbb{N}$ , so heißt  $i_j$  die *Vielfachheit* der Nullstelle  $\lambda_j$ . Ist  $P$  das charakteristische Polynom einer Matrix  $A$ , so heißt  $i_j$  die *algebraische Vielfachheit* des Eigenwerts  $\lambda_j$ .

Man beachte, dass notwendigerweise  $k \in \{1, \dots, n\}$  und  $i_1 + \dots + i_k = n$  ist.

**Beispiel 1.2.12.** Das Polynom  $Q(x) = x^2 - 2x + 1$  besitzt die Nullstelle 1 mit Vielfachheit 2 bzw.  $Q$  besitzt die zweifache Nullstelle 1.

Als Folgerung von Satz 1.2.8 und Theorem 1.2.9 erhalten wir die folgende Aussage:

**Satz 1.2.13.** Jede Matrix  $A \in \mathbb{C}^{n \times n}$  besitzt genau  $n$  nicht-notwendigerweise paarweise verschieden Eigenwerte. Sind  $\lambda_1, \dots, \lambda_k$  die paarweise verschiedenen Eigenwerte von  $A$ , so addieren sich deren algebraische Vielfachheiten  $i_1, \dots, i_k$  auf zu  $n$ .

Da jede reelle Matrix  $A \in \mathbb{R}^{n \times n}$  auch als komplexe Matrix aufgefasst werden kann folgt, dass auch jede reelle Matrix komplexe Eigenwerte besitzt. Wann können wir aber sicherstellen, dass die Eigenwerte reell sind? Eine Antwort gibt der nächste Satz.

**Satz 1.2.14.** Sei  $A \in \mathbb{C}^{n \times n}$  hermitesch, also  $A^* = \bar{A}^T = A$ . Dann sind alle Eigenwerte von  $A$  reelle Zahlen. Insbesondere besitzt eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  nur reelle Eigenwerte.

*Beweis.* Sei  $\lambda \in \mathbb{C}$  ein Eigenwert und  $v \in \mathbb{C}^n$  ein Eigenvektor zu  $\lambda$ . Dann ist

$$0 \leq \|v\|^2 |\lambda|^2 = (Av)^* Av = v^* A^* Av.$$

Man beachte, dass nach Voraussetzung  $A^* = \bar{A}^T = A$  ist, also

$$v^* A^* Av = v^* A^2 v = \lambda^2 v^* v = \lambda^2 \|v\|^2.$$

Also ist  $\lambda^2 = |\lambda|^2$  und damit ist  $\lambda$  reell. Eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  ist insbesondere hermitesch, woraus auch die zweite Aussage folgt. □

*Bemerkung 1.2.15.* Man beachte, dass wir nicht schließen können, dass jede reelle symmetrische Matrix auch nur reelle Eigenvektoren hat. Zum Beispiel gilt für die Einheitsmatrix, dass jeder Vektor  $v \in \mathbb{C}^n$  ein Eigenvektor zum Eigenwert 1 ist. Die Einheitsmatrix ist aber natürlich reell und symmetrisch. Zudem gilt: ist  $v$  ein reeller Eigenvektor zu einem Eigenwert  $\lambda$ , so ist  $iv$  ein komplexer Eigenvektor zu  $\lambda$ . Andererseits kann man aber schließen, dass zu jedem reellen Eigenwert  $\lambda$  auch ein reeller Eigenvektor existiert, wie wir in der folgenden Übungsaufgabe tun wollen.

**Übungsaufgabe 1.2.16.** Sei  $A \in \mathbb{R}^{n \times n}$  symmetrisch und  $\lambda \in \mathbb{R}$  ein Eigenwert von  $A$ . Sei

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + i \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = a + ib \in \mathbb{C}^n$$

ein komplexer Eigenvektor zu  $\lambda$  mit  $a, b \in \mathbb{R}^n$ . Man zeige, dass dann auch  $a$  und  $b$  Eigenvektoren zu  $\lambda$  sind.

## 1.2.2 Diagonalisierbarkeit

Wir betrachten nun eine Zerlegung für quadratische Matrizen, die eng mit der später folgenden Singulärwertzerlegung verwandt ist. Zuert definieren wir, was wir unter einer Diagonalmatrix verstehen werden.

**Definition 1.2.17.** Eine Matrix  $D \in \mathbb{C}^{n \times n}$ , auf der nur auf der Diagonalen Elemente ungleich 0 vorkommen, heißt *Diagonalmatrix*. Seien  $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ . Wir setzen dann

$$\text{diag}(\lambda_1, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Im Folgenden legen wir fest, was wir unter Diagonalisierbarkeit verstehen.

**Definition 1.2.18.** Eine Matrix  $A \in \mathbb{C}^{n \times n}$  heißt *diagonalisierbar*, falls es eine invertierbare Matrix  $S \in \mathbb{C}^{n \times n}$  gibt, so dass  $D = S^{-1}AS$  eine Diagonalmatrix ist. Äquivalent dazu ist, dass  $A$  eine Zerlegung der Form  $A = SDS^{-1}$  besitzt.

Gilt  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  und ist  $\lambda_i \neq 0$ , so ist  $\lambda_i$  ein Eigenwert von  $D$  zum Eigenvektor  $e_i$ , wobei  $e_i$  den  $i$ -ten Einheitsvektor bezeichnet.

**Satz 1.2.19.** Sei  $A$  diagonalisierbar mit Diagonalmatrix  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Dann sind  $\lambda_1, \dots, \lambda_n$  genau die Eigenwerte von  $A$ .

*Beweis.* Wir berechnen das charakteristische Polynom von  $A$ . Nach Voraussetzung gibt es eine invertierbare Matrix  $S \in \mathbb{C}^{n \times n}$  mit  $SDS^{-1} = A$ . Es gilt dann

$$\begin{aligned} p_A(z) &= \det(A - zI) = \det(SDS^{-1} - SzIS^{-1}) = \det(S(D - zI)S^{-1}) \\ &= \det(S) \det(D - zI) \det(S^{-1}) = \det(D - zI) \\ &= (\lambda_1 - z) \cdots (\lambda_n - z), \end{aligned}$$

## 1.2. EIGENWERTE, EIGENVEKTOREN UND DIAGONALISIERBARKEIT 21

wobei wir hier den Determinantenmultiplikationssatz verwendet haben. Also hat das charakteristische Polynom von  $A$  genau die Nullstellen  $\lambda_1, \dots, \lambda_n$  woraus die Aussage folgt.  $\square$

Der folgende Satz gibt an, wann eine Matrix  $A \in \mathbb{C}^{n \times n}$  diagonalisierbar ist.

