14

Mehrdimensionale Differenziation

Bisher haben wir, was Differenzierbarkeit betrifft, nur Funktionen *einer* reellen Variablen betrachtet. Im einfachsten Fall handelt es sich um *reellwertige Funktionen einer Variablen*, also Funktionen von der Form $\mathbb{R} \to \mathbb{R}$. Im vorangehenden Kapitel betrachteten wir allgemeiner *Kurven*, also Abbildungen der Form $\mathbb{R} \to \mathbb{R}^m$, wobei anstelle von $\mathbb{R}^m$ auch ein beliebiger Banachraum stehen kann. Dies ist adäquat, wenn wir Größen betrachten, die nur von einer Variablen abhängen, wie zum Beispiel der Zeit t.

Mindestens ebenso oft hat man es jedoch auch mit Abbildungen des Typs $\mathbb{R}^n \to \mathbb{R}$ zu tun, wo eine skalare Größe von mehreren Variablen abhängt, und noch allgemeiner mit Abbildungen des Typs $\mathbb{R}^n \to \mathbb{R}^m$, wo m ›abhängige Variablen‹ durch n ›unabhängigen Variablen‹ bestimmt werden. Ein bereits bekanntes Beispiel sind lineare Gleichungssysteme. Für solche Abbildungen können wir die Ableitung allerdings nicht mehr mithilfe von Differenzenquotienten erklären, da die Division durch einen Vektor nicht sinnvoll definiert werden kann – es existiert nur eine Vektorraum-, aber keine Körperstruktur. [1]

Statt dessen charakterisieren wir *Differenzierbarkeit* durch *Approximierbarkeit* durch eine affine Abbildung. Begriffe der linearen Algebra werden dabei eine wesentliche Rolle spielen. Diese enge Verzahnung der infinitesimalen Analysis mit der linearen Algebra ist es auch, was die mehrdimensionale Differenzialrechnung bei der ersten Begegnung schwierig macht.

[1] Eine Ausnahme gibt es – der $\mathbb{R}^2$ kann durch Identifikation mit $\mathbb{C}$ mit einer Körperstruktur versehen werden. Der daraus resultierende Ableitungsbegriff führt jedoch zu einer wesentlichen anderen Theorie, der sogenannten *Funktionentheorie*. Eine *einmal komplex differenzierbare* Funktion ist immer *unendlich oft differenzierbar* und lokal durch ihre Potenzreihe darstellbar, also eine *analytische* Funktion.

14.1 Elemente der Linearen Algebra

Wir betrachten zunächst lineare Abbildungen zwischen beliebigen Banachräumen V und W. Deren Normen bezeichnen wir mit $\|\cdot\|_V$ und $\|\cdot\|_W$ oder nur $\|\cdot\|$, wenn der Bezug aus dem Zusammenhang klar ist.

Definition *Eine lineare Abbildung $A\colon V \to W$ heißt beschränkt, falls*

$$\|A\|_{V,W} \coloneqq \sup_{\|x\|_V=1} \|Ax\|_W < \infty. \quad \rtimes$$

Aufgrund der positiven Homogenität jeder Norm gilt auch A-1

$$\|A\|_{V,W} = \sup_{0\neq x\in V} \frac{\|Ax\|_W}{\|x\|_V}.$$

Daher gilt auch *immer*

$$\|Ax\|_W \leqslant \|A\|_{V,W}\,\|x\|_V\,, \qquad x\in V.$$

Dies werden wir im Folgenden kommentarlos verwenden.

1 **Satz** *Für eine lineare Abbildung $A\colon V\to W$ sind äquivalent:*
(i) *A ist lipschitz auf V.*
(ii) *A ist stetig auf V.*
(iii) *A ist stetig im Nullpunkt.*
(iv) *A ist beschränkt.* $\rtimes$

⟨⟨⟨⟨⟨ (i) ⇒ (ii) und (ii) ⇒ (iii) sind trivial.
(iii) ⇒ (iv) Zu $\varepsilon = 1$ existiert ein $\delta > 0$, so dass

$$\|Ax\|_W \leqslant 1, \qquad \|x\|_V \leqslant \delta.$$

Für $x\in V$ mit $\|x\|_V = 1$ gilt dann

$$\|Ax\|_W = \delta^{-1}\,\|A(\delta x)\|_W \leqslant \delta^{-1}.$$

Also ist A beschränkt, genauer $\|A\|_{V,W} \leqslant \delta^{-1}$.
(iv) ⇒ (i) Folgt aus $\|Au - Av\|_W = \|A(u-v)\|_W \leqslant \|A\|_{V,W}\,\|u-v\|_V$. ⟩⟩⟩⟩⟩

Wir betrachten nun den Raum $L(V,W)$ aller stetigen, oder was dasselbe ist, aller beschränkten linearen Abbildungen $A\colon V\to W$.

