2

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 $\mathbb{R} \to \mathbb{R}$. In Kapitel 1 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 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 in keiner sinnvollen Weise definierbar ist – 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. Siehe dazu ›Noch mehr Analysis‹.

J. Pöschel, *Etwas mehr Analysis*, DOI 10.1007/978-3-658-05860-9_2,

2.1 Elemente der Linearen Algebra

▪ Beschränkte lineare Abbildungen

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

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

$$\|A\|_{V,W} \coloneqq \sup_{0 \neq x \in V} \frac{\|Ax\|_W}{\|x\|_V} < \infty. \qquad \rtimes$$

Aufgrund der positiven Homogenität jeder Norm ist auch

$$\|A\|_{V,W} = \sup_{\|x\|_V = 1} \|Ax\|_W .$$

Offensichtlich gilt *immer*

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

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

1 **Satz** *Für einen linearen Operator $A\colon V \to W$ 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

$$\|Ax\| < 1, \qquad \|x\| < \delta.$$

Sei $x \neq 0$. Für $\mu = \theta\delta / \|x\|$ mit $0 < \theta < 1$ gilt dann

$$\frac{\|Ax\|}{\|x\|} = \frac{\|A(\mu x)\|}{\|\mu x\|} \leqslant \frac{1}{\|\mu x\|} = \frac{1}{\theta\delta}.$$

Da dies für alle $0 < \theta < 1$ und auch für alle $x \neq 0$ gilt, folgt

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \leqslant \frac{1}{\delta}.$$

(iv) ⇒ (i) Auch klar, denn $\|Au - Av\| = \|A(u - v)\| \leqslant \|A\| \, \|u - v\|$. ⟫⟫

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*

$$\|A\|_{V,W} = \sup_{\|x\|_V=1} \|Ax\|_W$$

eine Norm, die sogenannte Operatornorm. Mit ihr wird $L(V,W)$ zu einem Banachraum. ⋊

⟪⟪ 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 $\|A_kx - A_lx\| \leqslant \|A_k - A_l\|\,\|x\|$ dann (A_kx) eine Cauchyfolge in W. Aufgrund der Vollständigkeit von W konvergiert also (A_kx), und wir können eine Abbildung $A\colon V \to W$ punktweise definieren durch

$$Ax \coloneqq \lim_{k\to\infty} A_kx.$$

Diese Abbildung ist linear, denn

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

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

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

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

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

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

$$V^* = \{L : V \to \mathbb{R} \text{ ist linear und stetig}\}$$

für jeden *normierten* Vektorraum V ein Banachraum. ⊸

▪ Hilberträume

Unter den Banachräumen spielen die Hilberträume eine besondere Rolle. Diese sind, wie wir bereits in Abschnitt 1.5.7 gesehen haben, charakterisiert durch die Existenz eines *Skalarprodukts*, also einer positiv definiten und symmetrischen Bilinearform

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

so dass die Norm gegeben ist durch 1.5.33 $\|x\| := \sqrt{\langle x, x \rangle}$.

Für ein Skalarprodukt gilt immer die Cauchy-Schwarzsche Ungleichung 1.5.32

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

Eine Konsequenz ist, dass jeder Vektor $v \in V$ ein stetiges lineares Funktional

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

definiert, denn wegen $|L_v x| \leqslant \|v\| \, \|x\|$ ist dieses 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 stetigen linearen Funktional $L \in V^*$ ein eindeutiger Vektor $v \in V$, so dass*

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

⟪⟪⟪ Wir können annehmen, dass $\|L\| = 1$. Dann existiert in V eine Folge (v_k) mit $\|v_k\| = 1$ und

$$0 < Lv_k \leqslant 1, \qquad Lv_k \to 1.$$

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

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

Mit der Parallelogrammgleichung A-1.5.33 folgt hiermit

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

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

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

Wir behaupten, dass

$$Lx = \langle v, x\rangle\,, \qquad x \in V.$$

Wir können annehmen, dass $Lx > 0$. Wegen $Lv = 1 = \|v\|$ gilt für $t > 0$

$$Lx = t^{-1}\,\{L(v + tx) - L(v)\} \leqslant t^{-1}\,(\|v + tx\| - \|v\|)$$

und

$$Lx = -t^{-1}\,\{L(v - tx) - L(v)\} \geqslant -t^{-1}\,(\|v - tx\| - \|v\|)\,.$$

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 1.11.1 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 $Lx = \langle v, x\rangle$. »»»

▪ 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-4. Dies ist auf einem endlich-dimensionalen Raum *nicht möglich*, da die Einheitssphäre dort kompakt ist 1.7.37, 1.7.39. Außerdem lassen sich lineare Abbildungen zwischen endlich-dimensionalen Räumen bequem durch *Matrizen* darstellen. Aus diesen Gründen werden wir uns auf lineare Abbildungen zwischen endlich-dimensionale Räumen beschränken.

Zunächst ein Wort zur Notation. Im Matrizenkalkül ist es üblich, Vektoren als *Spaltenvektoren* zu schreiben, also als

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Dabei verzichten wir auf einen Vektorpfeil oder sonstige Auszeichnungen. In einem horizontal laufenden Text ist dies natürlich platzraubend. Deshalb verwenden wir die Schreibweise

$$x = (x_1, \dots, x_n)^\top,$$

wobei $^\top$ die Transposition bezeichnet, wie sie für Matrizen erklärt ist. Dementsprechend ist $x^\top = (x_1, \dots, x_n)$ 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 endlicher Dimension. Sind Basisvektoren $v_1, \dots, v_n$ in V und $w_1, \dots, w_m$ in W gewählt, so wird A 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

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

die *Matrixdarstellung* von A bezüglich der gewählten Basen in V und W. Deren j-te Spalte 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

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

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.

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

$$x = \sum_{i=1}^{n} x_i v_i, \qquad y = \sum_{j=1}^{n} y_j v_j$$

wird ja

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

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 [2]

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

wodurch $e_1, .., e_n$ zu einer *Orthonormalbasis* des $\mathbb{R}^n$ wird. Die Koeffizienten des 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

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

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*.

2.2 Totale Ableitung

Für die folgenden Definitionen kommt es auf die Dimensionen der Räume nicht an. Somit bezeichnen V und W zunächst beliebige Banachräume. Deren Normen $|\cdot|_V$ und $|\cdot|_W$ bezeichnen wir meist einfach mit $|\cdot|$, wenn der Bezug aus dem Zusammenhang klar ist.

