Karlsruher Institut für Technologie

Einige Grundlagen aus der Linearen Algebra und der Analysis

Begleitmaterial zur Numerischen Mathematik
von

Prof. Dr. W. Dörfler

Institut für Angewandte und
Numerische Mathematik
an der Fakultät für Mathematik

Version: 3. Oktober 2021

Inhaltsverzeichnis

Der folgende Text soll Konzepte der linearen Algebra und der Analysis als Vorbereitung auf die Vorlesung *Numerische Mathematik* wiederholen. Im Detail sei auf die beiden Vorlesungen des ersten Studienjahres verwiesen.

1 Lineare Algebra

1.1 Vektorräume

Die abstrakten Voraussetzungen sind: Gegeben ist eine (nichtleere) Menge V die eine Gruppe bezüglich der „Addition +“ sei: zu v und w aus V sei ein eindeutiges Element $v+w \in V$ definiert. Außerdem gibt es ein Element $0 \in V$ und zu jedem Element das additiv inverse Element. Die Elemente der Menge V heißen *Vektoren*.

Es sei $\mathbb{K}$ ein Körper: Man nennt $\mathbb{K}$ auch den *Skalarkörper*, Elemente aus $\mathbb{K}$ auch einfach *Skalare*. Zu jedem $\lambda \in \mathbb{K}$ und Vektor $v \in V$ sei ein Element $\lambda v \in V$ erklärt. V, zusammen mit dem Skalarkörper und den genannten Verknüpfungen, heißt *Vektorraum* (*über* $\mathbb{K}$).

Im Folgenden wählen wir den Körper der reellen Zahlen $\mathbb{R}$. Später betrachten wir aber auch den Körper der komplexen Zahlen $\mathbb{C}$.

Eine besondere Bedeutung haben *Linearkombinationen*

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$$

zu $m \in \mathbb{N}$, $\lambda_i \in \mathbb{K}$ und $v_i \in V$ für $i = 1, \ldots, m$. Gibt es eine Menge $\{b_1, \ldots, b_n\} \subset V$ derart, dass *jedes* $v \in V$ eine *eindeutige* Darstellung als Linearkombination

$$v = \nu_1 b_1 + \cdots + \nu_n b_n \tag{1.1}$$

mit $\nu_i \in \mathbb{K}$ hat, so heißt diese Menge eine *Basis* von V. Die Zahl n ist eindeutig, indem jede andere Basis von V auch n Elemente hat, und heißt *Dimension* von V. Man sagt, V ist ein *n-dimensionaler* (insbesondere *endlich-dimensionaler*) *Vektorraum*. Für eine gegebene Basis schreiben wir v mithilfe der Zahlen $\nu_1, \ldots, \nu_n$ der eindeutigen Basisdarstellung (1.1) wie folgt als *Spaltenvektor*

$$v = [\nu_i]_{i=1,\ldots,n} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix}.$$

ν_i heißt *i-te Komponente von v*. Die Addition und die skalare Multiplikation sind in dieser Darstellung dann komponentenweise durchzuführen ($v + w := [\nu_i + \omega_i]_{i=1,n}$ und $\lambda v := [\lambda \nu_i]_{i=1,\ldots,n}$ für $v = [\nu_i]_{i=1,\ldots,n}$ und $w = [\omega_i]_{i=1,\ldots,n}$).

Beispiel 1 ($\mathbb{R}^n$)**.** *Das wichtigste Beispiel ist der Vektorraum aller geordneten n-Tupel reeller Zahlen, die wir in der Form*

$$v = [\nu_i]_{i=1,\ldots,n} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix}$$

schreiben und mit $\mathbb{R}^n$ bezeichnen. Die übliche Basis (d.h. soweit nicht anders vereinbart) ist die sogenannte euklidische Basis $e_1 := [1, 0, \ldots, 0]$, $\ldots$, $e_n := [0, \ldots, 0, 1]$*. Die Rechenoperationen sind komponentenweise durchzuführen. Die Dimension von $\mathbb{R}^n$ ist n. Im Fall $\mathbb{K} = \mathbb{C}$ ist die Bezeichnung entsprechend $\mathbb{C}^n$. Wie oben gesehen, kann man jeden endlich-dimensionalen Vektorraum über $\mathbb{K}$ mit dem Vektorraum $\mathbb{K}^n$ identifizieren.*

Beispiel 2 ($\mathbb{P}_n$)**.** *Wir betrachten nun die Menge der Polynome in einer reellen Variablen vom Grad kleiner gleich $n \in \mathbb{N}$,*

$$\Big\{p : \mathbb{R} \to \mathbb{R} \,\Big|\, p(x) = \sum_{i=0}^{n} a_i x^i \text{ für } a_i \in \mathbb{R},\ i = 0, \dots, n\Big\}.$$

Mit der punktweisen Addition von Funktionen und der punktweisen Multiplikation mit einer reellen Zahl wird diese Menge zu einem Vektorraum. Wir bezeichnen diesen mit $\mathbb{P}_n$. Die Dimension von $\mathbb{P}_n$ ist $n+1$, denn da sich das Null-Polynom nur für $a_0 = \cdots = a_n = 0$ ergibt, lässt sich kein Monom x^i durch die anderen ausdrücken. Die hier verwandte Basis $\{b_i \in \mathbb{P}_n \,|\, b_i(x) = x^i,\ i = 0, \dots, n\}$ heißt Monom-Basis.