2 **Satz** *Auf dem Raum $L(V,W)$ definiert $\|A\|_{V,W}$ eine Norm, die von den Normen auf V und W induzierte Operatornorm. Mit ihr wird $L(V,W)$ zu einem Banachraum.* $\rtimes$

Im Folgenden schreiben wir $\|A\|$ statt $\|A\|_{V,W}$, wenn die beteiligten Räume aus dem Zusammenhang klar sind.

⟨⟨⟨⟨⟨ Die Definitheit und positive Homogenität von $\|\cdot\|$ sind leicht zu sehen. Die Dreiecksungleichung ergibt sich mit

$$\begin{aligned}\|A+B\| &= \sup_{\|x\|=1} \|Ax+Bx\| \\ &\leqslant \sup_{\|x\|=1} (\|Ax\|+\|Bx\|) \\ &\leqslant \sup_{\|x\|=1} \|Ax\| + \sup_{\|x\|=1} \|Bx\| = \|A\|+\|B\|.\end{aligned}$$

Um die Vollständigkeit zu zeigen, sei (A_k) eine Cauchyfolge in $L(V,W)$. Für jedes $x \in V$ ist wegen $\|A_k x - A_l x\| \leqslant \|A_k - A_l\| \, \|x\|$ dann $(A_k x)$ eine Cauchyfolge in W. Aufgrund der Vollständigkeit von W konvergiert also $(A_k x)$, und wir können eine Abbildung $A\colon V \to W$ punktweise definieren durch

$$Ax \coloneqq \lim_{k\to\infty} A_k x.$$

Diese Abbildung ist linear, denn

$$\begin{aligned}A(\lambda x + \mu y) &= \lim \, (A_k(\lambda x + \mu y)) \\ &= \lim \, (\lambda A_k x + \mu A_k y) \\ &= \lambda \lim A_k x + \mu \lim A_k y = \lambda A x + \mu A y.\end{aligned}$$

Sie ist beschränkt, denn es existiert $M = \lim \|A_k\|$ 5.36, und damit gilt

$$\|Ax\| = \lim \|A_k x\| \leqslant \lim \|A_k\| \, \|x\| \leqslant M \, \|x\|, \qquad x \in V.$$

Bleibt noch zu zeigen, dass $A_k \to A$ in der Operatornorm. Da

$$\|(A_k - A)x\| = \lim_{l\to\infty} \|(A_k - A_l)x\| \leqslant \lim_{l\to\infty} \|A_k - A_l\| \, \|x\|$$

für jedes $x \in V$, gilt auch

$$\|A_k - A\| = \sup_{\|x\|=1} \|(A_k - A)x\| \leqslant \lim_{l\to\infty} \|A_k - A_l\| \leqslant \sup_{l \geqslant k} \|A_k - A_l\|.$$

Daraus folgt die Konvergenz in der Operatornorm. ⟩⟩⟩⟩⟩

Bemerkung Der Beweis verwendet an keiner Stelle die Vollständigkeit des Urbildraumes V. Tatsächlich gilt der Satz für jeden normierten Raum V, nur der Bildraum W muss vollständig sein. So ist zum Beispiel der Dualraum $V^* = L(V,\mathbb{R})$ für jeden *normierten* Vektorraum V ein Banachraum, da $\mathbb{R}$ vollständig ist. ⊸

Hilberträume

Unter den Banachräumen spielen die Hilberträume eine besondere Rolle. Diese sind, wie wir bereits in Abschnitt 5.7 gesehen haben, charakterisiert durch die Existenz eines *Skalarprodukts*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R},$$

so dass die Norm gegeben ist durch $_{5.33}$ $\|x\| := \sqrt{\langle x, x \rangle}$.

Für ein Skalarprodukt gilt immer die Cauchy-Schwarzsche Ungleichung $_{5.32}$

$$|\langle x, y \rangle| \leqslant \|x\| \, \|y\| \, .$$

Daher definiert jeder Vektor $v \in V$ ein stetiges lineares Funktional

$$L_v : V \to \mathbb{R}, \quad x \mapsto \langle v, x \rangle \, ,$$

denn wegen $|L_v x| \leqslant \|v\| \, \|x\|$ ist es beschränkt. Das Besondere an Hilberträumen ist unter anderem, dass umgekehrt *jedes* lineare Funktional auch auf diese Weise dargestellt werden kann.