**Satz 1.2.20.**  *$A \in \mathbb{C}^{n \times n}$  ist diagonalisierbar genau dann, wenn für jeden Eigenwert von  $A$  die algebraische Vielfachheit gleich der geometrischen Vielfachheit ist. Dies ist äquivalent dazu, dass sich die Dimensionen der Eigenräume zu den jeweiligen Eigenwerten von  $A$  aufaddieren zu  $n$ .*

Insbesondere folgt, dass  $A$  diagonalisierbar ist, wenn alle Eigenwerte von  $A$  unterschiedlich sind.

*Bemerkung 1.2.21.* Möchte man für eine diagonalisierbare Matrix  $A \in \mathbb{C}^{n \times n}$  eine Matrix  $S \in \mathbb{C}^{n \times n}$  mit  $D = S^{-1}AS$  berechnen, so kann man wie folgt vorgehen: Sind  $\lambda_1, \dots, \lambda_k$  die paarweise verschiedenen Eigenwerte von  $A$ , so wähle man für jeden Eigenraum  $E(\lambda_i)$  eine Basis und definiere  $S$ , indem man als Spaltenvektoren genau diese Basisvektoren wählt. Insbesondere folgt: Ist  $A \in \mathbb{R}^{n \times n}$  eine reelle diagonalisierbare Matrix, die nur reelle Eigenwerte besitzt, so kann man eine invertierbare reelle Matrix  $S \in \mathbb{R}^{n \times n}$  finden, so dass  $S^{-1}AS$  eine Diagonalmatrix ist.

Wir wollen nun reelle Matrizen  $A \in \mathbb{R}^{n \times n}$  betrachten. Wir wissen bereits, dass symmetrische reelle Matrizen auch reelle Eigenwerte besitzen. Wir wissen auch, dass diese Matrizen unter gewissen Annahmen diagonalisierbar sind. Wir werden nun sehen, dass symmetrische Matrizen tatsächlich *immer* diagonalisierbar sind. Es lässt sich sogar noch mehr zeigen. Bevor wir zu der nächsten Definition kommen, erinnern wir an den Begriff einer *orthogonalen Matrix*, siehe Definition B.2.27.

**Definition 1.2.22.** Eine Matrix  $A \in \mathbb{R}^{n \times n}$  heißt *orthogonal diagonalisierbar*, falls eine orthogonale Matrix  $S \in \mathbb{R}^{n \times n}$  existiert, so dass  $D = S^T A S$  eine Diagonalmatrix ist.

**Satz 1.2.23.** *Eine Matrix  $A \in \mathbb{R}^{n \times n}$  ist orthogonal diagonalisierbar genau dann wenn sie symmetrisch ist.*

*Beweis.* Wir nehmen zuerst an, dass  $A$  orthogonal diagonalisierbar ist, d.h. es existiere eine Diagonalmatrix  $D$  und eine orthogonale Matrix  $S$  mit  $A = SDS^T$ . Es gilt dann

$$A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T = A,$$

da  $D^T = D$  für jede Diagonalmatrix. Damit ist  $A$  also symmetrisch.

Die Rückrichtung beweisen wir nur in dem Fall, in dem alle Eigenwerte der symmetrischen Matrix  $A \in \mathbb{R}^{n \times n}$  paarweise verschieden sind (der allgemeine Fall wird z. B. in [Mei15, Korollar 2.33] betrachtet). Ist dies der Fall, so wissen wir bereits, dass eine invertierbare Matrix  $S \in \mathbb{R}^{n \times n}$  existiert, so dass  $D = S^{-1}AS$  eine Diagonalmatrix ist. Seien  $v_1, \dots, v_n \in \mathbb{R}^n$  die Spaltenvektoren von  $S$ . Dies sind genau die Eigenvektoren der Eigenwerte  $\lambda_1, \dots, \lambda_n$  von  $A$ , also  $Av_i = \lambda_i v_i$ , wobei  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Wir können annehmen, dass  $\|v_i\| = 1$

für alle  $i = 1, \dots, n$  ist, ansonsten ersetzen wir  $v_i$  durch  $\frac{v_i}{\|v_i\|}$ . Wir zeigen nun, dass  $v_i^T v_j = 0$  ist für  $i \neq j$ . In der Tat gilt

$$\lambda_i v_i^T v_j = (\lambda_i v_i)^T v_j = (Av_i)^T v_j = v_i^T A^T v_j = v_i^T A v_j = \lambda_j v_i^T v_j.$$

Da nach Voraussetzung  $\lambda_i \neq \lambda_j$  ist, muss  $v_i^T v_j = 0$  sein. Da  $v_i^T v_i = \|v_i\|^2 = 1$  ist, folgt damit  $S^T S = I$ , also ist  $S$  orthogonal und es gilt  $S^{-1} = S^T$ .  $\square$

Wir kommen nun noch einmal zurück zu komplexen Matrizen  $A \in \mathbb{C}^{n \times n}$ . Ähnlich wie für Satz 1.2.23 lässt sich das nun folgende Resultat beweisen. Hier brauchen wir den Begriff einer *unitären* Matrix, siehe Definition B.2.27.

**Satz 1.2.24.** *Eine Matrix  $A \in \mathbb{C}^{n \times n}$  ist hermitesch genau wenn sie unitär diagonalisierbar ist, d.h. wenn eine unitäre Matrix  $U \in \mathbb{C}^{n \times n}$  existiert, so dass  $D = U^* A U$  eine Diagonalmatrix ist.*

Mit diesen Sätzen können wir eine alternative Darstellung für die Operatornorm  $\|A\|$  einer Matrix  $A \in \mathbb{C}^{n \times n}$  herleiten. Zuerst definieren wir, was wir unter dem Spektralradius einer Matrix verstehen wollen.

