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

▪ Beschränkte lineare Abbildungen

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 $\|Au - Av\|_W = \|A(u-v)\|_W \leqslant \|A\|_{V,W}\,\|u-v\|_V$. Dies ist übrigens auch die bestmögliche Lipschitzkonstante. ⟩⟩⟩⟩

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 $L(V,W)$. Zu zeigen ist die Existenz einer linearen Abbildung $A \in L(V,W)$, so dass

$$\|A_k - A\|_{V,W} \to 0.$$

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

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

Sie ist beschränkt, denn es existiert $M = \sup \|A_k\|$ 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

$$\|A_k - A\| \leqslant \lim_{l\to\infty} \|A_k - A_l\| \leqslant \sup_{l\geqslant k} \|A_k - A_l\|.$$

Letzteres konvergiert gegen Null für $k \to \infty$. ⟩⟩⟩⟩⟩

Der Beweis verwendet an keiner Stelle die Vollständigkeit des Urbildraumes V. Tatsächlich muss nur der Bildraum W vollständig sein. Es gilt somit folgender

Zusatz *Der Satz gilt auch, wenn V nur ein normierter Raum ist. Insbesondere gilt er für den Dualraum*

$$V^* \coloneqq L(V, \mathbb{R})$$

eines normierten Raums V. ⋊

▪ Hilberträume

Unter den Banachräumen spielen die Hilberträume eine besondere Rolle. Diese sind – siehe Abschnitt 5.7 – charakterisiert durch die Existenz eines *Skalarprodukts*

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

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

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 lineares Funktional $L_v \in V^*$ durch

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

denn wegen $|L_v x| \leqslant \|v\| \, \|x\|$ ist es beschränkt. Das Besondere an Hilberträumen ist unter anderem, dass umgekehrt *jedes* lineare Funktional in V^* 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*

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

⟨⟨⟨⟨ Wir können $\|L\| = 1$ annehmen. Dann existiert in V eine Folge (v_k) mit $\|v_k\| = 1$ und $Lv_k \to 1$. Wir zeigen zunächst, dass dies eine Cauchyfolge ist.

Für jedes $0 < \varepsilon < 8$ existiert 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 $_{A\text{-}5.41}$ 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 8.20 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 darstellenden 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 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 Matrizen-Transposition bezeichnet. 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 $A: V \to W$ 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

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

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

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

Man nennt dann

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

14.6

die *Matrixdarstellung* von A bezüglich der gewählten Basen in V und W. Deren j-te Spalte $(a_{1j} \,..\, a_{mj})^\top$ besteht aus den Koeffizienten des Vektors Av_j bezüglich der Basis $w_1, .., w_m$.

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 wichtiger 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

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

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*

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

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* auf dem $\mathbb{R}^n$ ist erklärt durch

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

wobei man δ_{ij} als *Kronecker-Delta* bezeichnet. Damit wird $e_1, .., e_n$ zu einer *Orthonormalbasis* des $\mathbb{R}^n$, und die Koeffizienten eines Vektors x erhält man 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

$$a_{ij} = \langle e_i, Ae_j \rangle ,$$

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*.

14.2 Totale Ableitung

Wir betrachten zunächst weiterhin Abbildungen zwischen beliebigen Banachräumen. Da es vorerst nur um lokale Aspekte geht, verzichten wir auf eine explizite Bezeichnung des Definitionsbereichs und schreiben

$$f\colon\ V \hookrightarrow W,$$

wenn es eine nichtleere offene Teilmenge $\Omega \subset V$ gibt, so dass $f\colon \Omega \to W$. Zu jedem Punkt im Definitionsbereich von f existiert also eine offene Umgebung, auf der f definiert ist. Dies ist die für uns zur Zeit wesentliche Eigenschaft des Definitionsbereiches einer Abbildung.

Die Normen auf V und W bezeichnen wir nun kürzer mit $|\cdot|_V$ und $|\cdot|_W$, oder noch einfacher $|\cdot|$, wenn der Bezug aus dem Zusammenhang klar ist. Die Doppelstriche $\|\cdot\|$ reservieren wir für die Operatornorm.

4 **Definition** *Eine Abbildung $f\colon V \hookrightarrow W$ heißt im Punkt a total differenzierbar oder kurz differenzierbar, wenn es eine stetige lineare Abbildung $L\colon V \to W$ gibt, so dass*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - Lh|_W}{|h|_V} = 0.$$

In diesem Fall heißt L die totale Ableitung von f im Punkt a und wird mit $Df(a)$ bezeichnet. ⋊

Bemerkungen *a.* Die Eindeutigkeit dieser Ableitung und damit die Berechtigung der Bezeichnung $Df(a)$ zeigen wir gleich $_5$.

b. Diese Definition beinhaltet unsere früheren Definitionen der Differenzierbarkeit. Im Fall *einer* unabhängigen Variablen, also $V = \mathbb{R}$, ist f eine Kurve in W. Eine lineare Abbildung $L\colon \mathbb{R} \to W$ wird durch einen Vektor $v \in W$ dargestellt, und wir erhalten die Charakterisierung der Ableitung $Df(a) = \dot{f}(a)$ entsprechend dem zweiten Differenzierbarkeitssatz $_{13.1}$. Ist auch $W = \mathbb{R}$, so wird dieser Vektor durch eine reelle Zahl dargestellt, und wir erhalten die Ableitung einer reellen Funktion, $Df(a) = f'(a)$, einer Variable wie im ersten Differenzierbarkeitssatz $_{8.1}$.

c. Ist V von endlicher Dimension, so ist *jede* lineare Abbildung $A\colon V \to W$ stetig und diese Forderung redundant. Dies ist der einzige, aber wesentliche Unterschied zwischen dem endlich- und unendlich-dimensionalen Fall, was die Definition der totalen Ableitung angeht. ⊸

Die Landausymbole

Das Arbeiten mit Grenzwerten wie in der letzten Definition lässt sich mithilfe der *Landausymbole* erheblich vereinfachen.

Definition *Sei $\dot{\Omega}$ eine punktierte Umgebung von $0 \in V$, und $f, g\colon \dot{\Omega} \to W$ seien zwei Abbildungen mit $g \neq 0$ auf $\dot{\Omega}$. Dann ist*

$$f(h) = O(g(h)) \quad :\Leftrightarrow \quad \sup_{h\in\dot{\Omega}} \frac{|f(h)|}{|g(h)|} < \infty,$$

und man sagt, f ist von der Ordnung groß-O von g auf Ω. Ferner ist

$$f(h) = o(g(h)) \quad :\Leftrightarrow \quad \lim_{h\to 0} \frac{|f(h)|}{|g(h)|} = 0,$$

und man sagt, f ist von der Ordnung klein-o von g für $h \to 0$. ⋊

Es ist also f groß-O von g auf $\dot{\Omega}$, wenn es eine Konstante M gibt, so dass

$$|f(h)| \leqslant M\,|g(h)|, \qquad h \in \dot{\Omega}.$$

Sie ist klein-o von g, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass

$$|f(h)| \leqslant \varepsilon\, |g(h)|, \qquad 0 < |h| < \delta.$$

Die Landausymbole stehen übrigens nicht für eine bestimmte Funktion, sondern für *Klassen* von Funktionen. Eine Formulierung wie $O(h) + O(h) = O(h)$ ist daher zulässig und korrekt.