[2] Das hier erklärte Symbol δ_{ij} wird als *Kronecker-Delta* bezeichnet.

Wir betrachten Abbildungen, die eine nichtleere offene Teilmenge von V nach W abbilden. Da es zunächst nur um lokale Aussagen geht, ist eine explizite Bezeichnung für diese offene Menge nicht erforderlich. Wir schreiben daher

$$f : V \hookrightarrow W,$$

wenn es eine nichtleere offene Teilmenge $\Omega \subset V$ gibt, so dass $f : \Omega \to W$. Zu jedem Punkt a im Definitionsbereich von f existiert also eine offene Kugel

$$B_r(a) := \{x \in V : |x - a| < r\},$$

auf der f definiert ist. Dies ist die für uns zur Zeit wesentliche Eigenschaft des Definitionsbereiches einer Abbildung.

4 **Definition** *Eine Abbildung $f : V \hookrightarrow W$ heißt im Punkt a total differenzierbar oder kurz differenzierbar, wenn es eine stetige lineare Abbildung $L : V \to W$ 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 $_7$.

b. Diese Definition beinhaltet unsere früheren Definitionen der Differenzierbarkeit. Im Fall *einer* unabhängigen Variablen, also $\dim V = 1$, ist f eine Kurve in W. Eine lineare Abbildung $L : \mathbb{R} \to W$ wird durch einen Vektor $v \in W$ dargestellt, und wir erhalten die Charakterisierung der Ableitung entsprechend dem zweiten Differenzierbarkeitssatz $_{1.3}$. Ist auch $\dim W = 1$, so wird dieser Vektor durch eine reelle Zahl dargestellt, und wir erhalten die Ableitung einer reellen Funktion einer Variable wie im ersten Differenzierbarkeitssatz $_{1.9.1}$.

c. Sind V und W von endlicher Dimension, so ist jede lineare Abbildung $A : V \to W$ auch *stetig* und dieses Attribut somit redundant. Dies ist der einzige, aber auch wesentliche Unterschied zwischen dem endlich- und unendlich-dimensionalen Fall. ⊸

Wesentlich für die Existenz der Ableitung ist, dass der Approximationsfehler $|f(a+h) - f(a) - Lh|$ zwischen der im Allgemeinen nichtlinearen Abbildung $h \mapsto f(a+h)$ und der affinen Abbildung $h \mapsto f(a) + Lh$ schneller gegen Null konvergiert als $|h|$. Um solche unterschiedlichen Konvergenzgeschwindigkeiten bequem zum Ausdruck zu bringen, haben sich die *Landausymbole* bewährt.

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

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

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 $h \to 0$. ⋊

Bemerkungen *a.* Es ist also f groß-O von g auf Ω, wenn es eine Konstante M gibt, so dass

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

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

$$|f(h)| \leqslant \varepsilon\,|g(h)|, \qquad h \in B_\delta(0) \cap \Omega.$$

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

5 **Rechenregeln** *Für das Rechnen mit den Landausymbolen gilt:*

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

⟪⟪ Der Beweis ist als Übung überlassen. ⟫⟫

6 ▷ A. Für $t \in \mathbb{R}$ ist $\sin t = O(t)$ und $1 - \cos t = O(t^2)$.

B. Für $f \in C^{n+1}(I)$ ist aufgrund der Restgliedformel von Lagrange

$$f(a+h) = f(a) + f'(a)h + \ldots + \frac{1}{n!} f^{(n)}(a)h^n + O(h^{n+1}).$$

C. Ist $\langle\cdot,\cdot\rangle$ eine bilineare Form auf V und $A \in L(V)$, so ist

$$\langle Ah, h\rangle = O(|h|^2) = o(h). \qquad ◁$$

Mit den Landausymbolen können wir Differenzierbarkeit in einem Punkt wie folgt charakterisieren.

7 **Satz** *Eine Abbildung $f : V \hookrightarrow W$ ist im Punkt a differenzierbar genau dann, wenn*

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

mit einem $L \in L(V,W)$. In diesem Fall ist f im Punkt a auch stetig, und die Ableitung $Df(a) = L$ ist eindeutig bestimmt. ⋊

Somit ist eine Abbildung $f : V \hookrightarrow W$ in a *differenzierbar*, wenn sie dort *linear approximierbar* ist.

⟪⟪ Die erste Behauptung folgt aus den Definitionen der Differenzierbarkeit und des Landausymbols. Die Stetigkeit im Punkt a ist ebenso klar. 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-3. ⟫⟫

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 $Df : V \hookrightarrow W$ ist wieder eine Abbildung des gleichen Typs 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.

8 ⊳ *Affine Abbildung* 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 verschwindet also identisch. Die Ableitung ist damit $Df(a) = A$ in jedem Punkt a, also dieselbe lineare Abbildung A. Die Ableitung als *Abbildung*

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

ist somit konstant. ⊲

9 ▷▶ *Quadratische Form* Sei $\langle\cdot,\cdot\rangle$ eine bilineare Form auf V und $A \in L(V)$ bezüglich dieser Form symmetrisch, also $\langle Ax, y\rangle = \langle x, Ay\rangle$. Dann ist die *quadratische Form*

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

auf V differenzierbar. Denn es ist $\langle x, Ah\rangle = \langle Ax, h\rangle$ und damit

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

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

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

▪ Rechenregeln

Abbildungen in ein und denselben Vektorraum W können wir linear kombinieren. Diese Operation vertauscht mit der Differenziation, wie man leicht verifiziert.

10 **Satz** *Sind die Abbildungen* $f, g: V \hookrightarrow W$ *im Punkt* a *differenzierbar, so auch* $\lambda f + \mu g: V \hookrightarrow W$ *für alle* $\lambda, \mu \in \mathbb{R}$*, und es gilt*

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

Als nächstes betrachten wir die Verknüpfung zweier Abbildungen. Das Ergebnis ist die Verallgemeinerung der klassischen Kettenregel $_{1.9.8}$.