Beispiel 3 ($\mathcal{C}^0(I)$)**.** *Es sei $I \subset \mathbb{R}$ ein abgeschlossenes Intervall und V die Menge der stetigen Funktionen $I \to \mathbb{R}$. Mit der punktweisen Addition von Funktionen und der punktweisen Multiplikation mit einer reellen Zahl wird V zu einem Vektorraum. Wir bezeichnen diesen mit $\mathcal{C}^0(I)$. Da es keine endliche Basis gibt, ist dieser Vektorraum* nicht *endlich-dimensional. Unendlich dimensionale Funktionenräume sind Thema der* Funktionalanalysis.

1.2 Lineare Abbildungen

Eine Abbildung L zwischen 2 Vektorräumen V, W, $L : V \to W$, heißt *lineare Abbildung*, falls gilt

$$L(v + w) = L(v) + L(w),\ L(\lambda v) = \lambda L(v) \qquad \text{für alle } v, w \in V,\ \lambda \in \mathbb{K}.$$

Im Falle linearer Abbildungen lässt man die Argumentklammer oft weg, d.h. man schreibt $L(v) = Lv$.

Ist eine Basis $\{b_1, \dots, b_n\}$ von V gegeben und v wie in (1.1), so gilt für lineare Abbildungen

$$Lv = L\Big(\sum_{i=1}^{n} \nu_i b_i\Big) = \sum_{i=1}^{n} \nu_i L b_i.$$

Damit ist die Abbildung L definiert, wenn man $Lb_1, \dots, Lb_n$ kennt. Ist $\{d_1, \dots, d_m\}$ eine Basis von W und hat Lb_i die Darstellung

$$Lb_i = \sum_{j=1}^{m} a_{ij} d_j$$

mit Zahlen $a_{ij} \in \mathbb{K}$, so folgt

$$Lv = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \nu_i d_j.$$

Formulieren wir diese Identität bezüglich der vereinbarten Basen, so erhalten wir

$$v = [\nu_i]_i \in V \ \mapsto\ w = Lv = \Big[\sum_{i=1}^{n} a_{ij} \nu_i\Big]_j = \nu_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + \nu_m \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Bei gegebenen Basen ist L eindeutig durch die Zahlen a_{ij} bestimmt. Mit diesen Zahlen kann man L auch durch das Schema

$$L = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

darstellen, welches man als $m \times n$-*Matrix* bezeichnet. Die Anwendung der linearen Abbildung auf einen Vektor aus V ist nun wie folgt definiert

$$[\nu_i]_i \mapsto \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix} = \begin{bmatrix} a_{11}\nu_1 + a_{12}\nu_2 + \dots + a_{1n}\nu_n \\ a_{21}\nu_1 + a_{22}\nu_2 + \dots + a_{2n}\nu_n \\ \vdots \\ a_{m1}\nu_1 + a_{m2}\nu_2 + \dots + a_{mn}\nu_n \end{bmatrix}. \tag{1.2}$$

Bei der Hintereinanderschaltung zweier linearer Abbildungen müssen wir die Definitions- und Bildmengen beachten. Sei $A : \mathbb{R}^n \to \mathbb{R}^k$ und $B : \mathbb{R}^k \to \mathbb{R}^m$, dann ist die Abbildung $v \mapsto BAv : \mathbb{R}^n \to \mathbb{R}^m$ wohldefiniert. Für gegebene Basen in $\mathbb{R}^n$, $\mathbb{R}^k$ and $\mathbb{R}^m$ gilt in der Matrixdarstellung

$$BA = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mk} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{bmatrix} = C. \tag{1.3}$$

Dabei ist die j-te Spalte von C durch die Anwendung von B auf die j-te Spalte von A gegeben,

$$[c_{ij}]_i = \big[B[A_{kj}]_k\big]_i = \Big[\sum_{k=1}^n b_{ik}a_{kj}\Big]_i,$$

also

$$c_{ij} = \sum_{k=1}^n b_{ik}a_{kj}, \qquad \text{für } i = 1, \dots, m, \ j = 1, \dots, n. \tag{1.4}$$

Die Menge der linearen Abbildungen zwischen den Vektorräumen V und W bilden wieder einen Vektorraum (punktweise Addition und Multiplikation mit dem Skalar). Diesen bezeichnen wir mit $\mathbb{L}(V, W)$. Ist etwa $V = \mathbb{R}^n$ und $W = \mathbb{R}^m$ und sind Basen festgelegt, so identifizieren wir $\mathbb{L}(\mathbb{R}^n, \mathbb{R}^m)$ mit dem Raum der reellen $m \times n$-Matrizen, auch als $\mathbb{R}^{m,n}$ oder $\mathbb{R}^{m\times n}$ bezeichnet. Seine Dimension ist nm. Eine Basis ist dann etwa durch die Menge der $m \times n$-Matrizen gegeben die genau einen Eintrag 1 haben und 0 sonst.

Spezielle Matrizen sind die Nullmatrix und die Einheitsmatrix im $\mathbb{R}^n$

$$0_n = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} \quad \text{bzw.} \quad Id_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

manchmal auch einfach mit 0 und *Id* bezeichnet. Nicht dargestellte Elemente sind 0.