Rechenregeln *Für das Rechnen mit den Landausymbolen gilt:*

(i) *Eine $o(h)$-Abbildung ist auch $O(h)$.*
(ii) *Eine Linearkombination von $O(h)$-Abbildungen ist wieder $O(h)$.*
(iii) *Dasselbe gilt mit $o(h)$ anstelle von $O(h)$.*
(iv) *Sind f und g beide $O(h)$, so ist auch $(f \circ g)(h) = O(h)$.*
(v) *Ist sogar $f(h) = o(h)$ oder $g(h) = o(h)$, so ist $(f \circ g)(h) = o(h)$.* ⋊

⟨⟨⟨⟨⟨ Der Beweis ist als Übung überlassen. ⟩⟩⟩⟩⟩

▷ A. Es ist $\sin t = O(t)$ und $1 - \cos t = O(t^2)$.
B. Für lineare Abbildung A ist beschränkt genau dann, wenn $Ah = O(h)$.
C. Ist $\langle \cdot, \cdot \rangle$ eine bilineare Form auf V und $A \in L(V)$, so ist

$$\langle Ah, h \rangle = O(\langle h, h \rangle) = o(h).$$

D. Für $f \in C^{n+1}(I)$ ist aufgrund der Restgliedformeln

$$f(a+h) = T_a^n f(h) + O(h^{n+1}).$$ ◁

Mit den Landausymbolen können wir Differenzierbarkeit in einem Punkt auf folgende handliche Art charakterisieren.

5 **Satz** *Eine Abbildung $f : V \hookrightarrow W$ ist im Punkt a differenzierbar genau dann, wenn es ein $L \in L(V, W)$ gibt, so dass*

$$f(a+h) = f(a) + Lh + o(h).$$

Die Ableitung ist dann $Df(a) = L$, und sie ist eindeutig bestimmt. ⋊

⟨⟨⟨⟨⟨ Die erste Behauptung folgt aus den Definitionen der Differenzierbarkeit und des Landausymbols. Ist $\Lambda : V \to W$ eine weitere lineare Abbildung mit der geforderten Eigenschaft, also

$$f(a+h) = f(a) + \Lambda h + o(h),$$

so ergibt die Subtraktion beider Gleichungen $(L - \Lambda)(h) = o(h)$. Dies ist für eine lineare Abbildung nur möglich, wenn $L - \Lambda = 0$ A-5. ⟩⟩⟩⟩⟩

Nun definieren wir Differenzierbarkeit *in jedem Punkt* wie üblich.

Definition *Eine Abbildung $f: V \hookrightarrow W$ heißt total differenzierbar, wenn sie in jedem Punkt ihres Definitionsbereiches total differenzierbar ist.* ⋊

Die Ableitung von f definiert in diesem Fall eine Abbildung

$$Df: V \hookrightarrow L(V,W), \quad x \mapsto Df(x)$$

in den Vektorraum $L(V,W)$ aller linearen Abbildungen von V nach W. Ist $\dim V = 1$, also f eine Kurve, so ist $L(V,W) \simeq W$ und $Df: V \hookrightarrow W$ wiederum ein Kurve wie f. Ist dagegen $\dim V > 1$, so ist eine solche Identifikation nicht mehr möglich, und Df ist von einem wesentlich anderen Typ als die Abbildung f selbst. Mehr dazu am Ende von Abschnitt 6.

6 ⊳ A. *Affine Abbildungen:* Eine affine Abbildung

$$f: V \to W, \quad f(x) = Ax + b$$

mit $A \in L(V,W)$ und $b \in W$ ist in jedem Punkt differenzierbar. Denn es gilt

$$f(a+h) = A(a+h) = f(a) + Ah,$$

der o-Term ist also gar nicht erst vorhanden, und wir erhalten $Df(a) = A$ in jedem Punkt a. Die Ableitung als *Abbildung*

$$Df: V \to L(V,W), \quad a \mapsto Df(a) = A,$$

ist somit *konstant.*

B. *Quadratische Formen:* Ist $\langle \cdot , \cdot \rangle$ eine bilineare Form auf V und $A \in L(V)$, so nennt man

$$f: V \to \mathbb{R}, \quad f(x) = \langle Ax, x \rangle$$

eine *quadratische Form* auf V. Dabei kann man ohne Beschränkung der Allgemeinheit annehmen, dass A *symmetrisch* ist, also

$$\langle Ax, y \rangle = \langle x, Ay \rangle, \qquad x, y \in V,$$

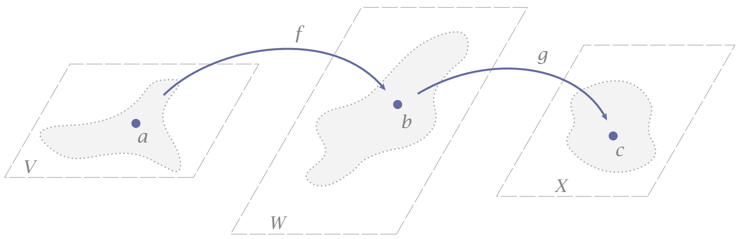
gilt A-7. Dann folgt

$$\begin{aligned} f(x+h) &= \langle A(x+h), x+h \rangle \\ &= \langle Ax, x \rangle + \langle Ax, h \rangle + \langle Ah, x \rangle + \langle Ah, h \rangle \\ &= f(x) + 2\langle Ax, h \rangle + o(h). \end{aligned}$$

Also ist f differenzierbar, und $Df(x)$ ist die lineare Abbildung $h \mapsto 2\langle Ax, h \rangle$. Die Ableitung selbst ist die Abbildung

$$Df: V \to L(V,\mathbb{R}), \quad x \mapsto Df(x) = 2\langle Ax, \cdot \rangle.$$ ⊲

Abb 1 Zur Kettenregel



▪ Kettenregel

Man sieht sofort, dass Differenziation eine lineare Operation ist. Sind also $f, g\colon V \hookrightarrow W$ im Punkt a differenzierbar, so auch $\lambda f + \mu g\colon V \hookrightarrow W$, und es gilt

$$D(\lambda f + \mu g)(a) = \lambda Df(a) + \mu Dg(a).$$

Interessanter ist die Verknüpfung zweier differenzierbarer Abbildungen. Für diese gilt die allgemeine

7 **Kettenregel** *Ist $f\colon V \hookrightarrow W$ im Punkt a und $g\colon W \hookrightarrow X$ im Punkt $f(a)$ differenzierbar, so ist auch $g \circ f\colon V \hookrightarrow X$ im Punkt a differenzierbar, und es gilt*

$$D(g \circ f)(a) = Dg(f(a))Df(a). \qquad \rtimes$$

Die Ableitung von $g \circ f$ ist also die Verknüpfung der beiden linearen Abbildungen

$$Dg(b)\colon\ W \to X, \qquad Df(a)\colon\ V \to W,$$

wobei $b = f(a)$. Offensichtlich kommt es hierbei auf die Reihenfolge der Faktoren an, denn für $V \neq X$ wäre eine umgekehrte Verknüpfung der linearen Abbildungen erst gar nicht definiert.