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

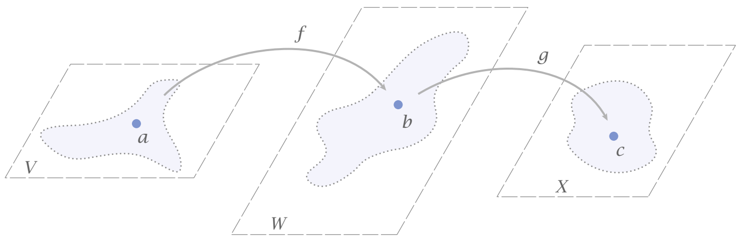
$$D(g \circ f)(a) = Dg(f(a))Df(a). \quad \rtimes$$

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

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

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.

Abb 1 **Zur Kettenregel**



Der Beweis ist praktisch identisch zum eindimensionalen Fall [1.9.8]. Aufgrund der Voraussetzungen ist f in einer Umgebung von a und g in einer Umgebung von $b = f(a)$ definiert. Für alle hinreichend kleinen h und k gilt somit

$$\begin{aligned} f(a+h) &= f(a) + Ah + o(h), \\ g(b+k) &= g(b) + Bk + o(k), \end{aligned}$$

mit $A = Df(a)$ und $B = Dg(b)$.

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

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

Nun zeigt man noch, dass $Bo(h) = o(h)$ und $o(Ah + o(h)) = o(h)$. Also ist

$$(g \circ f)(a+h)) = (g \circ f)(a) + BAh + o(h).$$

Somit [7] ist $g \circ f$ im Punkt a differenzierbar, und die Ableitung ist BA, was der Behauptung entspricht.

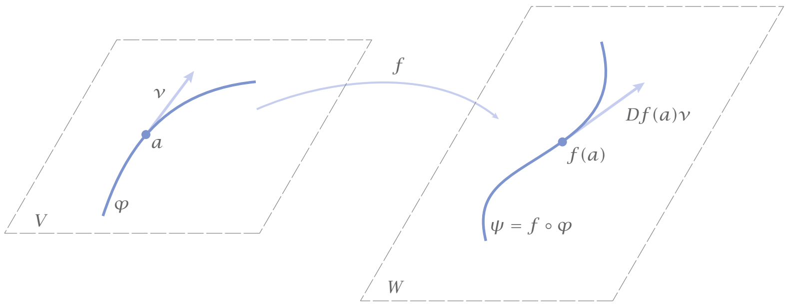
12 *Lineare Abbildung* Ist $f : V \hookrightarrow W$ differenzierbar und $\Lambda : W \to X$ eine lineare Abbildung, so ist auch

$$\Lambda f : V \hookrightarrow X$$

differenzierbar, und für jeden Punkt a im Definitionsbereich ist [8]

$$D(\Lambda f)(a) = (D\Lambda)(f(a))\,Df(a) = \Lambda Df(a).$$

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



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

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

eine differenzierbare Kurve in W mit Ableitung

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \varphi)(t) = Df(\varphi(t))\dot{\varphi}(t), \qquad t \in I.$$

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

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \varphi)(t)\bigg|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} f(a + tv)\bigg|_{t=0} = Df(a)v. \qquad ⊲$$

14 ⊳ *Quadratische Form entlang einer Kurve* Ist $\varphi : I \to V$ eine differenzierbare Kurve in V und

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

eine quadratische Form auf V [9], so ist

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

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 $\varphi(t) = a + tv$ ist $\Phi'(0) = 2\,\langle Aa, v \rangle$. ⊲

Es sei nochmals betont, dass alles bisher Gesagte unabhängig von der Dimension der beteiligten Räume gilt. Die einzige Forderung ist, dass alle beteiligten linearen Abbildungen auch beschränkt sind.

2.3 Richtungsableitungen

Neben dem Begriff der totalen Ableitung gibt es noch einen schwächeren Ableitungsbegriff, den der *Richtungsableitung*. Ist a ein Punkt im Definitionsbereich von $f : V \to W$ und $0 \neq v \in V$, so ist die Abbildung

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

in einer Umgebung von 0 wohldefiniert. Für Abbildungen dieser Art haben wir bereits den Begriff der Ableitung im vorangehenden Kapitel über Kurven erklärt, ohne Rückgriff auf die totale Ableitung.

Definition *Sei $f : V \to W$ im Punkt a definiert und $0 \neq v \in V$. Dann heißt*

$$\partial_v f(a) := \frac{\mathrm{d}}{\mathrm{d}t} f(a + tv)\Big|_{t=0},$$

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

▷ A. Für eine affine Abbildung $f : x \mapsto Ax + b$ ist

$$\partial_v f(u) = \frac{\mathrm{d}}{\mathrm{d}t}(A(u + tv) + b) = Av.$$

B. Für eine quadratische Form $f : x \mapsto \langle Ax, x \rangle$ ist

$$\partial_v f(u) = \frac{\mathrm{d}}{\mathrm{d}t} \langle A(u + tv), u + tv) \rangle = 2 \langle Av, u \rangle .$$ ◁

Wie der folgende Satz zeigt, impliziert totale Differenzierbarkeit die Existenz aller Richtungsableitungen.

15 **Satz** *Ist $f : V \to W$ im Punkt a differenzierbar, so existieren dort auch alle Richtungsableitungen, und es gilt*

$$\partial_v f(a) = Df(a)v. \quad ⋊$$

⟨⟨⟨⟨⟨ Da die Gerade $t \mapsto a + tv$ in $t = 0$ differenzierbar ist, ist die Kettenregel 11 anwendbar und ergibt

$$\partial_v f(a) = \frac{\mathrm{d}}{\mathrm{d}t} f(a + tv)\Big|_{t=0} = Df(a)\, \frac{\mathrm{d}}{\mathrm{d}t}(a + tv)\Big|_{t=0} = Df(a)v.$$ ⟩⟩⟩⟩⟩

Tatsächlich ist Richtungsdifferenzierbarkeit ein *schwächerer* Ableitungsbegriff als totale Differenzierbarkeit, wie das folgende Beispiel zeigt.

16 ⊳ Betrachte

$$f: \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{x^2 y}{x^2+y^2}, & (x,y) \neq 0, \\ 0 \quad, & (x,y) = 0. \end{cases}$$

Für jeden Vektor $0 \neq v = (\xi, \eta)$ gilt $f(tv) = tf(v)$, wie man leicht nachrechnet. Hieraus folgt

$$\partial_v f(0) = \lim_{t \to 0} f(tv)/t = f(v).$$

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

Die Ableitungen in Richtung der Vektoren einer fixierten Basis haben einen eigenen Namen.