**Definition 1.2.25.** Sei  $A \in \mathbb{C}^{n \times n}$  eine Matrix und  $\sigma(A)$  das Spektrum von  $A$ . Dann heißt

$$\rho(A) := \max\{|\lambda| : \lambda \in \sigma(A)\}$$

*Spektralradius* von  $A$ .

**Satz 1.2.26.** *Für  $A \in \mathbb{C}^{m \times n}$  gilt*

$$\|A\| = \sqrt{\rho(A^* A)}.$$

*Beweis.* Für  $A \in \mathbb{C}^{m \times n}$  ist  $A^* A \in \mathbb{C}^{n \times n}$  eine hermitesche Matrix. Nach Satz 1.2.24 existiert dann eine unitäre Matrix  $U \in \mathbb{C}^{n \times n}$  so dass  $\text{diag}(\lambda_1, \dots, \lambda_n) = U^* A^* A U$  ist, wobei  $\lambda_1, \dots, \lambda_n$  die (reellen) Eigenwerte von  $A^* A$  sind. Sind  $u_1, \dots, u_n$  die Spaltenvektoren von  $U$ , so ist  $u_i$  ein Eigenvektor zum Eigenwert  $\lambda_i$ . Da  $U$  unitär ist, bilden die Vektoren  $u_1, \dots, u_n$  eine Orthonormalbasis, siehe Satz B.2.29. Jedes  $x \in \mathbb{C}^n$  besitzt damit eine Darstellung

$$x = \alpha_1 u_1 + \dots + \alpha_n u_n.$$

Damit folgt

$$A^* A x = \alpha_1 \lambda_1 u_1 + \dots + \alpha_n \lambda_n u_n.$$

Somit gilt

$$\begin{aligned} \|Ax\|^2 &= \langle Ax, Ax \rangle = \langle x, A^* A x \rangle = \left\langle \sum_{i=1}^n \alpha_i u_i, \sum_{i=1}^n \alpha_i \lambda_i u_i \right\rangle \\ &= \sum_{i=1}^n \lambda_i |\alpha_i|^2 \leq \rho(A^* A) \sum_{i=1}^n |\alpha_i|^2 = \rho(A^* A) \|x\|^2. \end{aligned}$$



Da  $x$  beliebig war, folgt damit die Abschätzung

$$\|A\| \leq \sqrt{\rho(A^*A)}.$$

Um die umgekehrte Abschätzung zu zeigen betrachten wir einen Eigenvektor  $u_j$  zum betragsmäßig größten Eigenwert  $\lambda_j$ . Wir können o.B.d.A. annehmen, dass  $\|u_j\| = 1$  ist, andernfalls ersetzen wir  $u_j$  durch  $\frac{u_j}{\|u_j\|}$ . Ersetzen wir in der obigen Ungleichungskette  $x$  durch  $u_j$ , so erhalten wir

$$0 \leq \|Au_j\|^2 = \lambda_j.$$

Somit gilt, da  $\|u_j\| = 1$  ist,

$$\|A\|^2 \geq \frac{\|Au_j\|^2}{\|u_j\|^2} = \|Au_j\|^2 = \lambda_j = \rho(A^*A).$$

□

Wir formulieren noch eine weitere interessante Relation zwischen Spektralradius und Matrixnorm.

**Lemma 1.2.27.** *Sei  $A \in \mathbb{C}^{n \times n}$ . Dann gelten die folgenden Aussagen:*

- (i)  $\rho(A) \leq \|A\|$ .
- (ii) *Ist  $A$  hermitesch, so ist  $\rho(A) = \|A\|$ .*

*Beweis.* (i) Sei  $\lambda$  der betragsmäßig größte Eigenwert von  $A$  zum Eigenvektor  $u \in \mathbb{C}^n$ . Dann gilt

$$\|A\| = \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|} \geq \frac{\|Au\|}{\|u\|} = |\lambda| = \rho(A).$$

- (ii) *Ist  $A$  hermitesch, so sind alle Eigenwerte  $\lambda_1, \dots, \lambda_n$  reell. Nach Satz 1.2.24 existiert eine unitäre Matrix  $U \in \mathbb{C}^{n \times n}$ , so dass*

$$\text{diag}(\lambda_1, \dots, \lambda_n) = U^*AU$$

ist. Es folgt

$$A^*A = U \text{diag}(\lambda_1^2, \dots, \lambda_n^2)U^*$$

und damit  $\rho(A^*A) = \rho(A)^2$ . Die Aussage folgt damit aus Satz 1.2.26.

□

## 1.3 Singulärwertzerlegung

Wir wollen nun eine ähnliche Zerlegung wie die der Diagonalisierbarkeit einer Matrix kennenlernen, die auch für nicht-quadratische Matrizen funktioniert. Dies wird genau die Singulärwertzerlegung sein. Wir werden uns hier auf den Fall reellwertiger Matrizen beschränken.

**Definition 1.3.1.** Sei  $A \in \mathbb{R}^{m \times n}$  eine beliebige Matrix. Gibt es orthogonale Matrizen  $U \in \mathbb{R}^{m \times m}$  und  $V \in \mathbb{R}^{n \times n}$  sowie eine Matrix  $\Sigma \in \mathbb{R}^{m \times n}$  mit der Eigenschaft, dass für  $\Sigma = (\sigma_{ij})$  gilt, dass  $\sigma_{ij} = 0$  für  $i \neq j$  sowie  $\sigma_{ii} \geq 0$ , und gilt für diese Matrizen die Identität

$$A = U\Sigma V^T,$$

so nennen wir dies eine *Singulärwertzerlegung* von  $A$ . Die Diagonaleinträge  $\sigma_{ii}$  von  $\Sigma$  heißen *Singulärwerte*. Die Spaltenvektoren  $u_1, \dots, u_m$  von  $U$  heißen *links-Singulärvektoren*, die Spaltenvektoren  $v_1, \dots, v_n$  von  $V$  nennt man *rechts-Singulärvektoren*.