Beispiel 4 (Drehungen im $\mathbb{R}^3$). *Wir wollen eine Drehung im $\mathbb{R}^3$ mit Winkel $\phi \in [0, 2\pi)$ um die z-Achse beschreiben. Dazu berechnen wir die Bilder der Basis*

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} -\sin(\phi) \\ \cos(\phi) \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Damit ergibt sich die Matrix der Abbildung zu

$$\begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Beispiel 5 (Differentiation auf $\mathbb{P}_n$)**.** *Ist p ein Polynom n-ten Grades, $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, so gilt $d/dx\, p(x) = a_1 + a_2 x + \cdots + n a_n x^{n-1}$. Für die Monome gilt $d/dx\, x^k = k x^{k-1}$. Ordnen wir die Basis $\{x^0, x^1, \ldots, x^n\}$, so ist das k-te Monom durch den euklidischen Vektor e_{k+1} (Bsp. 1) repräsentiert. Die Ableitung d/dx ist also eine lineare Abbildung $\mathbb{P}_n \to \mathbb{P}_n$ die (wegen $e_{k+1} \mapsto k e_k$) in der Monombasis die Darstellung*

$$\begin{bmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & \ddots & \ddots & \\ & & & 0 & n \\ & & & & 0 \end{bmatrix}$$

als $(n+1) \times (n+1)$-Matrix hat.

Eine lineare Abbildung $L : \mathbb{R}^n \to \mathbb{R}$ heißt *Linearform* oder 1-*Form*. In der Matrixdarstellung bekommen wir

$$Lv = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix} = \alpha_1 \nu_1 + \alpha_2 \nu_2 + \cdots + \alpha_n \nu_n = \sum_{i=1}^{n} \alpha_i \nu_i. \tag{1.5}$$

Bezeichneten wir unsere bisherige Vektordarstellung als Spaltenvektoren, so können wir die Menge der Linearformen mit der Menge der *Zeilenvektoren* identifizieren.

Spaltenvektoren und Zeilenvektoren sind spezielle Matrizen in $\mathbb{R}^{n,1}$ bzw. $\mathbb{R}^{1,n}$. Eine eineindeutige Zuordnung eines Zeilenvektors zu dem Spaltenvektor mit gleichen Argumenten, bzw. umgekehrt, ist das *Transponieren*

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \quad \text{bzw.} \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{bmatrix}. \tag{1.6}$$

Allgemeiner definiert man zu $A \in \mathbb{R}^{m,n}$ die transponierte Matrix $A^{\mathrm{T}} \in \mathbb{R}^{n,m}$ durch

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{1m} \\ a_{12} & a_{22} & \ldots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix}. \tag{1.7}$$

Motiviert durch (1.5) definieren wir nun das *Skalarprodukt* zweier Spaltenvektoren $v, w \in \mathbb{R}^n$ durch

$$\langle v, w \rangle := v^{\mathrm{T}} w = \sum_{i=1}^{n} \nu_i \omega_i. \tag{1.8}$$

Zum Beispiel liest man aus (1.2): Die i-te Komponente von Av zu $A \in \mathbb{R}^{n,n}$ und $v \in \mathbb{R}^n$ ist das Skalarprodukt des i-ten Zeilenvektors der Matrix A (i.e., $[a_{ij}]_j$) mit v. Analog erhält man aus (1.4): Die ij-te Komponente von BA zu $B \in \mathbb{R}^{m,k}$ und $A \in \mathbb{R}^{k,n}$ ist das Skalarprodukt des i-ten Zeilenvektors der Matrix B (i.e., $[b_{ik}]_k$) mit dem j-ten Spaltenvektor der Matrix A (i.e., $[a_{kj}]_k$).

Beispiel 6 (Linearform auf $\mathcal{C}^0(I)$)**.** *Es sei $\mathcal{C}^0(I)$ wie in Beispiel 3. Die Abbildung $\mathcal{C}^0(I) \to \mathbb{R} : f \mapsto \int_I f$ ist wohldefiniert und linear.*

Beispiel 7 (Lineare Abbildung auf $\mathcal{C}^0(I)$)**.** *Es sei $\mathcal{C}^0(I)$ wie in Beispiel 3. Nun sei $K : I \times I \to \mathbb{R}$ stetig. Dann ist die Abbildung $\mathcal{C}^0(I) \to \mathcal{C}^0(I) : f \mapsto \int_I K(\,\cdot\,, y)f(y)\,\mathrm{d}y$ wohldefiniert und linear.*

Beispiel 8 (Projektion auf eine Gerade im $\mathbb{R}^n$)**.** *Wir geben einen Vektor $z = [\zeta]_i \neq 0$ im $\mathbb{R}^n$ vor und wir suchen eine lineare Abbildung P deren Bild nur aus Vielfachen von z besteht: $Pv = \alpha(v)z$ für $\alpha(v) \in \mathbb{R}$. α hängt linear von v ab und ist daher eine Linearform, d.h. es gibt einen Vektor $w \in \mathbb{R}^n$ mit $\alpha(v) = \langle w, v\rangle$. Also hat P die Form $Pv = \langle w, v\rangle\, z$.*

Wir wollen im Beispiel 8 gerne die Form $Pv = Av$ erhalten. Dazu berechnen wir, dass A die Form $a_{ij} = \zeta_i \omega_j$ hat. Zu Spaltenvektoren z, w erhält man dies durch $A = zw^{\mathsf{T}} \in \mathbb{R}^{n,n}$ (mit der Definition aus (1.4)). Ohne die Notation $(\,\cdot\,)^{\mathsf{T}}$ schreibt man dies als $z \otimes w$ („z tensor w“). Dieses Produkt heißt *dyadisches Produkt* oder *Tensor-Produkt* (zweier Vektoren). Es gilt also

$$Pv = \langle w, v\rangle\, z = zw^{\mathsf{T}}\, v = z \otimes w\, v.$$

Die Multiplikation von Matrizen ist nicht kommutativ, aber assoziativ. Man beachte $Pv = \langle w, v\rangle\, z = z\,\langle w, v\rangle = z(w^{\mathsf{T}}v) = (zw^{\mathsf{T}})v$. Mittels der euklidischen Basis $\{e_1, \dots, e_n\}$ des $\mathbb{R}^n$ können wir die Basis von $\mathbb{L}(\mathbb{R}^n, \mathbb{R}^n)$ auch als $\{e_i \otimes e_j = e_i e_j^{\mathsf{T}} \,|\, i, j = 1, \dots, n\}$ angeben.