Definition *Sei $f: V \hookrightarrow W$ im Punkt a definiert und $v_1, .., v_n$ eine Basis von V. Dann heißt*

$$\partial_j f(a) \coloneqq \frac{\mathrm{d}}{\mathrm{d}t} f(a + te_j)\Big|_{t=0} = \partial_{e_j} f(a),$$

falls diese Ableitung existiert, die j-te partielle Ableitung von f im Punkt a bezüglich dieser Basis. ⋊

Der wichtigste Fall ist natürlich der Standardraum $\mathbb{R}^n$ mit der Standardbasis $e_1, .., e_n$. Dann haben wir es mit einer Abbildung

$$f: \mathbb{R}^n \hookrightarrow W, \quad f(x) = f(x_1, .., x_n)$$

zu tun. In diesem Fall ist

$$\partial_j f(a) = \frac{\mathrm{d}}{\mathrm{d}t} f(a + te_j)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} f(a_1, .., a_j + t, .., a_n)\Big|_{t=0}.$$

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

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

Die letzte Schreibweise ist nützlich, wenn f selber noch einen Index besitzt, wie im nächsten Satz.

Wir bemerken noch, dass die Existenz der partiellen Ableitungen in einem Punkt *nicht* die Stetigkeit der Abbildung an dieser Stelle impliziert.

17 ⊳ Gegeben sei

$$f: \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{2xy}{x^2+y^2}, & (x,y) \neq 0, \\ 0, & (x,y) = 0. \end{cases}$$

Betrachten wir die Funktion nur auf der x- und y-Achse, so ist

$$f(t,0) = 0 = f(0,t), \qquad t \in \mathbb{R}.$$

Die Funktion ist dort also identisch Null. Somit existieren im Nullpunkt beide partiellen Ableitungen, und es ist

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

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 der Geraden ab und nimmt alle Werte im Intervall $[-1,1]$ an. Also ist f im Nullpunkt nicht stetig und damit auch nicht differenzierbar. ⊲

▪ Jacobimatrix

Bisher haben wir die totale Ableitung einer Abbildung als lineare Abbildung betrachtet, ohne Bezug auf explizite Koordinaten zu nehmen. Nun untersuchen wir die Frage, wie sich diese Ableitung im Standardfall konkret darstellt. Die Abbildung selbst hat dann die Gestalt

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

— Zunächst betrachten wir die Richtungsableitungen.

18 **Satz** *Ist $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ im Punkt a total differenzierbar, so ist die Richtungsableitung in Richtung von $v = (v_1, .., v_n)$ gegeben durch*

$$\partial_v f(a) = \sum_{j=1}^n \partial_j f(a) v_j.$$

Dabei sind die partiellen Ableitungen von f gegeben durch die partiellen Ableitungen seiner Komponentenfunktionen,

$$\partial_j f(a) = \partial_j \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} (a) = \begin{pmatrix} \partial_j f_1(a) \\ \vdots \\ \partial_j f_m(a) \end{pmatrix} = \begin{pmatrix} f_{1,j}(a) \\ \vdots \\ f_{m,j}(a) \end{pmatrix}. \qquad \rtimes$$

⟪⟪ Mit $v = \sum_{j=1}^n v_j e_j$ wird 15

$$\partial_v f(a) = Df(a)v = \sum_{j=1}^n v_j Df(a) e_j = \sum_{j=1}^n \partial_j f(a) v_j.$$

Dies ist die erste Behauptung. Die zweite Behauptung folgt aus dem entsprechenden Ergebnis für die Ableitung von Kurven 1.6 . ⟫⟫

Nun das Ergebnis für die totale Ableitung.

19 **Satz** *Ist $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ im Punkt a total differenzierbar, so existieren auch sämtliche partiellen Ableitungen von f in a, und die totale Ableitung $Df(a)$ wird dargestellt durch die Jacobi- oder Funktionalmatrix*

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

⟪⟪ Existiert die totale Ableitung $Df(a)$, so besteht die j-te Spalte ihrer Matrixdarstellung bezüglich den Standardbasen von $\mathbb{R}^n$ und $\mathbb{R}^m$ aus den Komponenten des Vektors $Df(a)e_j$. Dessen i-te Komponente ist

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

Für diese gilt aber

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

Das ist die Behauptung. ⟫⟫

▷ *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 $a \in \mathbb{R}^n$

$$\partial_j f(a) = (a_{ij})_{1 \leqslant i \leqslant m} = (a_{1j}, .., a_{mj})^\top$$

die j-te Spalte von A, und die Jacobimatrix von f ist

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

Für die allgemeine Richtungsableitung erhalten wir

$$\partial_v f(a) = Df(a)v = \sum_{j=1}^{n} \partial_j f(a) v_j = \left(\sum_{j=1}^{n} a_{ij} v_j \right)_{1 \leqslant i \leqslant m} = Av. \qquad \lhd$$

▷ *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 $a_{kl} = a_{lk}$. Nach der Produktregel für Funktionen einer Variablen ist

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

und die allgemeine Richtungsableitung ist

$$\partial_v f(x) = \sum_{j=1}^{n} \partial_j f(x) v_j = 2 \sum_{j=1}^{n} v_j \left(\sum_{k=1}^{n} a_{jk} x_k \right) = 2 \langle Ax, v \rangle. \qquad \lhd$$

Als weiteres Beispiel notieren wir noch den häufig benötigten Spezialfall der Kettenregel für die partiellen Ableitungen im Standardfall.

20 **Korollar** *Sind $f \colon \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ im Punkt a und $g \colon \mathbb{R}^m \hookrightarrow \mathbb{R}^s$ im Punkt $f(a)$ total differenzierbar, so ist*

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

⟨⟨⟨⟨ Es ist ja

$$\begin{aligned} \partial_j (g \circ f)(a) &= Dg(f(a)) \partial_j f(a) \\ &= \sum_{k=1}^{m} \partial_k g(f(a)) \partial_j f_k(a) = \sum_{k=1}^{m} \frac{\partial g}{\partial y_k}(f(a)) \frac{\partial f_k}{\partial x_j}(a). \end{aligned} \qquad ⟩⟩⟩⟩$$

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 15 . Wie aber verifiziert man die Existenz der totalen Ableitung? Die *Existenz* aller partiellen oder aller Richtungsableitungen reicht offensichtlich nicht aus 16 & 17 .