⟨⟨⟨⟨⟨ Der Beweis verläuft genau wie im eindimensionalen Fall 8.5. Nach Voraussetzung ist f in einer Umgebung von a, und g in einer Umgebung von $b = f(a)$ definiert. Für hinreichend kleine h und k gilt somit

$$f(a+h) = f(a) + Ah + o(h), \qquad A = Df(a),$$
$$g(b+k) = g(b) + Bk + o(k), \qquad B = Dg(b).$$

Wir setzen $k = Ah + o(h)$. Ist h hinreichend klein, so ist auch $Ah + o(h) = O(h)$ hinreichend klein, und g ist für k wohldefiniert. Mit dieser Wahl von k erhalten wir

$$\begin{aligned}(g\circ f)(a+h)) &= g(f(a)+Ah+o(h))\\ &= g(f(a))+B(Ah+o(h))+o(Ah+o(h))\\ &= (g\circ f)(a)+BAh+o(h),\end{aligned}$$

denn

$$Bo(h)=o(h), \qquad o(Ah+o(h))=o(h).$$

Somit 5 ist $g\circ f$ im Punkt a differenzierbar mit Ableitung BA, was der Behauptung entspricht. ⟫⟫⟫

8 ▷ A. *Lineare Abbildung:* Ist $f\colon V \hookrightarrow W$ differenzierbar und $A\colon W\to X$ eine stetige lineare Abbildung, so ist auch

$$Af\colon\ V\hookrightarrow X$$

differenzierbar, und für jeden Punkt x im Definitionsbereich ist 6

$$D(Af)(x)=(DA)(f(x))\,Df(x)=ADf(x).$$

B. *Abbildung einer Kurve:* Ist $\varphi\colon I\to V$ eine differenzierbare Kurve in V und $f\colon V\to W$ differenzierbar, so ist

$$f\circ\varphi\colon\ I\to W$$

eine differenzierbare Kurve in W mit Ableitung

$$(f\circ\varphi)^{\boldsymbol{\cdot}}(t)=Df(\varphi(t))\dot\varphi(t), \qquad t\in I.$$

Im Falle einer Geraden φ, mit $\varphi(t)=x+tv$, ist insbesondere

$$(f\circ\varphi)(t)^{\boldsymbol{\cdot}}\Big|_{t=0}=f(x+tv)^{\boldsymbol{\cdot}}\Big|_{t=0}=Df(x)v.$$

C. *Quadratische Form entlang einer Kurve:* Ist $\varphi\colon I\to V$ eine differenzierbare Kurve in V und

$$f\colon\ V\to R, \quad f(x)=\langle Ax,x\rangle$$

eine quadratische Form auf V 6 mit symmetrischem, beschränktem A, so ist

$$\Phi=f\circ\varphi\colon\ I\to\mathbb{R}, \quad \Phi(t)=\langle A\varphi(t),\varphi(t)\rangle$$

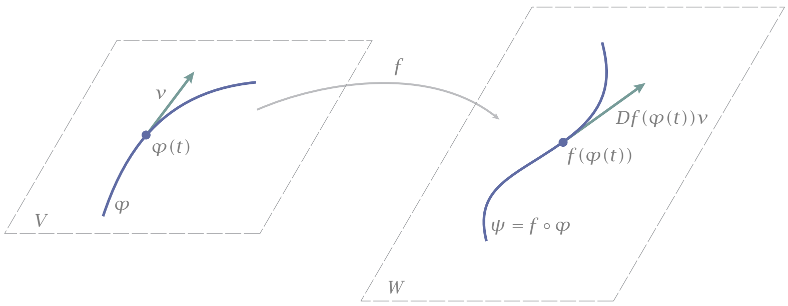
eine differenzierbare reelle Funktion einer Variablen. Ihre Ableitung ist

$$\Phi'(t)=Df(\varphi(t))\dot\varphi(t)=2\,\langle A\varphi(t),\dot\varphi(t)\rangle\,.$$

Speziell für eine Gerade φ mit $\varphi(t)=x+tv$ ist

$$\Phi'(0)=2\,\langle Ax,v\rangle\,. \qquad ◁$$

Abb 2 Die Kurve $f \circ \varphi$ und ihre Ableitung



Es sei nochmals betont, dass alles bisher Gesagte unabhängig von der Dimension der beteiligten Räume gilt. Im unendlich-dimensionalen Fall kommt einzig noch die Forderung hinzu, dass alle beteiligten linearen Abbildungen auch beschränkt sind.

14.3 Richtungsableitungen und Jacobimatrix

Neben dem Begriff der totalen Ableitung gibt es noch einen schwächeren Ableitungsbegriff, den der *Richtungsableitung*. Ist a ein Punkt im Definitionsbereich von $f : V \hookrightarrow W$ und $v \in V$, so ist die Abbildung

$$\mathbb{R} \hookrightarrow W, \quad t \mapsto f(a + tv)$$

in einer Umgebung von 0 wohldefiniert und stetig und definiert somit eine *Kurve in W*. Hierfür haben wir die Ableitung bereits ohne Rückgriff auf die totale Ableitung im vorangehenden Kapitel erklärt.

Definition *Sei $f : V \hookrightarrow W$ im Punkt a definiert und $v \in V$. Dann heißt*

$$\partial_v f(a) \coloneqq f(a + tv)^{\cdot}\Big|_{t=0} \in W,$$

falls diese Ableitung existiert, die Richtungsableitung von f im Punkt a in Richtung v. ⋊

▷ A. Für eine affine Abbildung erhält man

$$\begin{aligned}\partial_v(Ax + b) &= (A(x + tv) + b)^{\cdot}\Big|_{t=0} \\ &= (Ax + b + tAv)^{\cdot}\Big|_{t=0} = Av.\end{aligned}$$

B. Für eine quadratische Form gilt

$$\begin{aligned}\partial_v \langle Ax,x\rangle &= \langle A(x+tv),x+tv\rangle^{\cdot}\Big|_{t=0} \\ &= \left(\langle Ax,x\rangle + 2t\,\langle Ax,v\rangle + t^2\langle Av,v\rangle\right)^{\cdot}\Big|_{t=0} = 2\,\langle Ax,v\rangle\,.\end{aligned}$$

Totale Differenzierbarkeit in einem Punkt impliziert die Existenz aller Richtungsableitungen in diesem Punkt. Denn die Gerade $t \mapsto a+tv$ ist differenzierbar, und die Kettenregel $_7$ ergibt

$$\partial_v f(a) = f(a+tv)^{\cdot}\Big|_{t=0} = Df(a)(a+tv)^{\cdot}\Big|_{t=0} = Df(a)v.$$

Somit gilt folgender

9 **Satz** *Ist $f\colon V \hookrightarrow W$ im Punkt a total differenzierbar, so existieren dort auch alle Richtungsableitungen, und es gilt*

$$\partial_v f(a) = Df(a)v. \qquad \rtimes$$

Aus der Existenz aller Richtungsableitungen folgt allerdings nicht die totale Differenzierbarkeit, wie das folgende Beispiel zeigt.