1.3 Adjungierte Operatoren

Durch Transponieren (1.7) geht die lineare Abbildung $A : \mathbb{R}^n \to \mathbb{R}^m$ in die Abbildung $A^{\mathsf{T}} : \mathbb{R}^m \to \mathbb{R}^n$ über. Im Falle hintereinander ausgeführter Matrizen (wie in (1.3)) rechnet man leicht nach dass gilt $C^{\mathsf{T}} = (BA)^{\mathsf{T}} = A^{\mathsf{T}}B^{\mathsf{T}}$. Ein besonderer Fall sind quadratische Matrizen $A : \mathbb{R}^n \to \mathbb{R}^n$ für die $A^{\mathsf{T}} = A$ (d.h. $a_{ij} = a_{ji}$) gilt, diese heißen *symmetrisch*. Aufgrund der Symmetrie folgt $\langle Av, w\rangle = \sum_{ij} a_{ij} v_i w_j = \langle v, Aw\rangle$.

Ist $B \in \mathbb{R}^{n,n}$ eine invertierbare Matrix, so definieren wir das verallgemeinerte Skalarprodukt $\langle v, w\rangle_B := \langle Bv, w\rangle$ für alle $v, w \in \mathbb{R}^n$. Ist $A \in \mathbb{R}^{n,n}$, so ist der adjungierte Operator A^{T_B} definiert durch $\langle Av, w\rangle_B = \langle v, A^{\mathsf{T}_B} w\rangle_B$ für alle $v, w \in \mathbb{R}^n$. Der Zusammenhang mit A^{T} ist durch

$$\begin{aligned}\langle Av, w\rangle_B &= \langle BAv, w\rangle = \langle Av, B^{\mathsf{T}}w\rangle = \langle v, A^{\mathsf{T}}B^{\mathsf{T}}w\rangle = \langle B^{-1}Bv, A^{\mathsf{T}}B^{\mathsf{T}}w\rangle \\ &= \langle Bv, B^{-\mathsf{T}}A^{\mathsf{T}}B^{\mathsf{T}}w\rangle = \langle v, B^{-\mathsf{T}}A^{\mathsf{T}}B^{\mathsf{T}}w\rangle_B\end{aligned}$$

gegeben, d.h., es gilt $A^{\mathsf{T}_B} = B^{-\mathsf{T}}A^{\mathsf{T}}B^{\mathsf{T}}$. Das zeigt auch dass A^{T_B} eindeutig definiert ist. Man nennt nun $A \in \mathbb{R}^{n,n}$ *B-selbstadjungiert*, falls gilt $\langle Av, w\rangle_B = \langle v, Aw\rangle_B$ für alle $v, w \in \mathbb{R}^n$ (also $A^{\mathsf{T}_B} = A$). Damit ist A genau dann symmetrisch, wenn es (Id_n-)*selbstadjungiert* ist.

1.4 Multi-Lineare Abbildungen

Eine Abbildung $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ heißt *Bilinearform* wenn $B \equiv B(\,\cdot\,,\,\cdot\,)$ in beiden Argumenten jeweils linear ist. Zu festem $v \in \mathbb{R}^n$ ist $w \mapsto B(v, w)$ eine Linearform, d.h., es gibt eine Linearform L, welche von v abhängt, so dass $B(v, w) = Lw$ für alle $w \in \mathbb{R}^n$ gilt. Da diese Abhängigkeit linear ist muss $L = (Av)^{\mathsf{T}}$ für ein $A \in \mathbb{R}^{n,n}$ gelten. Damit gilt $B(v, w) = (Av)^{\mathsf{T}}w = \langle Av, w\rangle$. Verallgemeinerte Skalarprodukte $(v, w) \mapsto \langle v, w\rangle_B$ sind spezielle Bilinearformen.

Linear- und Bilinearformen lassen sich zu *Multi-Linearformen*

$$A : \bigotimes_{j=1}^{k} \mathbb{R}^n \to \mathbb{R},\ \ (v_1, \dots, v_k) \mapsto A(v_1, \dots, v_k),$$

A ist in jedem Argument linear,

verallgemeinern. Ein solches A heißt auch k-Form. Die Komponentendarstellung hat dann Koeffizienten $A_{i_1 \dots i_k}$ für $i_1, \dots, i_k \in \{1, \dots, n\}$. Spezielle k-Formen sind etwa die Tensorprodukte $w_1 \otimes w_2 \otimes \dots \otimes w_k$ zu $w_1, \dots, w_k \in \mathbb{R}^n$ mit

$$w_1 \otimes \dots \otimes w_k(v_1, \dots, v_k) := \langle w_1, v_1 \rangle \langle w_2, v_2 \rangle \dots \langle w_k, v_k \rangle.$$