*Bemerkung 1.3.2.* Die Matrix  $\Sigma$  ist von der Form

$$\Sigma = \begin{pmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & \dots & \sigma_{nn} \\ & & & \\ & & & 0 \\ \vdots & \ddots & \vdots & \\ 0 & \dots & 0 & \end{pmatrix}$$

für  $m > n$  und

$$\Sigma = \begin{pmatrix} \sigma_{11} & & 0 & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ 0 & & \sigma_{mm} & 0 & \dots & 0 \end{pmatrix}$$

für  $m < n$ .

Bevor wir uns weiter mit der Singulärwertzerlegung beschäftigen, betrachten wir eine Eigenschaft symmetrischer reellwertiger Matrizen:

**Definition 1.3.3.** Eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  heißt *positiv definit* wenn  $v^T A v > 0$  ist für alle  $v \in \mathbb{R}^n \setminus \{0\}$ . Die Matrix  $A$  heißt *positiv semi-definit* wenn  $v^T A v \geq 0$  ist für alle  $v \in \mathbb{R}^n$ .

Wir wissen bereits, dass symmetrische Matrizen reelle Eigenwerte haben. Im nächsten Satz sehen wir nun, dass die Definitheit einer solchen Matrix charakterisiert ist durch die Vorzeichen der Eigenwerte.

**Satz 1.3.4.** *Eine symmetrische Matrix  $A$  ist positiv definit genau dann wenn alle Eigenwerte reell und echt größer als 0 sind.  $A$  ist positiv semi-definit genau dann wenn alle Eigenwerte größer oder gleich 0 sind.*

*Beweis.* Da  $A$  symmetrisch ist besitzt  $A$  nur reelle Eigenwerte  $\lambda_1, \dots, \lambda_n$  und es existiert eine orthogonale Matrix  $S$  so dass  $A = SDS^T$  ist, wobei  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$  ist (siehe Satz 1.2.23). Sind alle  $\lambda_i$  echt größer als 0, so folgt für jeden Vektor  $v \in \mathbb{R}^n \setminus \{0\}$

$$v^T A v = v^T SDS^T v = (S^T v)^T D S^T v = w_1^2 \lambda_1 + \dots + w_n^2 \lambda_n > 0,$$

wobei  $w_i$  die  $i$ -te Komponente des Vektors  $S^T v$  bezeichne. Also ist  $A$  positiv definit. Die andere Richtung und der semi-definite Fall folgen ähnlich.  $\square$



Hier haben wir benutzt, dass

$$\|Av_i\|^2 = \langle Av_i, Av_i \rangle = \langle v_i, A^T Av_i \rangle = \langle v_i, \lambda_i v_i \rangle = \lambda_i \langle v_i, v_i \rangle = \lambda_i$$

ist wobei die letzte Gleichheit folgt, da  $V$  orthogonal ist. Nach Definition ist dann  $\langle u_i, u_i \rangle = 1$  und

$$\langle u_i, u_j \rangle = \frac{\langle Av_i, Av_j \rangle}{\|Av_i\| \|Av_j\|} = \frac{\langle v_i, A^T Av_j \rangle}{\|Av_i\| \|Av_j\|} = \frac{\lambda_j \langle v_i, v_j \rangle}{\|Av_i\| \|Av_j\|} = 0$$

für  $i \neq j$ . Insbesondere sind  $u_1, \dots, u_k$  linear unabhängig. Mit Hilfe des Gram-Schmidtschen Orthogonalisierungsverfahrens (siehe Definition B.2.25) können wir Vektoren  $u_{k+1}, \dots, u_m$  finden, so dass  $u_1, \dots, u_m$  eine Orthonormalbasis ist. Dann definieren wir  $U := (u_1, \dots, u_m) \in \mathbb{R}^{m \times m}$ , d.h.  $U$  wird gebildet aus den Spaltenvektoren  $u_1, \dots, u_m$ . Damit ist  $U$  eine Orthogonalmatrix. Weiter definieren wir  $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{m \times n}$  durch

$$\sigma_{ij} := \begin{cases} \sqrt{\lambda_i} & \text{für } i = j \text{ und } i \leq k, \\ 0 & \text{sonst.} \end{cases} \quad (1.1)$$

Nach Konstruktion ist dann  $AV = U\Sigma$ . Da  $V$  orthogonal ist, folgt  $A = U\Sigma V^T$  und die Existenz ist gezeigt.  $\square$

*Bemerkung 1.3.7.* Man beachte, dass der obige Beweis auch eine Möglichkeit zeigt, die Matrizen  $U, V$  und  $\Sigma$  zu konstruieren. Konkret kann man dies wie folgt tun:

1. Man bestimme die Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  von  $A^T A$  und berechne eine Orthonormalbasis  $v_1, \dots, v_n$  aus Eigenvektoren. Setze  $V := (v_1, \dots, v_n)$ .
2. Man definiere  $u_1, \dots, u_k$  durch  $u_i = Av_i / \|Av_i\|$  und nutze das Gram-Schmidtsche Orthogonalisierungsverfahren, um eine Orthonormalbasis  $u_1, \dots, u_m$  zu konstruieren. Setze  $U := (u_1, \dots, u_m)$ .
3. Definiere  $\Sigma$  wie in (1.1).

Der größte Singulärwert hat eine interessante Eigenschaft.

**Korollar 1.3.8.** Sei  $A \in \mathbb{R}^{m \times n}$  und  $\sigma$  der größte Singulärwert aus der Singulärwertzerlegung von  $A$ . Dann gilt

$$\|A\| = \sigma.$$

*Beweis.* Aus Satz 1.2.26 folgt

$$\|A\| = \sqrt{\rho(A^T A)}.$$

Nach Theorem 1.3.6 gilt aber auch  $\sqrt{\rho(A^T A)} = \sigma$ .  $\square$

Es stellt sich heraus, dass auch die Frobeniusnorm einer Matrix mit Hilfe der Singulärwerte dargestellt werden kann. Zunächst erinnern wir an die Definition der Spur einer Matrix.