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 diese auf dem *ganzen* Definitionsbereich stetig sind.

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

⟪⟪⟪ Betrachte f auf einer nichtleeren Kugel $B_r(a)$ in seinem Definitionsbereich. Ist $a + h \in B_r(a)$, so liegen auch die Punkte

$$x_i = a + h_1 e_1 + .. + h_i e_i, \qquad 0 \leqslant i \leqslant n,$$

sämtlich in $B_r(a)$, wobei $x_0 = a$ und $x_n = x + h$. 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 die Identität

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

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

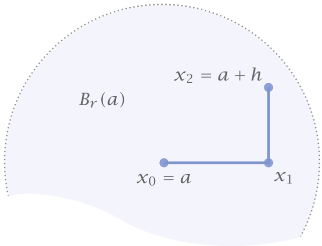
$$\partial_i f(x_{i-1} + th_i e_i) - \partial_i f(a) \to 0, \qquad h \to 0.$$

Das letzte Integral ist daher $o(h_i)$ und damit

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

Abb 3

Zum Differenzierbarkeitsbeweis im Fall $n = 2$



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. Die Klasse dieser Funktionen hat besonders praktische Eigenschaften und daher einen eigenen Namen.

Definition *Eine Abbildung $f : \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ heißt von der Klasse C^1, wenn sämtliche partiellen Ableitungen von f existieren und stetig sind. Die Klasse aller solchen, auf einer offenen Menge $\Omega \subset \mathbb{R}^n$ definierten Abbildungen wird mit $C^1(\Omega, \mathbb{R}^m)$ bezeichnet.* ⋊

▷ Sei $s \geqslant 1$ und betrachte

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) = \frac{1}{2s} \|x\|_e^{2s} = \frac{1}{2s}(x_1^2 + .. + x_n^2)^s.$$

Dann ist

$$\partial_j f(x) = (x_1^2 + .. + x_n^2)^{s-1} x_j = \|x\|_e^{2(s-1)} x_j, \qquad 1 \leqslant j \leqslant n.$$

Diese partiellen Ableitungen sind wegen $s \geqslant 1$ sämtlich stetig auf $\mathbb{R}^n$. Also ist f auch total differenzierbar, und es ist

$$Df(x) = \|x\|_e^{2(s-1)} (x_1, .., x_n) = \|x\|_e^{2(s-1)} x.$$

Man beachte, dass $Df(x)$ eine $1 \times n$-Matrix ist, also ein Zeilenvektor. ◁

2.4 Das Lemma von Hadamard

Der Mittelwertsatz der eindimensionalen Differenzialrechnung 1.9.14 bildet die Grundlage einiger wichtiger Sätze der eindimensionalen Analysis. Leider gilt er für Abbildung in Räume höherer Dimension *nicht mehr*. Für die Kreiskurve $\gamma : t \mapsto (\cos t, \sin t)$ ist zum Beispiel

$$\gamma(2\pi) - \gamma(0) = 0.$$

Für alle $t \in [0, 2\pi]$ gilt aber

$$2\pi\dot{\gamma}(t) = 2\pi(-\sin t, \cos t) \neq 0,$$

da Sinus und Cosinus keine gemeinsamen Nullstellen besitzen.

Es gibt aber ein allgemeineres Resultat, das sogar in beliebigen Dimensionen gilt. Dazu betrachten wir eine Funktion auf der *Verbindungsstrecke*

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

zwischen zwei gegebenen Punkten u und v ihres Definitionsbereiches.

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

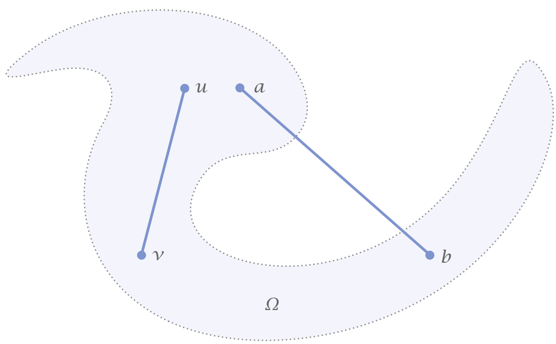
$$f(v) - f(u) = \Lambda(v - u)$$

mit der linearen Abbildung

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

Abb 4

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



Hierbei ist $t \mapsto Df((1-t)u + tv)$ eine Kurve im Vektorraum $L(V, W)$, dessen Integral wir in Abschnitt 1.8.5 erklärt haben. Im Standardfall ist dies eine von t abhängende $m \times n$-Matrix, deren Integral komponentenweise gebildet wird.

⟪⟪ Betrachte die Streckenparametrisierung

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

mit Anfangspunkt u und Endpunkt v. Nach Voraussetzung ist $f \circ \varphi$ wohldefiniert und aufgrund der Kettenregel $_{11}$ stetig differenzierbar. Mit dem Hauptsatz für Kurven $_{1.10}$ ergibt sich

$$f(v) - f(u) = f \circ \varphi \Big|_0^1 = \int_0^1 (f \circ \varphi)'(t)\,\mathrm{d}t = \int_0^1 Df(\varphi(t))\dot{\varphi}(t)\,\mathrm{d}t.$$

Hierbei ist $\dot{\varphi}(t) = v - u$ unabhängig von t, so dass wir diesen Term *hinter* das Integral ziehen können $_{\text{A-1.8.20}}$ [3]. Das ergibt die Behauptung. ⟫⟫

Das Integral über die Ableitung Df entlang der Verbindungsstrecke $[u, v]$ kann im Allgemeinen *nicht* durch die Ableitung an einer geeigneten Zwischenstelle ersetzt werden, wie das Beispiel der Kreiskurve zeigt.

Oft benötigt man den Mittelwertsatz jedoch nur als Grundlage des *Schrankensatzes* $_{1.9.16}$. Dieser gilt in folgender Form auch in höheren Dimensionen.

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

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

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

⟪⟪ Aufgrund des Hadamardschen Lemmas ist

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

Hierbei ist

$$\begin{aligned} \|\Lambda\| &= \left\| \int_0^1 Df((1-t)u + tv)\,\mathrm{d}t \right\| \\ &\leqslant \int_0^1 \|Df((1-t)u + tv)\|\,\mathrm{d}t \\ &\leqslant \max_{t \in [0,1]} \|Df((1-t)u + tv)\| = \max_{z \in [u,v]} \|Df(z)\|\,. \end{aligned}$$ ⟫⟫

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

Schließlich noch ein Korollar, das wir bereits aus dem eindimensionalen Fall kennen.