10 ⊳ Betrachte

$$f\colon\ \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) \triangleq \frac{x^2y}{x^2+y^2}.$$

Für jeden Vektor v gilt $f(tv) = tf(v)$, wie man sofort nachrechnet. Daher ist

$$\partial_v f(0) = f(tv)'|_{t=0} = f(v).$$

Diese Abbildung ist aber *nicht linear* in v. Dies müsste sie aber sein, wenn f im Punkt 0 total differenzierbar wäre. Also ist f in 0 nicht total differenzierbar. ◁

▪ Partielle Ableitungen

Im Standardraum $\mathbb{R}^n$ mit der Standardbasis $e_1,..,e_n$ spielen die Ableitungen in Richtung der Einheitsvektoren eine besondere Rolle.

Definition *Sei $f\colon \mathbb{R}^n \hookrightarrow W$ im Punkt a definiert. Dann heißt*

$$\partial_j f(a) \coloneqq \partial_{e_j} f(a) = f(a+te_j)^{\cdot}\Big|_{t=0} = f(a_1,..,a_j+t,..,a_n)^{\cdot}\Big|_{t=0},$$

falls diese Ableitung existiert, die j-te partielle Ableitung von f im Punkt a. $\rtimes$

Man betrachtet also f als Funktion *nur der j-ten Koordinate*, während alle anderen Koordinaten fixiert sind, und bildet hiervon die Ableitung wie im Fall einer Kurve. Andere übliche Bezeichnungen hierfür sind

$$\frac{\partial f}{\partial x_j}(a), \qquad \partial_{x_j} f(a), \qquad f_{x_j}(a).$$

Aus der Existenz aller partiellen Ableitungen in einem Punkt folgt allerdings nicht einmal die Stetigkeit der Abbildung an dieser Stelle:

11 ▷ Betrachte

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) \triangleq \frac{2xy}{x^2+y^2}.$$

Aus $f(x,0)=0$ und $f(0,y)=0$ folgt sofort

$$\partial_x f(0,0) = 0, \qquad \partial_y f(0,0) = 0.$$

Die Funktion ist im Nullpunkt aber nicht einmal *stetig*, denn entlang einer Nullpunktsgeraden $t \mapsto (t\cos\varphi, t\sin\varphi)$ gilt

$$f(t\cos\varphi, t\sin\varphi) = \frac{2\cos\varphi\sin\varphi}{\cos^2\varphi+\sin^2\varphi} = \sin 2\varphi, \qquad t \neq 0.$$

Also gilt auch

$$\lim_{t\to 0} f(t\cos\varphi, t\sin\varphi) = \sin 2\varphi.$$

Der Grenzwert hängt somit von der Richtung ab und nimmt alle Werte im Intervall $[-1,1]$ an. Also ist f im Nullpunkt nicht stetig und damit auch nicht differenzierbar. ◁

▪ Jacobimatrix

Im Standardfall haben wir es mit einer Abbildung der Gestalt

$$f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m, \quad f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} f_1(x_1,..,x_n) \\ \vdots \\ f_m(x_1,..,x_n) \end{pmatrix}$$

zu tun. Ihre totale Ableitung im Punkt a ist eine lineare Abbildung

$$Df(a)\colon \mathbb{R}^n \to \mathbb{R}^m.$$

Ihre Matrixdarstellung bezüglich der Standardbasen heißt die *Jacobimatrix* oder *Funktionalmatrix* von f im Punkt a.

Satz *Ist $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ im Punkt a total differenzierbar, so wird ihre totale Ableitung $Df(a)$ dargestellt durch die Jacobi- oder Funktionalmatrix*

$$Jf(a) := (\partial_j f_i(a))_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}. \qquad ⋊$$

⟨⟨⟨⟨⟨ Die ij-te Komponente a_{ij} dieser Matrix ist gegeben durch

$$a_{ij} = \langle e_i, Df(a)e_j\rangle = \langle e_i, \partial_j f(a)\rangle = \partial_j \langle e_i, f(a)\rangle,$$

und $\langle e_i, f(a)\rangle = f_i(a)$ ist die i-te Komponente von f. Das ergibt die Behauptung. ⟩⟩⟩⟩⟩

Andere Schreibweisen für die Jacobimatrix sind

$$Jf(a) = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{mn} = (f_{i,x_j}(a))_{mn} = \begin{pmatrix} f_{1,1}(a) & \cdots & f_{1,n}(a) \\ \vdots & & \vdots \\ f_{m,1}(a) & \cdots & f_{m,n}(a) \end{pmatrix}.$$

A. *Affine Abbildung in Koordinaten:* Im Standardfall besitzt eine affine Abbildung $f\colon \mathbb{R}^n \to \mathbb{R}^m$ die Darstellung

$$f(x) = Ax + b = \left(\sum_{j=1}^{n} a_{ij}x_j + b_i\right)_{1\leqslant i\leqslant m}.$$

Dann ist in jedem Punkt

$$\partial_j f_i(x) = a_{ij},$$

die Jacobimatrix von f ist somit

$$Jf(x) = (a_{ij})_{mn} = A.$$

B. *Quadratische Form in Koordinaten:* Eine quadratische Form $f\colon \mathbb{R}^n \to \mathbb{R}$ besitzt im Standardfall die Darstellung

$$f(x) = \langle Ax, x\rangle = \sum_{1\leqslant k,l\leqslant n} a_{kl}x_k x_l$$

mit symmetrischen Koeffizienten $a_{kl} = a_{lk}$. Die Jacobimatrix einer solchen skalaren Funktion ist ein $1\times n$-Zeilenvektor mit Komponenten $\partial_j f(x)$. Für diese finden wir mit der Produktregel für Funktionen einer Variablen

$$\partial_j f(x) = \sum_{k=1}^{n} a_{kj}x_k + \sum_{l=1}^{n} a_{jl}x_l = 2\sum_{k=1}^{n} a_{jk}x_k = 2\,\langle Ax, e_j\rangle.$$

Somit ist

$$Jf(x) = 2(Ax)^\top.$$

Die Jacobimatrix der Verknüpfung zweier linearer Abbildungen ist das Matrixprodukt ihrer Jacobimatrizen. Die Kettenregel erhält damit im Standardfall die folgende Form.

12 **Kettenregel im Standardfall** *Ist $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ im Punkt a und $g\colon \mathbb{R}^m \hookrightarrow \mathbb{R}^s$ im Punkt $f(a)$ differenzierbar, so ist auch $g\circ f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^s$ im Punkt a differenzierbar, und es gilt*

$$J(g\circ f)(a) = Jg(f(a))\,Jf(a).$$

▹ A. Wendet man diese Formel auf die Standard-Einheitsvektoren an, so erhält man für die partiellen Ableitungen

$$\partial_j(g \circ f)(a) = \sum_{k=1}^{m} \frac{\partial g}{\partial y_k}(f(a))\frac{\partial f_k}{\partial x_j}(a), \qquad 1 \leqslant j \leqslant n.$$

Schreibt man f einfacher als $y = y(x)$ und $b = f(a)$, so erhält man die leicht zu merkende Formel

$$\partial_j(g \circ f)(a) = \sum_{k=1}^{m} \frac{\partial g}{\partial y_k}(b)\frac{\partial y_k}{\partial x_j}(a), \qquad 1 \leqslant j \leqslant n.$$

B. Betrachte

$$f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}, \quad f(x) = |x|_e .$$

Für $x \neq 0$ ist dann

$$\partial_j f(x) = \frac{x_j}{|x|_e}, \qquad 1 \leqslant j \leqslant n,$$

und damit

$$Jf(x) = \frac{x^{\top}}{|x|_e}. \qquad ◃$$

▪ Ein Differenzierbarkeitskriterium

Wir kennen nun die Begriffe der totalen Ableitung, der Richtungsableitung und der partiellen Ableitung. Dabei zieht die Existenz der totalen Ableitung diejenige aller anderen Ableitungen nach sich 9. Wie aber verifiziert man die Existenz der totalen Ableitung? Die *Existenz* aller partiellen oder aller Richtungsableitungen reicht offensichtlich nicht aus 10 & 11.