**Definition 1.3.9.** Sei  $A \in \mathbb{R}^{n \times n}$ . Dann ist die *Spur von  $A$*  definiert als die Summe der Diagonaleinträge, geschrieben als

$$\operatorname{tr}(A) := \sum_{i=1}^n a_{ii}.$$

Für die Spur gilt die folgende nützliche Gleichheit.

**Lemma 1.3.10.** Seien  $A, B \in \mathbb{R}^{n \times n}$ . Dann gilt

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

*Beweis.* Man rechnet nach, dass sowohl  $\operatorname{tr}(AB)$  als auch  $\operatorname{tr}(BA)$  gegeben ist durch

$$\sum_{i,j=1}^n a_{ij}b_{ji}.$$

□

Die Frobeniusnorm einer Matrix lässt sich auch mit Hilfe der Spur schreiben:

**Lemma 1.3.11.** Für jede Matrix  $A \in \mathbb{R}^{m \times n}$  gilt

$$\|A\|_F = \sqrt{\operatorname{tr}(A^T A)} = \sqrt{\operatorname{tr}(A A^T)}.$$

*Beweis.* Folgt direkt aus der Definition der Frobeniusnorm und der Definition des Matrixprodukts. □

Damit können wir den folgenden Satz beweisen:

**Satz 1.3.12.** Sei  $A \in \mathbb{R}^{m \times n}$ ,  $r = m \wedge n$  und seien  $\sigma_{11}, \dots, \sigma_{rr}$  die Singulärwerte von  $A$ . Dann gilt

$$\|A\|_F = \sqrt{\sigma_{11}^2 + \dots + \sigma_{rr}^2}.$$

*Beweis.* Sei  $A = U \Sigma V^T$  eine Singulärwertzerlegung von  $A$ . Mit Hilfe von Lemma 1.3.10 und Lemma 1.3.11 folgt dann

$$\begin{aligned} \|A\|_F &= \sqrt{\operatorname{tr}((U \Sigma V^T)^T U \Sigma V^T)} = \sqrt{\operatorname{tr}(V \Sigma^T \Sigma V^T)} = \sqrt{\operatorname{tr}(\Sigma^T \Sigma V^T V)} \\ &= \sqrt{\operatorname{tr}(\Sigma^T \Sigma)} = \sqrt{\sigma_{11}^2 + \dots + \sigma_{rr}^2}. \end{aligned}$$

□

### 1.3.1 Reduzierte Singulärwertzerlegung

Wir betrachten nun eine alternative Form der Singulärwertzerlegung. Zuerst ein Lemma.

**Lemma 1.3.13.** Sei  $A \in \mathbb{R}^{m \times n}$ . Dann gilt  $\operatorname{rank}(A) = \operatorname{rank}(A^T A) = \operatorname{rank}(A A^T)$ .

*Beweis.* Ist  $x \in \text{Ker}(A)$ , so ist  $A^T Ax = A^T 0 = 0$ , also  $x \in \text{Ker}(A^T A)$ . Umgekehrt gilt für  $x \in \text{Ker}(A^T A)$ , dass  $A^T Ax = 0$  ist, also auch

$$0 = x^T A^T Ax = (Ax)^T Ax = \langle Ax, Ax \rangle = \|Ax\|^2,$$

woraus  $Ax = 0$  und somit  $x \in \text{Ker}(A)$  folgt. Es gilt also  $\text{Ker}(A) = \text{Ker}(A^T A)$ . Nach dem Rangsatz (Satz B.1.13) ist

$$n = \dim(\text{Ker}(A)) + \text{rank}(A) = \dim(\text{Ker}(A^T A)) + \text{rank}(A^T A),$$

woraus  $\text{rank}(A) = \text{rank}(A^T A)$  folgt. Angewendet auf  $A^T$  erhalten wir auch  $\text{rank}(A^T) = \text{rank}(AA^T)$ . Da nach Satz B.1.9 aber  $\text{rank}(A^T) = \text{rank}(A)$  ist, folgt daraus die Aussage. □

Wir zitieren noch ein Resultat, dass wir allerdings nicht beweisen werden.

**Satz 1.3.14.** *Ist  $A \in \mathbb{R}^{m \times n}$  und  $r = \text{rank}(A)$ , so besitzen  $A^T A$  (und  $AA^T$ ) genau  $r$  Eigenwerte echt größer als 0.*

**Korollar 1.3.15.** *Ist  $A \in \mathbb{R}^{m \times n}$  und  $r = \text{rank}(A)$ , so besitzt  $A$  genau  $r$  Singulärwerte echt größer als 0.*

Sei  $A \in \mathbb{R}^{m \times n}$  und  $r := \text{rank}(A) \leq m \wedge n$ . Sei  $A = U\Sigma V^T$  eine Singulärwertzerlegung von  $A$  der Form, bei der die Singulärwerte in absteigender Form angeordnet sind, also  $\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{rr} > 0$ . Die  $(m-r)$ -letzten Zeilen und die  $(n-r)$ -letzten Spalten der Matrix  $\Sigma$  sind also Nullzeilen bzw. Nullspalten. Definieren wir  $\tilde{U} \in \mathbb{R}^{m \times r}$ , indem wir die letzten  $m-r$  Spalten von  $U$  streichen und  $\tilde{V} \in \mathbb{R}^{n \times r}$ , indem wir die letzten  $n-r$  Spalten von  $V$  streichen, so erhalten wir die Zerlegung

$$A = \tilde{U} \tilde{\Sigma} \tilde{V}^T,$$

wobei  $\tilde{\Sigma} = \text{diag}(\sigma_{11}, \dots, \sigma_{rr})$ . Man beachte, dass  $\tilde{U}$  und  $\tilde{V}$  weiterhin orthogonale Spaltenvektoren besitzen. Diese Zerlegung nennen wir reduzierte Singulärwertzerlegung:

**Definition 1.3.16.** Sei  $A \in \mathbb{R}^{m \times n}$  und  $r = \text{rank}(A)$ . Existieren Matrizen  $U \in \mathbb{R}^{m \times r}$  und  $V \in \mathbb{R}^{n \times r}$  mit orthogonalen Spalten und eine Diagonalmatrix  $\Sigma \in \mathbb{R}^{r \times r}$ , so dass  $A = U\Sigma V^T$  gilt, so nennen wir dies eine *reduzierte Singulärwertzerlegung*.