24 **Satz** *Ist $f : V \hookrightarrow W$ von der Klasse C^1, so ist f lokal lipschitz.* $\rtimes$

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 die Abbildung $Df : V \hookrightarrow L(V, W)$ stetig. Also ist auch

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

stetig. Zu jedem Punkt a existiert daher eine Kugel $B_r(a)$ im Definitionsbereich von f und ein $M < \infty$, so dass

$$\sup_{x \in B_r(a)} \|Df(x)\| \leqslant M.$$

Ist nun $u, v \in B_r(a)$, so ist auch $[u, v] \subset B_r(a)$ wegen der Konvexität jeder Kugel A-1.5.34. Aufgrund des Schrankensatzes 23 gilt dann

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

Somit ist f auf $B_r(a)$ M-lipschitz. ⟫⟫

2.5 Gradient

Für *skalare* Funktionen gibt es schließlich noch einen dritten Ableitungsbegriff, den des *Gradienten*. Ist $f : V \hookrightarrow \mathbb{R}$ 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. Liegt nun ein Skalarprodukt $\langle \cdot, \cdot \rangle$ vor – das heißt, V ist ein Hilbertraum –, 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 $f : V \hookrightarrow \mathbb{R}$ 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*

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

25 ⊳▸ a. Sind $l \in V$ und $x_0 \in V$, so ist

$$L : V \to \mathbb{R}, \quad L(x) = \langle l, x - x_0 \rangle$$

differenzierbar mit $DL(x)h = \langle l, h \rangle$. Also ist $\nabla L(x) = l$ für alle $x \in V$.

b. Ist $A : V \to V$ linear und symmetrisch, also $\langle Av, w \rangle = \langle v, Aw \rangle$ für alle $v, w \in V$, 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

$$\nabla f(x) = 2Ax. \quad ◂⊲$$

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 aber natürlich der Standardraum $\mathbb{R}^n$, der meist stillschweigend mit dem Standardskalarprodukt versehen ist.

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

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

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.

⟨⟨⟨⟨ Ist $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt x differenzierbar, so gilt

$$Df(x)h = \sum_{j=1}^{n} \partial_j f(x) h_j = \langle \nabla f(x), h \rangle$$

genau dann, wenn $\nabla f(x)$ aus den Komponenten $\partial_j f(x)$ besteht. ⟩⟩⟩⟩

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*

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

▪ Tangentialebene

Mithilfe des Gradienten lässt sich die Tangentialebene an den Graphen einer skalaren Funktionen einfach beschreiben. Zunächst die

Definition *Ist $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so heißt der Graph der affinen Funktion*

$$T : V \to \mathbb{R}, \quad z = f(a) + Df(a)(x - a)$$

die Tangentialebene an den Graphen von f im Punkt $(a, f(a))$. ⋊

⊳ Im vorangehenden Beispiel 25 ist die Tangentialebene des linearen Funktionals $x \mapsto \langle l, x \rangle$ im Punkt a gegeben durch

$$z = \langle l, a \rangle + \langle l, x - a \rangle = \langle l, x \rangle,$$

und die der symmetrischen quadratischen Form $x \mapsto \langle Ax, x \rangle$ im Punkt x_0 durch

$$z = \langle Ax_0, x_0 \rangle + 2 \langle Ax_0, x - x_0 \rangle = \langle Ax_0, 2x - x_0 \rangle. \qquad ⊲$$

Die Tangentialebene ist also eine Hyperebene im Raum $\mathbb{R}^n \times \mathbb{R}$. Diese lässt sich durch den Gradienten wie folgt beschreiben.

27 **Satz** *Ist $f: \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so ist*

$$N(a) \coloneqq (-\nabla f(a), 1) \in \mathbb{R}^n \times \mathbb{R}$$

der Normalenvektor der Tangentialebene an den Graphen von f über a. ⋊

⟪⟪ Ausgedrückt mithilfe des Gradienten lautet die Gleichung der Tangentialebene $z = f(a) + \langle \nabla f(a), x - a \rangle$, oder

$$\langle -\nabla f(a), x - a \rangle_n + \langle 1, z - f(a) \rangle_1 = 0.$$

Dies ist gerade die *Normalengleichung* dieser Ebene im Raum $\mathbb{R}^n \times \mathbb{R}$ mit dem Standardskalarprodukt $\langle \cdot, \cdot \rangle_{n+1}$ und dem Normalenvektor

$$N(a) = (-\nabla f(a), 1).$$ ⟫⟫

▪ Richtungen des steilsten Anstiegs und Abstiegs

Der Graph einer stetig differenzierbaren Funktion $f: \mathbb{R}^n \hookrightarrow \mathbb{R}$ bildet eine *Hyperfläche* im $\mathbb{R}^{n+1}$, also eine Fläche der Kodimension 1. Bezüglich dieser Fläche erlaubt der Gradient von f eine interessante Interpretation.

Sei f 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. Wird die Ableitung $f_e'(0)$ für einen Vektor e *maximal*, so nennen wir e eine *Richtung des steilsten Anstiegs*. Wird $f_e'(0)$ *minimal*, so heißt e eine *Richtung des steilsten Abstiegs*.