Von besonderer Bedeutung sind hierbei *alternierende Formen* die durch die Relation $A(\dots, v, w, \dots) = -A(\dots, w, v, \dots)$ definiert sind. Eine alternierende n-Form auf dem $\mathbb{R}^n$ ist die *Determinantenform* $(v_1, \dots, v_n) \mapsto \det(v_1, \dots, v_n)$. Für eine Matrix $A \in \mathbb{R}^{n,n}$ aus Spaltenvektoren $a_1, \dots, a_n$ setzen wir $\det(A) := \det(a_1, \dots, a_n)$. Da det genau dann verschwindet wenn die Argumentvektoren *keine* Basis des $\mathbb{R}^n$ bilden, gilt $\det(A) = 0$ genau dann wenn A nicht surjektiv ist. det ist mit der Normierung $\det(Id_n) = 1$ eindeutig. det wird auch als *Volumenform* bezeichnet, denn $|\det(a_1, \dots, a_n)|$ ist gleich dem Volumen des durch $a_1, \dots, a_n$ aufgespannten (n-dimensionalen) Parallelepipeds.

1.5 Eigenwerte

Eine Matrix $A \in \mathbb{R}^{n,n}$ hat einen Eigenwert $\lambda \in \mathbb{R}$ zum Eigenvektor $u \in \mathbb{R}^n \setminus \{0\}$, wenn gilt

$$Au = \lambda u.$$

λ ist genau dann Eigenwert von A, wenn $p(\lambda) = \det(A - \lambda Id_n) = 0$ gilt, d.h., λ eine Nullstelle von p ist. Da p aufgrund der Definition von det ein Polynom n-ten Grades ist, hat A nicht immer Eigenwerte, und wenn, dann weniger als n Eigenwerte. Falls A die n Eigenwerte $\lambda_1, \dots, \lambda_n$ hat, so ist A *diagonalisierbar*, d.h. es gibt eine Basis in der A die *Diagonalgestalt*

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

hat. Die Basis besteht aus der Menge der dazugehörigen Eigenvektoren im $\mathbb{R}^n$. Ist A symmetrisch, so ist A diagonalisierbar.

Ersetzen wir in der Definition $\mathbb{R}$ durch $\mathbb{C}$, so hat A immer Eigenwerte, da Polynome in $\mathbb{C}$ immer Nullstellen haben. Nach geeigneter Basiswahl ergibt sich daraus die Jordan-Normalform (auf die wir hier nicht näher eingehen).

1.6 Komplexität

Für Anwendungszwecke ist es interessant folgendes zu wissen: Wie aufwändig ist die Speicherung eines Objektes und wie aufwändig, etwa gemessen in der Zahl der Multiplikationen, ist die Auswertung einer linearen Abbildung. Meist kommt dieselbe Anzahl an Additionen hinzu, wir erwähnen die Additionen aber nur wenn Sie die Zahl der Multiplikationen weit überschreiten.

Ein Vektor im $\mathbb{R}^n$ benötigt i.A. einen Speicherplatz von n Gleitkommazahlen, eine Matrix im $\mathbb{R}^{n,n}$ i.A. den Speicherplatz von n^2 Kommazahlen. Die Auswertung ‚Matrix mal Vektor' benötigt n^2 Multiplikationen, die Auswertung ‚Matrix mal Matrix' benötigt n^3 Multiplikationen.

Hat die Matrix Diagonalgestalt, so benötigen wir nur n Speicherstellen und n Multiplikationen für die Auswertung ‚Matrix mal Vektor'. Für die Speicherung der Basis müsste man aber ggf. den Speicherplatz von n^2 Kommazahlen zur Verfügung stellen.

Ist A ein Tensorprodukt von Vektoren, etwa $A = v \otimes w$, so benötigt man nur $2n$ Speicherplätze und die Anwendung $Au = \langle u, w \rangle v$ benötigt nur $2n$ Multiplikationen.

Diese Beispiele machen deutlich wie wichtig die Struktur von Matrizen für die Anwendungen sind, denn in praktischen Anwendungen kann n *sehr* groß sein.

1.7 Diverse Notationen

Ein so umfangreiches Gebiet wie die lineare Algebra kennt viele Notationen, je nach Umfeld. Hier werden einige Varianten aufgelistet, mit denen man in der Literatur rechnen muss.

Für Vektoren nimmt man meist lateinische Buchstaben. Oft sind diese besonders hervorgehoben, z.B. indem man Fettdruck und/oder einen anderen Font verwendet: v, v, $\boldsymbol{v}$, $\mathfrak{v}$. Üblich sind auch Vektorpfeile oder Unterstreichungen: $\vec{v}$, $\underline{v}$. In der Physik (Quantenmechanik) gibt es auch die bra-ket-Notation: Das Analogon zum Zeilenvektor ist der *bra-Vektor* $\langle v|$ und das Analogon zum Spaltenvektor ist der *ket-Vektor* $|v\rangle$.

Für Matrizen nimmt man eher große Buchstaben, es folgt analog: A, A, $\boldsymbol{A}$, $\mathfrak{A}$, $\underline{\underline{A}}$.

Für Matrizen und Vektoren werden entweder runde ( ) oder eckige Klammern [ ] verwendet, in der Ingenieurliteratur manchmal auch { }. Meist wird ein Vektor als Spaltenvektor definiert. Im Text schreibt man dann aus Platzgründen den Zeilenvektor oder benutzt $^{\mathrm{T}}$. Analog zu Matlab kann man auch $[\xi_1, \xi_2]$ für die Zeile und $[\xi_1; \xi_2]$ für die Spalte schreiben.