Die obige Überlegung zeigt, dass die Existenz der Singulärwertzerlegung auch die Existenz der reduzierten Singulärwertzerlegung impliziert (und umgekehrt).

### 1.3.2 Anwendung: Matrixapproximation

Die für uns wichtigste Anwendung der Singulärwertzerlegung betrifft die Approximation einer gegebenen Matrix  $A \in \mathbb{R}^{m \times n}$ , die wir nun beschreiben wollen. Wir nehmen dazu eine Singulärwertzerlegung

$$A = U\Sigma V^T \tag{1.2}$$

an. Zudem nehmen wir an, dass die Singulärwerte in  $\Sigma$  absteigend geordnet sind, also  $\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{rr} \geq 0$ . Setze  $\sigma_i := \sigma_{ii}$ . Wie üblich bezeichnen  $u_1, \dots, u_m$  die links-Singulärvektoren und  $v_1, \dots, v_n$  die rechts-Singulärvektoren. Wir können dann (1.2) umschreiben zu

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T = \sum_{i=1}^r \sigma_i A_i,$$

wobei  $A_i := u_i v_i^T \in \mathbb{R}^{m \times n}$ . Jede Matrix  $A_i$  hat Rang 1, da die Spaltenvektoren von  $A_i$  Vielfache des Vektors  $v_i$  sind. Die abgeschnittene Summe

$$A(k) := \sum_{i=1}^k \sigma_i A_i \in \mathbb{R}^{m \times n} \quad (1.3)$$

hat damit Rang  $k$ . Wir haben also durch  $A(k)$  eine Approximation von  $A$  konstruiert, die einen geringeren Rang als  $A$  selber besitzt. Wir werden nun sehen, dass diese Approximation in gewissem Sinne optimal ist. Genauer zeigt das folgende Theorem, dass die Approximation  $A(k)$  aus (1.3) optimal unter allen Rang- $k$  Approximationen ist. Dies gilt sowohl dann, wenn man den Abstand der Matrizen in der üblichen Operatornorm misst als auch dann, wenn man die Frobeniusnorm zugrunde legt.

**Theorem 1.3.17** (Eckart-Young Theorem). *Für jede Matrix  $A \in \mathbb{R}^{m \times n}$  gilt*

$$A(k) = \arg \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B)=k}} \|A - B\| = \arg \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B)=k}} \|A - B\|_F.$$

*Mit anderen Worten: Für jede Matrix  $B$  mit  $\text{rank}(B) = k$  haben wir*

$$\|A - B\| \geq \|A - A(k)\| \quad \text{und} \quad \|A - B\|_F \geq \|A - A(k)\|_F$$

*Beweis.* Siehe [BHK20, Theorem 3.6 und Theorem 3.9] □

Mit dem Eckart-Young Theorem haben wir nun eine mathematische Beschreibung davon, dass die Bildkompression aus der Motivation in Abschnitt 1.1 tatsächlich funktioniert. Insbesondere können wir auch den Approximationsfehler mithilfe der Singulärwerte angeben. Es gilt nämlich

$$A - A(k) = \sum_{i=k+1}^r \sigma_i A_i$$

und aus Korollar 1.3.8 folgt damit

$$\|A - A(k)\| = \sigma_{k+1}.$$

Weiter folgt aus Satz 1.3.12, dass

$$\|A - A(k)\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$$

ist. Zur Erinnerung: Abbildung 1.5 zeigt die Singulärwerte  $\sigma_1 \geq \sigma_2 \geq \dots$  des Fotos aus Abbildung 1.3. Wir sehen hier, dass die Singulärwerte vor allem am Anfang stark abfallen (etwa bis  $k = 75$ ). Intuitiv ist es also nicht erstaunlich, dass das komprimierte Bild für  $k = 100$  dem Original bereits recht ähnlich sieht.

## 1.4 Beispielaufgabe: Singulärwertzerlegung und Gram-Schmidt

Wir nehmen an, wir bekommen die Aufgabe gestellt, die Singulärwertzerlegung  $A = U\Sigma V^T$  der Matrix

$$A = \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix}$$

zu berechnen. Wir betrachten in diesem Abschnitt, wie wir im Allgemeinen systematisch an eine solche „Übungsaufgabe“ herangehen können und wie wir diese konkrete Aufgabe lösen können. Wir gehen in diesem Abschnitt sehr detailliert vor und wollen keine Kleinigkeiten auslassen. Wundern Sie sich daher nicht, falls Sie schneller zu einer Lösung kommen oder das Beispiel sehr lang wirkt. Überspringen Sie gerne die Teile, die Sie schon kennen.

**Schritt 0:** Wir suchen im Skript nach relevanten Stichworten und Informationen um die Aufgabenstellung zu verstehen und eine Idee für die Lösung zu bekommen.

In unserem Fall finden wir dabei zunächst die Definition 1.3.1 der Singulärwertzerlegung. Diese besagt, dass die Matrix  $\Sigma$  die *Singulärwerte* auf der Hauptdiagonalen enthält.  $U$  und  $V$  sollen *orthogonale* Matrizen sein. Da  $A$  eine  $3 \times 2$ -Matrix ist, muss  $V$  eine  $2 \times 2$ -Matrix,  $U$  eine  $3 \times 3$ -Matrix und  $\Sigma$  eine  $3 \times 2$ -Matrix sein. Die Definition sagt uns aber noch nicht, wie wir diese berechnen können. Dazu finden wir Bemerkung 1.3.7. Aus dieser schließen wir, dass wir zunächst *Eigenwerte*  $\lambda_1 \geq \dots \geq \lambda_n$  von  $A^T A$  bestimmen (diese werden laut Theorem 1.3.6 die Singulärwerte) und eine zugehörige *Orthonormalbasis* aus *Eigenvektoren* berechnen sollen. Mithilfe dieser Vektoren können wir  $V$  und  $U$  bestimmen und benötigen gegebenenfalls noch das *Gram-Schmidtsche Orthogonalisierungsverfahren*.