Es stellt sich heraus, dass die *Stetigkeit* aller partiellen Ableitungen eine hinreichende Bedingung darstellt. Dabei beschränken wir uns auf den Standardfall und die Annahme, dass die partiellen Ableitungen auf dem *ganzen* Definitionsbereich stetig sind.

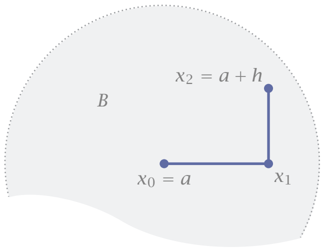
13 **Differenzierbarkeitskriterium** *Existieren sämtliche partiellen Ableitungen von $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ und sind diese stetig, so ist f total differenzierbar, und die Abbildung $Df\colon \mathbb{R}^n \hookrightarrow L(\mathbb{R}^n, \mathbb{R}^m)$ ist ebenfalls stetig.* ⋊

⟪⟪ Betrachte f auf einer nichtleeren Kugel $B = \{|x-a| < r\}$ in seinem Definitionsbereich. Ist $a + h \in B$, so liegen auch die Punkte

$$x_k = a + h_1 e_1 + .. + h_k e_k, \qquad 0 \leqslant k \leqslant n,$$

Abb 3

Zum Beweis des Differenzierbarkeitskriteriums



sämtlich in B, wobei $x_0 = a$ und $x_n = a + h$. Es gilt dann

$$f(a+h) - f(a) = f(x_n) - f(x_0) = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})).$$

Für jeden Summanden gilt

$$\begin{aligned} f(x_i) - f(x_{i-1}) &= f(x_{i-1} + th_ie_i)\Big|_0^1 \\ &= \int_0^1 f(x_{i-1} + th_ie_i)'\,\mathrm{d}t \\ &= \int_0^1 \partial_i f(x_{i-1} + th_ie_i)h_i\,\mathrm{d}t \\ &= \partial_i f(a)h_i + h_i \int_0^1 (\partial_i f(x_{i-1} + th_ie_i) - \partial_i f(a))\,\mathrm{d}t. \end{aligned}$$

Aufgrund der Definition der x_i und der Stetigkeit der partiellen Ableitungen gilt

$$\partial_i f(x_{i-1} + th_ie_i) - \partial_i f(a) \to 0, \qquad h \to 0,$$

gleichmäßig für $0 \leqslant t \leqslant 1$. Das letzte Integral ist daher $o(1)$, und damit

$$f(x_i) - f(x_{i-1}) = \partial_i f(a)h_i + o(h_i).$$

Insgesamt erhalten wir

$$f(a+h) - f(a) = \sum_{i=1}^{n} (\partial_i f(a)h_i + o(h_i)) = \sum_{i=1}^{n} \partial_i f(a)h_i + o(h).$$

Da die Summe eine lineare Abbildung in h darstellt, ist f in a total differenzierbar mit

$$Df(a)h = \sum_{i=1}^{n} \partial_i f(a)h_i.$$

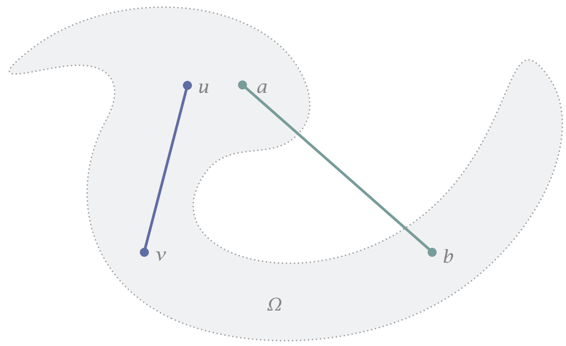
Die Stetigkeit der Ableitung folgt aus der Stetigkeit der $\partial_i f$. ⟩⟩⟩⟩⟩

Die Differenzierbarkeit einer Abbildung $f : \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ stellt man somit fest, indem man die Existenz sämtlicher partiellen Ableitungen und deren Stetigkeit nachweist. Solche Abbildungen nennt man *von der Klasse* C^1 6 .

14.19

Abb 4

$[u,v] \subset \Omega$,
$[a,b] \not\subset \Omega$



14.4 Der Mittelwertsatz

Der Mittelwertsatz der eindimensionalen Differenzialrechnung 8.10 gilt in höheren Dimensionen nur für *skalare Funktionen.* Voraussetzung hierfür ist, dass mit zwei Punkten u und v auch die gesamte *Verbindungsstrecke*

$$[u,v] \coloneqq \{(1-t)u + tv : 0 \leqslant t \leqslant 1\}$$

zum Definitionsbereich gehört. Dann gilt der folgende

14 **Mittelwertsatz für skalare Funktionen** *Sei $f : V \hookrightarrow \mathbb{R}$ stetig differenzierbar. Gehört $[u,v]$ zum Definitionsbereich von f, so gilt*

$$f(v) - f(u) = Df(x)(v-u)$$

mit einem $x \in [u,v]$. $\rtimes$

⟨⟨⟨⟨⟨ Man bilde die Hilfsfunktion $\phi\colon\ [0,1] \to \mathbb{R}$ mit $\phi(t) = f((1-t)u + tv)$ und wende den klassischen Mittelwertsatz 8.10 an A-15 . ⟩⟩⟩⟩⟩

Ist der Zielraum dagegen höherdimensional, so gilt der Mittelwertsatz nicht mehr. Für die Kreiskurve $\gamma\colon\ t \mapsto (\cos(2\pi t), \sin(2\pi t))$ ist zum Beispiel

$$\gamma(1) - \gamma(0) = 0 \neq \dot{\gamma}(t), \qquad t \in [0,1],$$

da Sinus und Cosinus keine gemeinsamen Nullstellen besitzen. Es gibt aber eine etwas schwächere, ebenfalls sehr nützliche Identität, die in beliebigen Dimensionen gilt.

15 **Lemma von Hadamard** *Sei $f : V \hookrightarrow W$ stetig differenzierbar. Gehört $[u,v]$ zum Definitionsbereich von f, so gilt*

$$f(v) - f(u) = \Lambda(v-u)$$

mit der linearen Abbildung $\Lambda = \int_0^1 Df((1-t)u + tv)\,\mathrm{d}t\,.$ ⋊

Hierbei ist $t \mapsto Df((1-t)u+tv)$ eine stetige Kurve im Vektorraum $L(V,W)$, dessen Integral 10.7 ebenfalls ein Element von $L(V,W)$ ergibt. Im Standardfall ist dies eine von t abhängende $m \times n$-Matrix, deren Integral komponentenweise gebildet wird.