Die Skalare sollen sich von den Vektoren unterscheiden. Man wählt dann getrennte Buchstabenmengen oder andere Fonts. Skalare werden auch oft mit griechischen Buchstaben bezeichnet, während man für Vektoren dann eher lateinische nimmt.

Die ‚Matrix mal Vektor'-Multiplikation und die ‚Matrix mal Matrix'-Multiplikation werden ohne Punkt geschrieben. Für das Skalarprodukt zweier Spaltenvektoren schreibt man $v^{\mathrm{T}}w$ (bei Zeilenvektoren vw^{T}) oder $v \cdot w$ bzw. verwendet die Klammernotationen (v, w), $\langle v, w\rangle$, im Fall der bra-ket-Notation das Bra(c)ket $\langle v \mid w\rangle$. Das Tensorprodukt zweier Spaltenvektoren ist dann vw^{T} oder $v \otimes w$, in der bra-ket-Notation $|v\rangle\langle w|$ (etwa $(|v\rangle\langle w|)|y\rangle = |v\rangle\langle w|y\rangle = \langle w|y\rangle\ |v\rangle$).

Ist der zugrundeliegende Körper $\mathbb{C}$ so benötigen wir zudem die Abbildung $z = a + \mathrm{i}b \mapsto \overline{z} = a - \mathrm{i}b$ die als *komplexe Konjugation* bezeichnet wird. Dies wird entsprechend komponentenweise auf $\mathbb{C}^n$ erweitert. Neben $^{\mathrm{T}}$ definieren wir die *hermitsche Transponierte* einer Matrix $A = [a_{ij}]_{ij} \in \mathbb{C}^{m,n}$ als $A^{\mathrm{H}} = [\overline{a}_{ji}]_{ji} \in \mathbb{C}^{n,m}$, also $A^{\mathrm{H}} = \overline{A}^{\mathrm{T}}$. Insbesondere benötigen wir das *hermitsche Skalarprodukt* $w^{\mathrm{H}}v = \overline{w} \cdot v = \sum_{i=1}^{n} \overline{\omega}_i \xi_i$ zweier Spaltenvektoren. Indem wir allgemein $\langle w, v\rangle = w^{\mathrm{H}}v$ definieren, können wir die bisherige Notation ohne Änderung verwenden.

Eine andere Schreibweise für komplexe Konjugation und die hermitsche Transponierte verwendet * statt $^{\mathrm{H}}$.

2 Analysis

2.1 Normierte Räume

Wir versehen nun den $\mathbb{R}^n$ mit einem Abstandsmaß $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ mit den Voraussetzungen der *Definitheit* ($\|x\| = 0 \Leftrightarrow x = 0$), der Homogenität ($\|\lambda x\| = |\lambda|\,\|x\|$) und der *Dreiecksungleichung* ($\|x + y\| \leq \|x\| + \|y\|$). $\|\cdot\|$ nennt man dann *Norm*. Den Abstand zwischen „2 Punkten" x und y im $\mathbb{R}^n$ definiert man als $\|x - y\|$. Mittels dieser Norm definiert man „offene Mengen" und den Begriff der „Konvergenz" im $\mathbb{R}^n$. Für Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m$ lassen sich damit die Begriffe „Stetigkeit" und „Differenzierbarkeit" definieren.

Eine übliche Wahl ist die *euklidische Norm*,

$$\|x\|_2 = \big\|[x_i]_{i=1,\ldots,n}\big\|_2 := \Big(\sum_{i=1}^{n} x_i^2\Big)^{1/2}.$$

Hier ergibt sich der Zusammenhang $\|x\|_2 = \langle x, x\rangle^{1/2}$ zum obigen Skalarprodukt. Es gibt auch andere gebräuchliche Normen, zum Beispiel die *p-Normen* $\|x\|_p$, wo wir die ‘2' durch $p \in [1, \infty)$ ersetzen. Für $p = \infty$ setzen wir $\|x\|_\infty = \max\{|x_i| \,|\, i = 1, \ldots, n\}$. Nach dem Satz von Heine–Borel ergeben sich aber „vergleichbare" Abstände, das heißt zu gegebenen Normen $\|\cdot\|, \|\cdot\|'$ gibt es positive Konstanten c, C mit $c\|x\|' \leq \|x\| \leq C\|x\|'$ für alle

$x \in \mathbb{R}^n$. Im Allgemeinen hängen diese Konstanten von n ab und für große n sehr groß oder sehr klein werden.

2.2 Differenzierbarkeit

Ziel dieses Abschnittes ist es die Differentiation mehrdimensionaler Funktionen im Zusammenhang mit der Linearen Algebra zu entwickeln.

Es sei Ω eine offene Teilmenge des $\mathbb{R}^n$. Wir studieren differenzierbare Funktionen $F : \Omega \to \mathbb{R}^m$.