Wenn wir uns an die Bedeutung der kursiv gedruckten Begriffe erinnern, haben wir nun einen guten Überblick, was zu tun ist. Andernfalls nehmen wir die unbekanntesten Begriffe als neue Stichworte, nach denen wir zunächst wieder im Skript suchen. Auf der Basis der gefundenen Informationen machen wir uns einen Plan, was genau gefragt ist und welche Schritte dafür nötig sind. Bemerkung 1.3.7 hilft uns im aktuellen Fall sehr. Für manche theoretischen oder praktischen Aufgaben ist der Plan weniger offensichtlich vorgegeben. Dann können wir uns aus verschiedenen Lemmata und Theoremen eine (Beweis-)„Kette“ zusammensetzen, die uns dann vorgibt, was wir in welcher Reihenfolge zeigen müssen.

Nachdem wir uns im obigen Abschnitt einen Plan davon gemacht haben, wie wir vorgehen können, führen wir die Berechnung der Singulärwertzerlegung jetzt Schritt für Schritt durch.

**Schritt 1:** Wir bestimmen die Matrix  $V$ , bzw.  $V^T$ .



1.4. BEISPIELAUFGABE: SINGULÄRWERTZERLEGUNG UND GRAM-SCHMIDT31

i) Wir berechnen zuerst  $A^T A$ :

$$\begin{aligned} A^T A &= \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix}^T \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -4 & \frac{21}{5} & 0 \\ -3 & -\frac{28}{5} & 0 \end{pmatrix} \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{841}{25} & -\frac{288}{25} \\ -\frac{288}{25} & \frac{1009}{25} \end{pmatrix} \end{aligned}$$

wobei der markierte Eintrag als Skalarprodukt der markierten Zeile und der markierten Spalte, d.h.  $(-4) \cdot (-3) + \frac{21}{5} \cdot (-\frac{28}{5}) + 0 \cdot 0 = \frac{841}{25}$ , berechnet wird.

ii) Nun berechnen wir die Eigenwerte von  $A^T A$ . Dafür stellen wir das charakteristische Polynom  $p_{A^T A}$  auf und berechnen seine Nullstellen (siehe Definition 1.2.6 und Satz 1.2.8).

$$\begin{aligned} p_{A^T A}(z) &= \det(A^T A - zI) = \det \begin{pmatrix} \frac{841}{25} - z & -\frac{288}{25} \\ -\frac{288}{25} & \frac{1009}{25} - z \end{pmatrix} \\ &= \left( \frac{841}{25} - z \right) \left( \frac{1009}{25} - z \right) - \left( -\frac{288}{25} \right)^2 \\ &= z^2 - 74z + 1225. \end{aligned}$$

Die Nullstellen  $\lambda$  sind die Lösungen der quadratischen Gleichung  $p_{A^T A}(\lambda) = \lambda^2 - 74\lambda + 1225 = 0$ . Mit der  $p$ - $q$ -Formel finden wir  $\lambda_1 = 49$  und  $\lambda_2 = 25$ . Da die Singulärwerte  $\sigma_i = \sqrt{\lambda_i}$  nach der Größe absteigend geordnet sein sollen, sortieren wir die Eigenwerte ebenfalls schon hier.

iii) Wir bestimmen eine Orthonormalbasis aus Eigenvektoren. Dafür machen wir zunächst folgende Beobachtungen. Erstens,  $A^T A$  ist symmetrisch, d.h. alle Eigenräume sind orthogonal zueinander. Zweitens, die Eigenräume hier haben höchstens (und mindestens) die Dimension 1. Das bedeutet, wir berechnen hier also jeweils einen Eigenvektor und normieren diesen dann. Das Ergebnis ist dann bereits eine Orthonormalbasis. Falls (in einem anderen Beispiel) ein Eigenraum eine höhere Dimension hätte, könnten wir das Gram-Schmidt-Verfahren anwenden. Wir werden diesem Fall weiter unten begegnen.

Nach Definition 1.2.1 ist  $\lambda_i$  ein Eigenwert zum Eigenvektor  $v_i$ , falls  $A^T A v_i = \lambda_i v_i$  gilt. Wir müssen also das Gleichungssystem  $(A^T A - \lambda_i I)v_i = 0$  lösen. Wir betrachten zunächst  $\lambda_1 = 49$ .

$$\begin{aligned} \begin{pmatrix} \frac{841}{25} - 49 & -\frac{288}{25} \\ -\frac{288}{25} & \frac{1009}{25} - 49 \end{pmatrix} \begin{pmatrix} v_1^1 \\ v_1^2 \end{pmatrix} &= \begin{pmatrix} -\frac{384}{25} & -\frac{288}{25} \\ -\frac{288}{25} & -\frac{216}{25} \end{pmatrix} \begin{pmatrix} v_1^1 \\ v_1^2 \end{pmatrix} \\ &= \begin{pmatrix} -\frac{384}{25} v_1^1 - \frac{288}{25} v_1^2 \\ -\frac{288}{25} v_1^1 - \frac{216}{25} v_1^2 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Aus der ersten Zeile des Gleichungssystems schließen wir, dass  $v_1^2 = -\frac{4}{3}v_1^1$  gelten muss. Setzen wir dies in die zweite Zeile des Gleichungssystems ein, erhalten wir

$$-\frac{288}{25}v_1^1 + \frac{216}{25} \frac{4}{3}v_1^1 = 0.$$

Also haben Eigenvektoren zu  $\lambda_1$  die allgemeine Form  $(v_1^1, -\frac{4}{3}v_1^1)^T$ ,  $v_1^1 \in \mathbb{R}$ . Insbesondere wird der Eigenraum also nur durch einen einzelnen Basisvektor beschrieben und hat Dimension 1 (wie wir schon zuvor wussten). Wir wählen nun zum Beispiel  $v_1^1 = 1$  und setzen  $v_1 = (1, -\frac{4}{3})$ . Für  $\lambda_2 = 25$  gehen wir genauso vor. Die Eigenvektoren haben die Form  $(v_2^1, \frac{3}{4}v_2^1)^T$ ,  $v_2^1 \in \mathbb{R}$ . Wir wählen  $v_2 = (1, \frac{3}{4})^T$ . Wir wissen schon, dass die  $v_1, v_2$  orthogonal sind, können dies aber auch nachrechnen:

$$\langle v_1, v_2 \rangle = \left\langle \begin{pmatrix} 1 \\ -\frac{4}{3} \end{pmatrix}, \begin{pmatrix} 1 \\ \frac{3}{4} \end{pmatrix} \right\rangle = 1 \cdot 1 - \frac{4}{3} \cdot \frac{3}{4} = 1 - 1 = 0.$$