28 **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 Richtungen des steilsten Anstiegs respektive steilsten Abstiegs von f an der Stelle a. Beide Richtungen sind eindeutig.* ⋊

⟪⟪ Es ist

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

Aufgrund der Cauchy-Schwarzschen Ungleichung gilt

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

Ist $\nabla f(a) \neq 0$, so tritt Gleichheit hierbei *genau dann* ein, wenn die Vektoren $\nabla f(a)$ und e kollinear sind $_{1.5.32}$. 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. ⟫⟫

2.6 Höhere Ableitungen

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

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

wobei wir hier verschiedene Schreibweisen für dieselbe Sache aufführen. Dies kann man eventuell fortsetzen und führt zu folgender

Definition *Sei $f\colon V \hookrightarrow W$ eine Abbildung und $v_1, .., v_n$ eine Basis von V. Dann ist die r-te partielle Ableitung*

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

von f für $r \geqslant 2$ 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

$$f\colon \mathbb{R}^3 \to \mathbb{R}, \quad f(x, y, z) = x^2 y^4 \sin z$$

ist

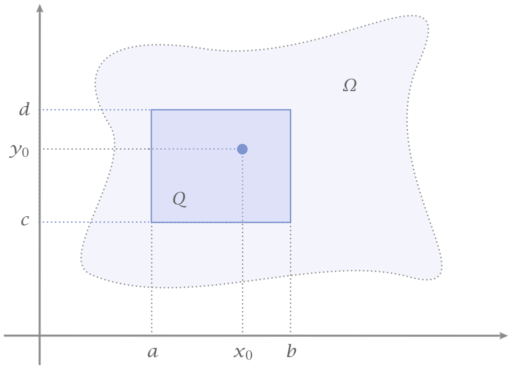
$$\begin{aligned} f_x &= 2xy^4 \sin z, & f_y &= 4x^2y^3 \sin z, & f_z &= x^2y^4 \cos z, \\ f_{xy} &= 8xy^3 \sin z, & f_{yz} &= 4x^2y^3 \cos z, & f_{zx} &= 2xy^4 \cos z, \\ f_{xyz} &= 8xy^3 \cos z, & f_{yzx} &= 8xy^3 \cos z, & f_{zxy} &= 8xy^3 \cos z. \end{aligned}$$ ◁

Bemerkung Schon hier wird deutlich, dass auf die Dauer eine praktischere Notation für solche Ableitungen erforderlich ist. Im nächsten Kapitel führen wir hierfür die *Multiindex-Notation* ein. ⊸

Im vorangehenden Beispiel kommt es auf die Reihenfolge der partiellen Ableitungen *nicht an*. Das ist aber nicht immer so.

Abb 5

Zum Lemma von Schwarz



▷▶ Betrachte

$$f: \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{x^3y - xy^3}{x^2+y^2}, & (x,y) \neq 0, \\ 0, & (x,y) = 0. \end{cases}$$

Wegen $f(0,y) = 0 = f(x,0)$ gilt

$$f_x(0,y) = \lim_{x\to 0} f(x,y)/x = -y,$$
$$f_y(x,0) = \lim_{y\to 0} f(x,y)/y = x,$$

und damit

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

Im Punkt $(0,0)$ stimmen diese beiden Ableitungen somit nicht überein. ◀◁

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

29 **Lemma von Schwarz** *Sei $\Omega \subset \mathbb{R}^n$ offen, $f: \Omega \to \mathbb{R}^m$ 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*

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

⟪⟪ Da nur die beiden Koordinaten x und y involviert sind und alle anderen fixiert werden können, beschränken wir uns auf den Fall $f = f(x,y)$.

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

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

Nach Voraussetzung ist f_{xy} auf Q stetig. Aufgrund des anschließend bewiesenen Lemmas 30 definiert das Integral daher eine in y differenzierbare Funktion, deren Ableitung man durch Differenzieren ›unter dem Integral‹ erhält. Da natürlich auch $f(a, y)$ nach y differenzierbar ist, gilt also

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

Da der Integrand f_{xy} nach Voraussetzung ebenfalls stetig ist, definiert dieses Integral eine nach x differenzierbare Funktion. Da der andere Term $f_y(a, y)$ nicht von x abhängt, ist somit $f_y(x, y)$ nach x differenzierbar, und es gilt

$$f_{yx}(x, y) = \frac{\partial}{\partial x} \int_a^x f_{xy}(s, y)\,\mathrm{d}s = f_{xy}(x, y).$$

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.

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

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

differenzierbar, und es gilt

$$\phi' = \int_a^b f_y(s, y)\,\mathrm{d}s. \qquad \rtimes$$

⟪⟪ Sei zuerst y ein innerer Punkt von $[c, d]$. Aufgrund des Hauptsatzes der Differenzial- und Integralrechnung 1.9.23 ist

$$\begin{aligned} \phi(y + h) - \phi(y) &= \int_a^b [f(s, y + h) - f(s, y)]\,\mathrm{d}t \\ &= \int_a^b \int_0^1 f_y(s, y + th)h\,\mathrm{d}t\,\mathrm{d}s. \end{aligned}$$

Damit folgt

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

Wegen der Kompaktheit von Q ist h_y auf Q *gleichmäßig stetig* 1.7.43. Zu jedem $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass für *alle* $s \in [a,b]$ und $t \in [0,1]$

$$\left|f_y(s,y+th)-f_y(s,y)\right| < \varepsilon, \qquad |h| < \delta.$$

Für diese h erhalten wir

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

Also gilt

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

Der Fall eines Randpunktes y wird entsprechend behandelt. ⟫⟫⟫

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 $f\colon \Omega \to \mathbb{R}^m$, 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

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

den Raum der unendlich oft partiell differenzierbaren Abbildungen $\Omega \to \mathbb{R}^m$. Diese werden auch *glatt* genannt [4]. Alle diese Räume sind *lineare Vektorräume.*

[4] *Hinreichend glatt* bedeutet dagegen *genügend oft differenzierbar*!

Im Fall skalarer Funktionen schreibt man noch kürzer

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

Mit Induktion beweist man die folgende unmittelbare Folgerung aus dem Satz von Schwarz.

31 **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.* $\rtimes$

Totale Ableitungen

Wir haben bisher partielle Ableitungen höherer Ordnung definiert, aber noch nicht die entsprechenden totalen Ableitungen

$$D^2 f, \quad D^3 f, \quad .. \ .$$

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 $f : \mathbb{R}^n \hookrightarrow \mathbb{R}^m$ total differenzierbar, so ist

$$Df : \ \mathbb{R}^n \hookrightarrow L(\mathbb{R}^n, \mathbb{R}^m),$$

da jedem Punkt im Definitionsbereich eine lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ zugeordnet wird. Nun ist der Zielraum wieder ein endlich-dimensionaler Raum, denn $L(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{nm}$. Also ist auch die zweite totale Ableitung aufgrund unserer allgemeinen Definition erklärt – wenn sie existiert –, und es ist

$$D^2 f = D(Df) : \ \mathbb{R}^n \hookrightarrow L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \simeq L(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m).$$

Somit ist $D^2 f$ eine *bilineare Abbildung* des $\mathbb{R}^n$ in den $\mathbb{R}^m$.