3 **Rieszscher Darstellungssatz** *Sei V ein Hilbertraum. Dann existiert zu jedem $L \in V^*$ ein eindeutiger Vektor $v \in V$, so dass*

$$L = L_v = \langle v, \cdot \rangle \, . \qquad \rtimes$$

⟨⟨⟨⟨⟨ Wir können $\|L\| = 1$ annehmen. Dann existiert in V eine Folge (v_k) mit $\|v_k\| = 1$ und $Lv_k \to 1$. Für jedes $0 < \varepsilon < 8$ existiert dann ein K, so dass

$$Lv_k > 1 - \varepsilon/8 > 0, \qquad k \geqslant K.$$

Wegen $\|L\| = 1$ ist dann

$$\|v_k + v_l\| \geqslant |L(v_k + v_l)| = Lv_k + Lv_l > 2 - \varepsilon/4, \qquad k, l \geqslant K.$$

Mit der Parallelogrammgleichung $_{\text{A-5.41}}$ folgt hiermit

$$\begin{aligned} \|v_k - v_l\|^2 &= 2\,\|v_k\|^2 + \|v_l\|^2 - \|v_k + v_l\|^2 \\ &\leqslant 4 - (2 - \varepsilon/4)^2 = \varepsilon - \varepsilon^2/16 < \varepsilon. \end{aligned}$$

Somit ist (v_k) eine Cauchyfolge und aufgrund der Vollständigkeit von V konvergent. Für den Grenzwert $v = \lim_{k\to\infty} v_k$ gilt dann

$$\|v\| = 1, \qquad Lv = 1.$$

Wir zeigen, dass $L = L_v$.

Sei $x \neq 0$ mit $Lx \geqslant 0$. Wegen $\|L\| = 1$ und $Lv = \|v\|$ gilt für $t > 0$

$$Lx = \frac{L(v + tx) - L(v)}{t} \leqslant \frac{\|v + tx\| - \|v\|}{t}$$

und
$$Lx = \frac{L(v - tx) - L(v)}{-t} \geqslant -\frac{\|v - tx\| - \|v\|}{t}.$$
Also ist
$$-\frac{\|v - tx\| - \|v\|}{t} \leqslant Lx \leqslant \frac{\|v + tx\| - \|v\|}{t}.$$
Für $t \to 0$ haben beide Seiten aufgrund der Regel von l'Hospital $_{8.19}$ denselben Grenzwert
$$\frac{\mathrm{d}}{\mathrm{d}t} \|v + tx\| \Big|_{t=0} = \frac{\langle v, x \rangle}{\|v\|} = \langle v, x \rangle .$$
Somit folgt durch Grenzübergang auf beiden Seiten, dass $Lx = \langle v, x \rangle$.

Die Eindeutigkeit des Vektors v ist eine leichte Übung. ⟩⟩⟩⟩⟩

▪ Endlich dimensionale Räume

Alles bisher Gesagte gilt unabhängig von der Dimension der betrachteten Räume. In einem unendlich-dimensionalen Raum ist es jedoch möglich, dass eine lineare Abbildung *unbeschränkt* und damit *unstetig* ist $_{\text{A-6}}$. Dies ist auf einem endlich-dimensionalen Raum *nicht möglich*, da die Einheitssphäre dort kompakt ist $_{7.27,\,7.29}$. Außerdem lassen sich lineare Abbildungen zwischen endlich-dimensionalen Räumen bequem durch *Matrizen* darstellen. Aus diesen Gründen werden wir uns jetzt auf lineare Abbildungen zwischen endlich-dimensionale Räumen beschränken.

Zunächst ein Wort zur Notation. Im Matrizenkalkül ist es üblich, Vektoren als *Spaltenvektoren*
$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$
zu schreiben. Dabei verzichten wir auf einen Vektorpfeil oder sonstige Auszeichnungen. In einem horizontal laufenden Text ist dies natürlich platzraubend. Deshalb verwenden wir die Schreibweise
$$x = (x_1, .., x_n)^\top,$$
wobei $^\top$ die Transposition bezeichnet, wie sie für Matrizen erklärt ist. Umgekehrt ist $x^\top = (x_1, .., x_n)$ ein Zeilenvektor. Ein Zeilenvektor ist zudem nichts anderes als eine $1 \times n$-Matrix, ein Spaltenvektor eine $n \times 1$-Matrix, und die Transposition überführt das eine in das andere.