⟨⟨⟨⟨⟨ Betrachte die Streckenparametrisierung

$$\varphi : \quad [0,1] \to [u,v], \quad \varphi(t) = (1-t)u + tv,$$

mit Anfangspunkt u und Endpunkt v. Nach Voraussetzung ist $f \circ \varphi$ wohldefiniert und aufgrund der Kettenregel 7 stetig differenzierbar. Mit dem Hauptsatz für Kurven 13.6 ergibt sich

$$f(v) - f(u) = f \circ \varphi \Big|_0^1 = \int_0^1 (f \circ \varphi)^{\cdot}(t)\,\mathrm{d}t = \int_0^1 Df(\varphi(t))\dot{\varphi}(t)\,\mathrm{d}t.$$

Hierbei ist $\dot{\varphi}(t) = v - u$ unabhängig von t, so dass wir diesen Term *hinter* das Integral ziehen können A-10.29 [2]. Das ergibt die Behauptung. ⟩⟩⟩⟩⟩

Aus dem Lemma von Hadamard folgt nun der klassiche Schrankensatz auch in höheren Dimensionen.

16 **Schrankensatz** *Sei $f : V \hookrightarrow W$ stetig differenzierbar. Gehört $[u,v]$ zum Definitionsbereich von f, so gilt*

$$|f(v) - f(u)| \leqslant \max_{x \in [u,v]} \|Df(x)\|\,|v-u|$$

mit der durch die Vektorraumnormen induzierten Operatornorm $\|\cdot\|$. ⋊

⟨⟨⟨⟨⟨ Setze $\varphi(t) = (1-t)u + tv$ wie zuvor. Aufgrund des Hadamardschen Lemmas ist dann

$$|f(v) - f(u)| \leqslant \|\Lambda\|\,|v-u|$$

mit

$$\|\Lambda\| = \left\| \int_0^1 Df(\varphi(t))\,\mathrm{d}t \right\| \leqslant \int_0^1 \|Df(\varphi(t))\|\,\mathrm{d}t$$
$$\leqslant \max_{t \in [0,1]} \|Df(\varphi(t))\| = \max_{x \in [u,v]} \|Df(x)\|\,.$$ ⟩⟩⟩⟩⟩

[2] Wir dürfen $v - u$ nicht nach vorne ziehen, da es das Argument von Λ ist.

Korollar *Ist $f\colon V \hookrightarrow W$ von der Klasse C^1, so ist f lokal lipschitz.* ⋊

Dabei heißt eine Abbildung *lokal lipschitz*, wenn jeder Punkt ihres Definitionsbereiches eine Umgebung besitzt, auf der die Abbildung lipschitz ist. Die L-Konstanten dürfen dabei von der Umgebung abhängen.

⟨⟨⟨⟨⟨ Nach Voraussetzung ist $Df\colon V \hookrightarrow L(V,W)$ stetig. Also ist auch

$$\|Df\| : V \hookrightarrow \mathbb{R}, \quad x \mapsto \|Df(x)\|$$

stetig. Um jedem Punkt existiert daher eine Kugel B im Definitionsbereich von f, so dass

$$\sup_{x\in B} \|Df(x)\| = M < \infty.$$

Sind nun $u,v \in B$, so ist auch $[u,v] \subset B$ wegen der Konvexität jeder Kugel $_{\text{A-5.42}}$, und aufgrund des Schrankensatzes $_{16}$ gilt

$$|f(v) - f(u)| \leqslant M\,|v-u|\,.$$

Somit ist f auf B M-lipschitz. ⟩⟩⟩⟩⟩

14.5 Gradient

Für *skalare* Funktionen gibt es schließlich noch einen dritten Ableitungsbegriff, den des *Gradienten*. Ist $f\colon V \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so definiert die totale Ableitung

$$Df(a) : V \to \mathbb{R}, \quad h \mapsto Df(a)h$$

ein *lineares Funktional* auf V. Ist V nun ein *Hilbertraum* mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$, so existiert aufgrund des Rieszschen Darstellungssatzes $_3$ genau ein Vektor $\phi \in V$ mit der Eigenschaft, dass

$$Df(a)h = \langle \phi, h\rangle\,, \qquad h \in V.$$

Man sagt, *ϕ stellt $Df(a)$ bezüglich $\langle\cdot,\cdot\rangle$ dar.* Dieser Vektor ist der *Gradient* von f im Punkt a.

Definition *Sei V ein Hilbertraum. Ist $f\colon V \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so ist der Gradient von f an der Stelle a der eindeutig bestimmte und mit $\nabla f(a)$ bezeichnete Vektor in V mit der Eigenschaft, dass*

$$Df(a) = \langle \nabla f(a), \cdot\rangle\,.$$ ⋊

14.22

17 ⊳ A. Für jedes $v \in V$ ist

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

differenzierbar mit $DL_v(x)h = \langle v, h \rangle$. Also ist

$$\nabla L_v = v$$

in jedem Punkt von V.

B. Ist $A : V \to V$ linear und symmetrisch, so ist

$$f : V \to \mathbb{R}, \quad f(x) = \langle Ax, x \rangle$$

differenzierbar mit Ableitung $Df(x)h = 2\langle Ax, h \rangle$. Somit ist

$$\nabla f(x) = 2Ax, \qquad x \in V. \quad ⊲$$

Der Gradient ist also die *Darstellung* der totalen Ableitung einer skalaren Funktion bezüglich eines Skalarproduktes. Ohne Bezug auf ein Skalarprodukt macht es keinen Sinn, von einem Gradienten zu sprechen. Besonders wichtig ist natürlich der Standardraum $\mathbb{R}^n$ mit dem Standardskalarprodukt.

Satz *Im Standardraum $\mathbb{R}^n$ mit dem Standardskalarprodukt ist der Gradient einer differenzierbaren Funktion $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ der Spaltenvektor*

$$\nabla f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix} = Df^\top. \quad ⋊$$

⟨⟨⟨⟨ Die j-te Komponente von ∇f ist ja

$$\langle \nabla f, e_j \rangle = Df e_j = \partial_j f,$$

während $Df = (\partial_1 f, .., \partial_n f)$. ⟩⟩⟩⟩

Der Gradient ∇f ist also ein *Spaltenvektor*, während die Ableitung Df einer skalaren Funktion durch einen *Zeilenvektor* dargestellt wird. Dies ist ein wichtiger Unterschied, der zum Beispiel Folgen hat für deren Verhalten unter Koordinatentransformationen.