Die Definition der Fréchet-Ableitung besagt: F ist in einem Punkt $x_0 \in \mathbb{R}^n$ differenzierbar, falls es eine lineare Abbildung $T_{x_0} : \mathbb{R}^n \to \mathbb{R}^m$ gibt, so dass für hinreichend kleine $\|x - x_0\|_2$

$$\|F(x) - F(x_0) - T_{x_0}(x - x_0)\|_2 \le a(\|x - x_0\|_2)\|x - x_0\|_2$$

gilt, wobei $a : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ eine stetige Funktion ist mit $a(s) \to 0$ für $s \to 0$. Dieses T_{x_0} bezeichnet man dann als $F'(x_0)$, das *Differential* von F in x_0. Als Abbildung gilt $F' : \Omega \to \mathbb{L}(\mathbb{R}^n, \mathbb{R}^m)$, also $F'(x_0) \in \mathbb{L}(\mathbb{R}^n, \mathbb{R}^m)$. Andere Bezeichnungen sind dF, DF oder ∇F. Für $F'(x_0)(x - x_0)$ schreibt man auch $F'(x_0)[x - x_0]$ ($[\cdot]$ für lineare Argumente) oder $F'(x_0; x - x_0)$ (lineare Argumente nach ;), insbesondere wenn es wie hier 2 Sorten Argumente gibt.

Die Idee dieser Definition der Ableitung ist, dass die einfache Formel $F(x_0) - F'(x_0)[x - x_0]$ nahe x_0 eine gute Approximation an $F(x)$ liefert.

Die Frage die wir hier diskutieren wollen ist: Wie berechnet man konkret F'?

Eine Funktion $F : \Omega \to \mathbb{R}^m$ ist ein m-Tupel von m skalaren differenzierbaren Funktionen $f_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, m$, so dass $F(x) = [f_1(x), \ldots, f_m(x)]^{\mathrm{T}}$ gilt. Das Differential $F'(x)$ wird nun durch die *Jakobi-Matrix* dargestellt,

$$F'(x) = \begin{bmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \ldots & \partial_n f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \ldots & \partial_n f_2(x) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \partial_2 f_m(x) & \ldots & \partial_n f_m(x) \end{bmatrix}$$

(falls die f_i stetig differenzierbar sind). ∂_i bezeichnet die partielle Ableitung $\partial_{x_i} = \partial/\partial x_i$. Die Darstellung von F' bezieht auf die gewählten Basen für die Vektordarstellung von x und F. Diese dürfen aber nicht selbst x-abhängig sein.

Beispiel 9 (Ableitung einer linearen Abbildung)**.** *Es sei $F(x) = Ax$ zu $A \in \mathbb{R}^n$. Dann ist $f_i(x) = \sum_{j=1}^n a_{ij} x_j$ und $\partial_j f_i(x) = a_{ij}$. Also gilt: $F'(x) = A$ und $F'(x)[v] = Av$. Die Ableitung ist, wie zu erwarten war, konstant.*

Es sei $f : \mathbb{R}^n \to \mathbb{R}$ differenzierbar. Dann ist $f'(x) = [\partial_j f(x)]_j$ eine Matrix im $\mathbb{R}^{1,n}$ (Zeilenvektor). Da $f'(x)$ eine Linearform ist, können wir $f'(x)[v] = \langle \nabla f(x), v \rangle$ schreiben. $\nabla f(x)$ heißt *Gradient von f in x* und definiert sich durch das Differential und das gewählte Skalarprodukt.

Beispiel 10 (Ableitung einer Funktion $\mathbb{R}^n \to \mathbb{R}$)**.** *Es sei $f(x) = \langle x, x \rangle$. Dann ist $f'(x)[v] = \langle 2x, v \rangle$.*

Beispiel 11 (Ableitung der Determinante)**.** *Wir wollen die Ableitung der Funktion* $\det : \mathbb{R}^{n,n} \to \mathbb{R}$ *in A berechnen. Wir wissen $\det'(A) : \mathbb{R}^{n,n} \to \mathbb{R}$. Wir können $\mathbb{R}^{n,n}$ mit $\mathbb{R}^{n^2}$ identifizieren und aus Beispiel 10 schließen wir*

$$\det{}'(A)[B] = \sum_{i,j=1}^{n} \partial_{ij} \det(A)\, B_{ij}.$$

Aufgrund des Laplaceschen Entwicklungssatzes können wir zu gegebenem ij die Determinante nach der i-ten Zeile entwickeln:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A|ij).$$

Dabei ist $\det(A|ij)$ *die Determinante der* $(n-1)\times(n-1)$*-Untermatrix von* A *die entsteht wenn man die* i*-te Zeile und* j*-te Spalte von* A *streicht. Damit ist die Abhängigkeit von* a_{ij} *linear und es gilt*

$$\partial_{ij} \det(A) = (-1)^{i+j} \det(A|ij).$$

Die Matrix $[(-1)^{i+j}\det(A|ij)]_{ij}$ *heißt* Kofaktormatrix *zu* A, $\mathrm{Cof}(A)$. *Die transponierte dazu heißt* Adjunkte *von* A, $\mathrm{Adj}(A)$, *und diese ist definiert als* $\mathrm{Adj}(A) := \det(A)A^{-1}$. *Somit gilt:*

$$\det{}'(A)[B] = \det(A) \sum_{i,j=1}^{n} A^{-1}_{ji} B_{ij} = \det(A)(A^{-1})^{\mathrm{T}} : B = \det(A)A^{-\mathrm{T}} : B.$$

Letzteres verwendet das Frobenius-Skalarprodukt von Matrizen $C : B := \sum_{i,j=1}^{n} C_{ij}B_{ij}$.