Wir normalisieren die Vektoren noch:

$$w_1 = \frac{v_1}{\|v_1\|} = \frac{\begin{pmatrix} 1 \\ -\frac{4}{3} \end{pmatrix}}{\sqrt{1^2 + (-4/3)^2}} = \frac{3}{5} \begin{pmatrix} 1 \\ -\frac{4}{3} \end{pmatrix} = \begin{pmatrix} \frac{3}{5} \\ -\frac{4}{5} \end{pmatrix},$$

$$w_2 = \frac{v_2}{\|v_2\|} = \frac{\begin{pmatrix} 1 \\ \frac{3}{4} \end{pmatrix}}{\sqrt{1^2 + (3/4)^2}} = \frac{4}{5} \begin{pmatrix} 1 \\ \frac{3}{4} \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix},$$

und  $\{w_1, w_2\}$  ist eine Orthonormalbasis des  $\mathbb{R}^2$  aus Eigenvektoren. Dies ergibt die Matrix

$$V = \begin{pmatrix} | & | \\ w_1 & w_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{pmatrix}.$$

**Schritt 2:** Wir berechnen die Matrix  $U$ .

i) Wir setzen zunächst  $u_i = \frac{Aw_i}{\|Aw_i\|}$ , also

$$Aw_1 = \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{5} \\ -\frac{4}{5} \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \\ 0 \end{pmatrix} \Rightarrow u_1 = \frac{Aw_1}{\|Aw_1\|} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$Aw_2 = \begin{pmatrix} -5 \\ 0 \\ 0 \end{pmatrix} \Rightarrow u_2 = \frac{Aw_2}{\|Aw_2\|} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

Die Vektoren  $u_i$  werden die ersten beiden Spalten der Matrix  $U$ . Wir wollen  $U$  zu einer  $3 \times 3$ -Matrix ergänzen. Dies tun wir, indem wir  $\{u_1, u_2\}$  zu einer ONB des  $\mathbb{R}^3$  ergänzen. Man kann in dieser einfachen Situation möglicherweise einen Vektor raten, der die Basis ergänzt. Im Allgemeinen ist das nicht möglich. Wir wenden daher nun das Gram-Schmidt-Verfahren an, um den Vektor zu finden.

ii) Das Gram-Schmidt-Verfahren macht aus einer Basis des  $\mathbb{R}^3$  eine Orthonormalbasis  $\{q_1, q_2, q_3\}$  des  $\mathbb{R}^3$ . Wir starten also, indem wir zu  $\{u_1, u_2\}$  einen beliebigen, linear unabhängigen Vektor hinzufügen, um eine Basis  $\{b_1 = u_1, b_2 = u_2, b_3\}$  des  $\mathbb{R}^3$  zu erhalten. Wir setzen  $b_3 = (2, 1, 3)^T$ . Nun gehen wir die Schritte 1.-5. des Verfahrens aus dem Anhang durch, Definition B.2.25. Wir setzen  $q_1 = b_1$ , da

1.4. BEISPIELAUFGABE: SINGULÄRWERTZERLEGUNG UND GRAM-SCHMIDT33

$b_1$  bereits die Länge 1 hat (Schritt 1 aus Definition B.2.25). Nun berechnen wir das Lot von  $b_2$  auf die Gerade durch  $q_1$  (in der Notation des Anhangs)

$$l_2 = b_2 - \langle b_2, q_1 \rangle q_1 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} - \underbrace{\left\langle \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle}_{=0} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

Das war Schritt 2 aus Definition B.2.25. Da das Lot bereits Länge 1 hat, haben wir  $q_2 = l_2 = b_2$  (Schritt 3). Nun berechnet man  $l_k$  und  $q_k$  auf ähnliche Weise für  $k = 3, \dots, n$  (Schritte 4 und 5). Bei uns gilt  $n = 3$ , also betrachten wir nur noch einen Vektor:

$$\begin{aligned} l_3 &= b_3 - \langle b_3, q_1 \rangle q_1 - \langle b_3, q_2 \rangle q_2 \\ &= \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} - \left\langle \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \left\langle \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \cdot \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} - 1 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}. \end{aligned}$$

Insbesondere wird hier das Lot auf die Ebene gefällt, die von  $q_1$  und  $q_2$  aufgespannt wird. Wir normieren den Vektor, den wir hier berechnet haben:

$$q_3 = \frac{l_3}{\|l_3\|} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

iii) Nun definieren wir die Matrix  $U$ . Wir haben

$$U = \begin{pmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

**Schritt 3:** Wir berechnen die Matrix  $\Sigma$ . Es gilt

$$\sigma_{ij} = \begin{cases} \sqrt{\lambda_i}, & \text{falls } i = j, \\ 0 & \text{sonst.} \end{cases}$$

also

$$\Sigma = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \sqrt{49} & 0 \\ 0 & \sqrt{25} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & 5 \\ 0 & 0 \end{pmatrix}.$$

Damit sind wir fertig.

Insgesamt wird die Singulärwertzerlegung

$$A = \begin{pmatrix} -4 & -3 \\ \frac{21}{5} & -\frac{28}{5} \\ 0 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_U \underbrace{\begin{pmatrix} 7 & 0 \\ 0 & 5 \\ 0 & 0 \end{pmatrix}}_\Sigma \underbrace{\begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}}_{V^T}.$$