Bemerkung Das Symbol ∇ selbst wird *Nablaoperator* genannt und bezeichnet in der klassischen Vektoranalysis den vektoriellen Differenzialoperator

$$\nabla = \begin{pmatrix} \partial_1 \\ \vdots \\ \partial_n \end{pmatrix}.$$

Der *Gradient* von f kann damit – rein formal interpretiert – verstanden werden als das Produkt des Vektors ∇ mit dem Skalar f:

$$\nabla f = \begin{pmatrix} \partial_1 \\ \vdots \\ \partial_n \end{pmatrix} f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix}.$$

Skalarprodukte und Kreuzprodukte von ∇ mit Vektorfunktionen sind ebenfalls erklärt, sie werden uns später – im dritten Band – als *Rotation* und *Divergenz* begegnen. Das Skalarprodukt von ∇ mit sich selbst ergibt den *Laplaceoperator*

$$\nabla \bullet \nabla = \sum_{j=1}^{n} \partial_j^2 =: \Delta. \qquad \multimap$$

▪ Richtung des steilsten An- und Abstiegs

Sei $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ um den Punkt a definiert. Für jeden Einheitsvektor $e \in \mathbb{R}^n$ ist dann

$$f_e : \mathbb{R} \hookrightarrow \mathbb{R}, \quad t \mapsto f(a + te)$$

in einer Umgebung von 0 definiert und stetig differenzierbar. Die Ableitung $f_e'(0)$ beschreibt dann die *Anstiegsrate* von f im Punkt a in Richtung e. Wird diese für einen Vektor e *maximal* respektive minimal, so nennen wir e eine *Richtung des steilsten Anstiegs* respektive eine *Richtung des steilsten Abstiegs*. Diese Richtungen müssen existieren, da die Einheitssphäre im $\mathbb{R}^n$ kompakt ist.

Satz *Sei $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar. Ist $\nabla f(a) \neq 0$, so bezeichnen $\nabla f(a)$ und $-\nabla f(a)$ die in diesem Falle eindeutigen Richtungen des steilsten An- respektive Abstiegs von f im Punkt a.* ⋊

⟨⟨⟨⟨ Es ist

$$f_e'(0) = f(a + te)^{\cdot}|_{t=0} = Df(a)e = \langle \nabla f(a), e \rangle.$$

Aufgrund der Cauchy-Schwarzschen Ungleichung gilt

$$|f_e'(0)| = |\langle \nabla f(a), e \rangle| \leqslant |\nabla f(a)|\,|e| = |\nabla f(a)|.$$

Ist $\nabla f(a) \neq 0$, so tritt Gleichheit hierbei *genau dann* ein, wenn die Vektoren $\nabla f(a)$ und e kollinear sind$_{5.32}$. Somit wird $f_e'(0)$ maximal genau für $e \upuparrows \nabla f(a)$, und minimal genau für $e \updownarrows \nabla f(a)$. In allen anderen Richtungen ist die Ableitung nicht extremal. Das ist gerade die Behauptung. ⟩⟩⟩⟩

Zu beachten ist, dass ∇f in die Richtung des steilsten Anstiegs im *Definitionsbereich von f* weist, und nicht am Graphen von f.

14.6 Höhere Ableitungen

Wir betrachten nun höhere partielle Ableitungen. Existiert die partielle Ableitung einer Abbildung $f: \mathbb{R}^n \hookrightarrow W$ nach einer Koordinate x_k auf dem ganzen Definitionsgebiet von f, so erhalten wir wieder eine Abbildung $\partial_k f: \mathbb{R}^n \hookrightarrow W$. Man kann also eventuell ein weiteres Mal partiell differenzieren, sagen wir nach x_l. Dann erhalten wir eine *zweite partielle Ableitung*

$$\partial_l \partial_k f \coloneqq \partial_l(\partial_k f) = (f_{x_k})_{x_l} = f_{x_k x_l} = \frac{\partial^2 f}{\partial x_k \partial x_l} : \mathbb{R}^n \hookrightarrow W,$$

wobei wir hier verschiedene Schreibweisen für dieselbe Sache aufführen. Und so weiter ... Zum Beispiel ist

$$\partial_m \partial_l \partial_k f = \partial_m(\partial_l \partial_k f) = f_{x_k x_l x_m} = \frac{\partial^3 f}{\partial x_k \partial x_l \partial x_m}.$$

Definition *Für eine Abbildung $f: \mathbb{R}^n \hookrightarrow W$ ist die r-te partielle Ableitung*

$$\partial_{k_r} .. \partial_{k_2} \partial_{k_1} f = f_{x_{k_1} x_{k_2} .. x_{k_r}} = \frac{\partial^r f}{\partial x_{k_1} .. \partial x_{k_r}}$$

rekursiv erklärt durch

$$\partial_{k_r} .. \partial_{k_2} \partial_{k_1} f \coloneqq \partial_{k_r}(\partial_{k_{r-1}} .. \partial_{k_1} f),$$

sofern alle Zwischenableitungen existieren. ⋊

▹ Für $f: \mathbb{R}^3 \to \mathbb{R}$ mit $f(x, y, z) = x^2 y^4 \sin z$ ist beispielsweise

$$\begin{aligned} f_x &= 2xy^4 \sin z, & f_z &= x^2 y^4 \cos z, \\ f_{xy} &= 8xy^3 \sin z, & f_{zy} &= 4x^2 y^3 \cos z, \\ f_{xyz} &= 8xy^3 \cos z, & f_{zyx} &= 8xy^3 \cos z. \end{aligned}$$ ◃

In diesem Beispiel kommt es auf die Reihenfolge der partiellen Ableitungen *nicht an*, es ist $f_{xyz} = f_{zyx}$. Das ist aber nicht immer so.

▹ Betrachte

$$f: \mathbb{R}^2 \to \mathbb{R}, \quad f(x, y) \mathrel{\hat{=}} \frac{x^3 y - xy^3}{x^2 + y^2}.$$

Wegen $f(0, \cdot) \equiv 0$ und $f(\cdot, 0) \equiv 0$ gilt

$$f_x(0, y) = \lim_{h \to 0} \frac{f(h, y)}{h} = -y, \qquad f_y(x, 0) = \lim_{h \to 0} \frac{f(x, h)}{h} = x.$$

Somit ist

$$f_{xy}(0, y) = -1, \qquad f_{yx}(x, 0) = 1,$$

für alle $x, y \in \mathbb{R}$ und damit $f_{xy}(0, 0) \neq f_{yx}(0, 0)$. ◃

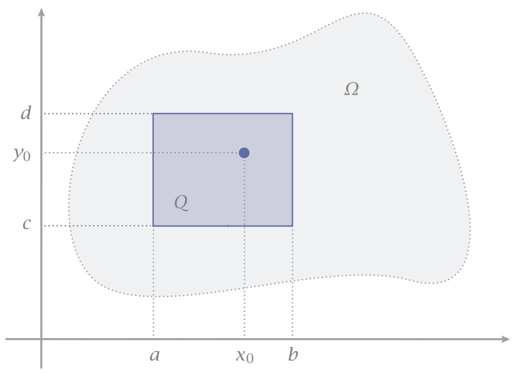


Abb 5

Zum Lemma von Schwarz

Zum Glück reicht die *Stetigkeit* der partiellen Ableitungen, um sie unabhängig von deren Reihenfolge zu machen. Die Quintessenz ist das