Ein wichtiger Fall ist die Ableitung einer zusammengesetzten Funktion. Zu $G : \mathbb{R}^k \to \mathbb{R}^m$ und $F : \mathbb{R}^n \to \mathbb{R}^k$ betrachten wir $H : \mathbb{R}^n \to \mathbb{R}^m$, $H = G \circ F$, d.h. $H(x) = G(F(x))$. Sind F und G differenzierbar, so ist auch H differenzierbar und es gilt die *Kettenregel*

$$H'(x) = G'(F(x))F'(x).$$

Die Jakobi-Matrix von H ist das Matrix-Produkt der Jakobi-Matrix von G mit der von F.

Beispiel 12 (Koordinatentransformation)**.** *Es sei* $g : D \subseteq \mathbb{R}^n \to \mathbb{R}$ *eine skalare Funktion. Wir wollen nun eine Koordinatentransformation auf dem Definitionsbereich anwenden, d.h. zu* $F : \mathbb{R}^n \to \mathbb{R}^n$ *definieren wir nun* $h : \mathbb{R}^n \to \mathbb{R}$ *durch* $h(x) = g(F(x))$. *Wir erhalten als Formel für die Ableitung*

$$h'(x) = g'(F(x))F'(x).$$

Man beachte, dass h' und g' Zeilenvektoren sind (also $1 \times n$). Wollen wir den Gradienten als Spaltenvektor auffassen, so lautet die Formel $\nabla h(x)^{\mathrm{T}} = \nabla g(F(x))^{\mathrm{T}} F'(x)$, also nach transponieren $\nabla h(x) = F'(x)^{\mathrm{T}} \nabla g(F(x))$!

Höhere Ableitungen sind Multilinearformen, d.h. falls $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ zweimal differenzierbar ist, so ist $F'' : \Omega \to \mathbb{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m)$. Damit ist $F''(x)$ eine Bilinearform mit Werten im $\mathbb{R}^m$, also ein Funktionsvektor (der Länge m) von Bilinearformen. Entsprechend ist die k-te Ableitung in einem Punkt x, $F^{[k]}(x)$, ein Vektor (der Länge m) von k-Formen, $(v_1, \dots, v_k) \mapsto F^{[k]}(x)[v_1, \dots, v_k]$.

Für eine zweimal differenzierbare Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ ist die zweite Ableitung gegeben durch die *Hesse-Matrix*

$$Hf(x) := f''(x) = [\partial_{ij} f(x)]_{i,j=1,n} \in \mathbb{R}^{n,n}.$$

Da $\partial_{ij} f(x) = \partial_{ji} f(x)$ gilt, ist $f''(x)$ eine symmetrische Matrix.

Beispiel 13 (Ableitungen einfacher Formen)**.** *Für die Ableitung von* $F : \mathbb{R}^n \to \mathbb{R}^n$, $F(x) = Ax$ *mit* $A \in \mathbb{R}^{n,n}$, *hatten wir* $F'(x) = A$. *Die weiteren Ableitungen sind offenbar* 0, *als entsprechende Form. Wir betrachten nun noch den nächst einfacheren Fall* $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \langle Ax, x\rangle$. *Statt mit Koeffizienten arbeiten wir mit der* Produktregel*, so dass wir nur lineare Funktionen differenzieren müssen. Für alle* $x, v \in \mathbb{R}^n$ *erhalten wir*

$$f'(x)[v] = \langle Av, x\rangle + \langle Ax, v\rangle = \langle (A + A^{\mathrm{T}})x, v\rangle$$

und das zeigt $f'(x) = (A + A^{\mathrm{T}})x$ *und* $f''(x) = A + A^{\mathrm{T}}$. *Ist* A *symmetrisch, so folgt* $f'(x) = 2Ax$ *und* $f''(x) = 2A$.

2.3 Der Satz von Taylor

Ein sehr wichtiges Argument in der Analysis ist die lokale (d.h. in einer kleinen Umgebung eines gegebenen Punktes x) Approximation einer Funktion durch eine einfachere Funktion, hier ein Polynom. Für $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, welches 2-mal stetig differenzierbar ist, und $x_0 \in \Omega$ folgt

$$F(x) = F(x_0) + F'(x_0)[x - x_0] + \mathcal{O}(\|x - x_0\|_2^2)$$

für alle x die nahe x_0 sind. Dies entspricht der Definition von $F'(x_0)$. Ist F 3-mal stetig differenzierbar so gilt

$$F(x) = F(x_0) + F'(x_0)[x - x_0] + \frac{1}{2}F''(x_0)[x - x_0, x - x_0] + \mathcal{O}(\|x - x_0\|_2^3).$$

Beispiel 14 (Nullstellenapproximation)**.** *Angenommen, wir suchen zu einem stetig differenzierbarem $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ ein x mit $F(x) = 0$ und wir hätten schon eine gewisse Näherung x_0. Dann könnte man versuchen eine Nullstelle x_1 der Näherung an F in x_0 zu finden, die möglicherweise eine noch bessere Approximation ist. Also lösen wir*

$$F(x_0) + F'(x_0)[x_1 - x_0] \overset{!}{=} 0$$

nach x_1 auf. Damit das geht muss die $n \times n$-Matrix $F'(x_0)$ offenbar invertierbar sein. Wenn das geht, erhalten wir

$$x_1 = x_0 - F'(x_0)^{-1}F(x_0).$$

*Dies führt tatsächlich zu Ziel wenn x_0 nahe genug an einer existierten Nullstelle liegt (*Newton-Verfahren*).*

In der Analysis lernt man auch Abschätzungen kennen, die es erlauben die Differenz zwischen F und dieser Approximation, die in $\mathcal{O}(\,\cdot\,)$ versteckt ist, genauer abzuschätzen.