Matrizendarstellung

Seien V und W Vektorräume der Dimension n und m, respektive. Sind Basisvektoren $v_1, .., v_n$ in V und $w_1, .., w_m$ in W gewählt, so wird eine lineare Abbildung $A\colon V \to W$ durch eine $m \times n$-Matrix wie folgt dargestellt. — Für jeden Basisvektor v_j gilt

$$Av_j = \sum_{i=1}^{m} a_{ij} w_i$$

mit eindeutig bestimmten Koeffizienten a_{ij}. Für $x = \sum_{j=1}^{n} x_j v_j$ gilt dann

$$\begin{aligned} Ax = \sum_{j=1}^{n} x_j A v_j &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} w_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) w_i = \sum_{i=1}^{m} y_i w_i \end{aligned}$$

mit

$$y_i = \sum_{j=1}^{n} a_{ij} x_j .$$

Diesen Zusammenhang zwischen den Koeffizienten in der Gleichung $y = Ax$ schreibt man im Matrizenkalkül bekanntlich als

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} .$$

Man nennt dann

$$(a_{ij})_{mn} := (a_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$$

die *Matrixdarstellung* von A bezüglich der gewählten Basen in V und W. Deren j-te Spalte $(a_{1j} \,..\, a_{mj})^{\top}$ besteht gerade aus den Koeffizienten des Vektors Av_j.

Eine Matrix ist somit immer eine *Darstellung* einer linearen Abbildung bezüglich einer bestimmten Basis. Wählen wir eine andere Basis, ändert sich auch die Matrix. Die entsprechenden Transformationen werden ausführlich in der Linearen Algebra diskutiert.

Ein Spezialfall ist ein *lineares Funktional* $L\colon V \to \mathbb{R}$. Hier ist $m = 1$, und L wird durch eine $1 \times n$-Matrix, sprich einen n-dimensionalen Zeilenvektor

$$l = (l_1, .., l_n)$$

dargestellt, wobei $l_j = Lv_j$. Für $x = \sum_{j=1}^{n} x_j v_j$ ist dann

14.6

$$Lx = \sum_{j=1}^{n} x_j L v_j = \sum_{j=1}^{n} l_j x_j = (l_1, .., l_n)(x_1, .., x_n)^\top.$$

Die rechte Seite ist das Produkt eines Zeilenvektors mit einem Spaltenvektor, oder, was dasselbe ist, einer $1 \times n$-Matrix mit einer $n \times 1$-Matrix. Das Ergebnis ist ein Skalar.

Ein Skalarprodukt $\langle \cdot, \cdot \rangle$ wird bezüglich einer Basis $v_1, .., v_n$ dargestellt durch die *symmetrische* $n \times n$-Matrix

$$A = (a_{ij})_{1 \leqslant i,j \leqslant n}, \qquad a_{ij} = \langle v_i, v_j \rangle .$$

Denn für $x = \sum_{i=1}^{n} x_i v_i$ und $y = \sum_{j=1}^{n} y_j v_j$ wird ja

$$\langle x, y \rangle = \sum_{i,j=1}^{n} x_i \langle v_i, v_j \rangle y_j = (x_1, .., x_n)(a_{ij})(y_1, .., y_n)^\top = x^\top A y.$$

Die $1 \times n$-Matrix $x^\top$ wird also mit der $n \times 1$-Matrix Ay multipliziert.

▪ Der Standardfall

Der *Standardvektorraum* der Dimension n ist der $\mathbb{R}^n$, und jeder andere n-dimensionale Vektorraum ist zu diesem isomorph. Die *Standardbasis* des $\mathbb{R}^n$ besteht aus den *Standardeinheitsvektoren*

$$e_j := (0, .., 1, .., 0)^\top, \qquad 1 \leqslant j \leqslant n,$$

mit der 1 an der j-ten Stelle. Jeder Vektor hat dann die eindeutige Darstellung

$$x = \sum_{j=1}^{n} x_j e_j = (x_1, .., x_n)^\top.$$

Das *Standardskalarprodukt* ist erklärt durch

$$\langle e_i, e_j \rangle = \delta_{ij} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} .$$

Dadurch wird $e_1, .., e_n$ zu einer *Orthonormalbasis* des $\mathbb{R}^n$. Die Koeffizienten eines Vektors x erhält man hiermit als

$$x_i = \langle e_i, x \rangle , \qquad 1 \leqslant i \leqslant n.$$

Ist $A\colon \mathbb{R}^n \to \mathbb{R}^m$ linear und auch der $\mathbb{R}^m$ mit der Standardbasis versehen, so erhält man die Koeffizienten der Matrixdarstellung von A als

$$a_{ij} = \langle e_i, A e_j \rangle ,$$

denn die j-te Spalte von A enthält ja gerade die Koeffizienten des Vektors Ae_j bezüglich der Standardbasis.

Die vom Standardskalarprodukt induzierte Norm ist die euklidische Norm,

$$\|x\|^2 = \langle x, x \rangle = x_1^2 + .. + x_n^2.$$

Falls Verwechslungsgefahr mit anderen Normen und Skalarprodukten besteht, schreiben wir hierfür genauer $\|\cdot\|_e$ und $\langle \cdot, \cdot \rangle_e$.

Diese Situation bezeichnen wir im Folgenden als den *Standardfall*.

14.8