18 **Lemma von Schwarz** *Sei* $f: \mathbb{R}^n \supset \Omega \hookrightarrow \mathbb{R}^m$ *stetig differenzierbar, und seien* x *und* y *zwei beliebige Koordinaten in* Ω*. Existiert die zweite Ableitung* f_{xy} *auf* Ω *und ist sie dort stetig, so existiert auch* f_{yx} *auf* Ω*, und es gilt*

$$f_{xy} = f_{yx}. \quad \rtimes$$

⟪⟪ Da nur die beiden Koordinaten x und y involviert sind und alle anderen fixiert werden können, beschränken wir uns auf den Fall $\Omega \subset \mathbb{R}^2$. Fixiere einen Punkt $(x_0, y_0) \in \Omega$ und wähle Intervalle $[a,b]$ um x_0 und $[c,d]$ um y_0 so, dass

$$Q \coloneqq [a,b] \times [c,d] \subset \Omega.$$

Da f stetig differenzierbar in x ist, gilt

$$f(x,y) - f(a,y) = \int_a^x f_x(t,y)\,\mathrm{d}t, \qquad (x,y) \in Q.$$

Nach Voraussetzung ist f_{xy} auf Q stetig. Aufgrund des anschließend bewiesenen Lemmas $_{19}$ definiert das Integral daher eine in y differenzierbare Funktion, deren Ableitung man durch Differenzieren ›unter dem Integral‹ erhält. Es gilt also

$$f_y(x,y) - f_y(a,y) = \int_a^x f_{xy}(t,y)\,\mathrm{d}t.$$

Ebenfalls wegen der Stetigkeit von f_{xy} definiert dieses Integral für festes y eine nach x differenzierbare Funktion. Somit ist auch $f_y(x,y)$ nach x differenzierbar, und es gilt

$$\partial_x f_y(x,y) = \frac{\partial}{\partial x} \int_a^x f_{xy}(t,y)\,\mathrm{d}t = f_{xy}(x,y).$$

Das war zu zeigen. ⟫⟫

Das folgende Lemma macht eine Aussage darüber, wann ein sogenanntes *parameterabhängiges Integral* eine differenzierbare Funktion definiert, deren Ableitung man durch ›Differenziation unter dem Integral‹ erhält.

19 **Lemma** *Sei $Q = [a,b] \times [c,d]$ mit Koordinaten (x,y). Ist $f\colon Q \to \mathbb{R}$ stetig, nach y partiell differenzierbar und f_y ebenfalls stetig auf Q, so ist auch die Funktion*

$$\phi\colon\ [c,d] \to \mathbb{R}, \quad \phi(y) = \int_a^b f(t,y)\,\mathrm{d}t$$

differenzierbar, und es gilt

$$\phi'(y) = \int_a^b f_y(t,y)\,\mathrm{d}t. \qquad ⋊$$

⟨⟨⟨⟨⟨ Sei y ein innerer Punkt von $[c,d]$ — Randwerte gehen analog. Aufgrund des Hauptsatzes der Differenzial- und Integralrechnung $_{10.16}$ ist

$$\begin{aligned}\phi(y+h) - \phi(y) &= \int_a^b f(t,y+sh)\Big|_0^1\,\mathrm{d}t\\ &= h\int_a^b \left(\int_0^1 f_y(t,y+sh)\,\mathrm{d}s\right)\mathrm{d}t.\end{aligned}$$

Damit folgt

$$\begin{aligned}&\frac{1}{h}\left(\phi(y+h) - \phi(y)\right) - \int_a^b f_y(t,y)\,\mathrm{d}t\\ &\quad = \int_a^b\int_0^1 f_y(t,y+sh)\,\mathrm{d}s\,\mathrm{d}t - \int_a^b\int_0^1 f_y(t,y)\,\mathrm{d}s\,\mathrm{d}t\\ &\quad = \int_a^b\int_0^1 \left(f_y(t,y+sh) - f_y(t,y)\right)\mathrm{d}s\,\mathrm{d}t.\end{aligned}$$

Wegen der Kompaktheit von Q ist f_y auf Q *gleichmäßig stetig* $_{7.32}$. Zu jedem $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass für *alle* $t \in [a,b]$ und *alle* $s \in [0,1]$

$$|f_y(t,y+sh) - f_y(t,y)| < \varepsilon, \qquad |h| < \delta.$$

Für diese h erhalten wir

$$\begin{aligned}&\left|\frac{1}{h}\left(\phi(y+h) - \phi(y)\right) - \int_a^b f_y(t,y)\,\mathrm{d}t\right|\\ &\quad \leqslant \int_a^b\int_0^1 |f_y(t,y+sh) - f_y(t,y)|\,\mathrm{d}s\,\mathrm{d}t\\ &\quad\quad \leqslant \int_a^b\int_0^1 \varepsilon\,\mathrm{d}s\,\mathrm{d}t = \varepsilon(b-a).\end{aligned}$$

Also gilt

$$\lim_{h\to 0}\frac{\phi(y+h) - \phi(y)}{h} = \int_a^b f_y(t,y)\,\mathrm{d}t. \qquad ⟩⟩⟩⟩⟩$$

Ein entsprechender Satz gilt für die Vertauschbarkeit höherer partieller Ableitungen, wenn die entsprechenden Stetigkeitsbedingungen erfüllt sind. Auch aus diesem Grund sind Funktionen mit *stetigen* Ableitungen so wichtig.

Definition *Sei $\Omega \subset \mathbb{R}^n$ offen und $r \geqslant 1$. Dann bezeichnet $C^r(\Omega, \mathbb{R}^m)$ den Raum aller Abbildungen $f : \Omega \to \mathbb{R}^m$, die auf Ω sämtliche partiellen Ableitungen bis zur Ordnung r besitzen und diese dort auch stetig sind. Eine Abbildung $f \in C^r(\Omega, \mathbb{R}^m)$ heißt von der Klasse C^r oder C^r-Abbildung.* ⋊

Weiter setzt man

$$C^\infty(\Omega, \mathbb{R}^m) := \bigcap_{r \geqslant 1} C^r(\Omega, \mathbb{R}^m),$$

den Raum der unendlich oft partiell differenzierbaren Abbildungen $\Omega \to \mathbb{R}^m$. Diese werden auch *glatt*Alle diese Räume sind *lineare Vektorräume*. Im Fall skalarer Funktionen schreibt man noch kürzer

$$C^r(\Omega) := C^r(\Omega, \mathbb{R}).$$

Satz *Ist $f \in C^r(\Omega, \mathbb{R}^m)$, so ist jede partielle Ableitung von f der Ordnung r unabhängig von der Reihenfolge der Differenziationen.* ⋊

▪ Totale Ableitungen

Wir haben partielle Ableitungen höherer Ordnung im Standardfall definiert, jedoch nicht die entsprechenden totalen Ableitungen $D^2 f$, $D^3 f$, ... *Dies werden wir hier auch nicht tun* – weil wir es nicht unmittelbar benötigen, und weil dies auch konzeptionell komplizierter ist.

Im Prinzip geht es um Folgendes. Ist $f : V \hookrightarrow W$ total differenzierbar, so ist

$$Df : V \hookrightarrow L(V, W).$$

Nun ist der Zielraum wieder ein Banachraum. Also ist auch die zweite totale Ableitung aufgrund unserer allgemeinen Definition erklärt – wenn sie existiert –, und es ist

$$D^2 f = D(Df) : V \hookrightarrow L(V, L(V, W)) \simeq L(V \times V, W).$$

Somit ist $D^2 f$ eine *bilineare Abbildung* von V nach W.

Und so weiter ... Die r-te totale Ableitung von f kann identifiziert werden mit einer Abbildung

$$D^r f : V \hookrightarrow L(V \times \cdot\cdot \times V, W),$$

die jedem Punkt im Definitionsbereich eine r-lineare Abbildung von V nach W zuordnet. Auf die Details verzichten wir hier.