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

$$D^r f : \ \mathbb{R}^n \hookrightarrow L(\mathbb{R}^n \times \cdot\cdot \times \mathbb{R}^n, \mathbb{R}^m),$$

die jedem Punkt im Definitionsbereich eine r-lineare Abbildung des $\mathbb{R}^n$ in den $\mathbb{R}^m$ zuordnet. Ist f eine C^r-Abbildung, so ist $D^r f$ in eindeutiger Weise durch die partiellen Ableitungen der Ordnung r von f definiert. Auf die Details verzichten wir hier.

Aufgaben

1 Für ein beliebiges Skalarprodukt $\langle\cdot,\cdot\rangle$ auf dem $\mathbb{R}^n$ gilt

$$\langle u,v\rangle = \langle Au,v\rangle_e$$

mit der symmetrischen Darstellungsmatrix $A = (\langle e_i,e_j\rangle)_{1\leqslant i,j\leqslant n}$.

2 Sei $f\colon V \hookrightarrow W$ im Punkt a differenzierbar $_4$. Dann ist f auch differenzierbar in a mit derselben Ableitung, wenn man die Normen auf V und W durch äquivalente Normen ersetzt.

3 Sei $A\colon V \to W$ eine lineare Abbildung zwischen zwei normierten Vektorräumen. Dann gilt $Ah = o(h)$ genau dann, wenn $A = 0$.

4 Für $a > 0$ sei

$$\ell_\infty^a \coloneqq \{x \in \ell^\infty : \|x\|_{a,\infty} \coloneqq \sup_{n\geqslant 1} \mathrm{e}^{an}\,|x_n| < \infty\}.$$

Dieser Raum ist mit $\|\cdot\|_{a,\infty}$ ein Banachraum. Der Operator

$$D:\ \ell_\infty^a \to \ell_\infty^a, \quad Dx = D(x_n)_{n\geqslant 1} = (nx_n)_{n\geqslant 1}$$

ist linear und bijektiv, aber nicht beschränkt.

5 Gegeben seien Funktionen $\varphi, f, g\colon [-1,1] \to \mathbb{R}$. Dabei sei φ stetig und

$$f(h) = O(h), \qquad g(h) = o(h).$$

Bestimmen sie (natürlich mit Beweis) die Ordnung von
a. $(f\circ g)(h)$ *b.* $g^2(h)$ *c.* $(\varphi f)(h)$ *d.* $(fg)(h)$ *e.* $g(h+f(h))$

6 Bestimmen sie die partiellen Ableitungen der folgenden Funktionen $\mathbb{R}^2 \to \mathbb{R}$.
a. $ax^2 + 2bxy + cx^2$ *b.* $\exp(x^2+y^2)$
c. $\sin(xy) + \cos(xy) + \exp(x^2y^2)$.

7 *Allgemeineres Differenzierbarkeitskriterium* Existieren sämtliche partiellen Ableitungen von $f\colon V \hookrightarrow W$ bezüglich einer Basis $v_1,..,v_n$ von V und sind diese stetig, so ist f total differenzierbar und $Df\colon V \hookrightarrow L(V,W)$ ist ebenfalls stetig.

8 Sei $f\colon V \to W$ differenzierbar. Für $\alpha \in \mathbb{R}$ und $v,w \in V$ gilt dann

$$\partial_{\alpha v} f(a) = \alpha\partial_v f(a), \qquad \partial_{v+w} f(a) = \partial_v f(a) + \partial_w f(a).$$

9 Die Funktion

$$f:\ \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{2xy^2}{x^2+y^4}, & (x,y) \neq 0, \\ 0 \quad , & (x,y) = 0. \end{cases}$$

besitzt im Nullpunkt sämtliche Richtungsableitungen, ist dort aber nicht stetig.

10 Gegeben ist die Funktion

$$f\colon\ \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{x^3y - xy^3}{x^2+y^2}, & (x,y) \neq 0, \\ 0, & (x,y) = 0. \end{cases}$$

Zeigen sie, dass f alle partiellen Ableitungen *zweiter* Ordnung besitzt. Bestimmen sie diese Ableitungen, insbesondere im Nullpunkt.

11 Zeigen sie, dass $f\colon \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x) = \begin{cases} |x|_e^2 \sin |x|_e^{-1}, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

überall differenzierbar ist. Besitzt f überall stetige partielle Ableitungen?

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

$$f(v) - f(u) = Df(\xi)(v-u)$$

mit einem $\xi \in [u,v]$.

13 Mit $r > 0$ ist

$$\varphi\colon\ \mathbb{R}^n \to \mathbb{R}, \quad \varphi(x) = \begin{cases} 0, & |x| \leqslant r \\ \exp((|x|-r)^{-1}), & |x| > r \end{cases}$$

stetig differenzierbar.

14 Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ differenzierbar und $\lambda \in \mathbb{R}$. Gilt $f(tx) = t^\lambda f(x)$ für alle x und $t > 0$, so folgt

$$Df(x)\,x = \lambda f(x).$$

Hiervon gilt auch die Umkehrung. *Hinweis:* Bestimmen sie eine DGl für

$$\varphi = t^\lambda f(x) - f(tx).$$

15 Konstruieren sie eine Kurve $\varphi \in C^0([0,1],\mathbb{R}^2)$ und eine Funktion $f \in C^1(\mathbb{R}^2)$ so, dass

$$Df(\varphi(t)) \equiv 0,$$

aber $f(\varphi(t))$ nicht konstant ist. (H. Whitney, *Duke Math. J.* 1 (1935) 514–517)

16 Sei $\langle\cdot,\cdot\rangle : V \to \mathbb{R}$ eine symmetrische Bilinearform und $A\colon V \to V$ linear, aber *nicht* notwendigerweise symmetrisch. Zeigen sie, dass

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

differenzierbar ist, und bestimmen sie die Ableitung.

17 *Produktregel* Seien $\langle\cdot,\cdot\rangle : V \to \mathbb{R}$ eine symmetrische Bilinearform und $f, g \colon V \to V$ differenzierbare Abbildungen. Zeigen sie, dass

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

differenzierbar ist, und bestimmen sie die Ableitung.

18 Bestimmen sie die Jacobimatrizen der folgenden Abbildungen.

a. $f \colon \mathbb{R}^n \to \mathbb{R}^1$ mit $f(x) = \|x\|_e^2$

b. $f \colon \mathbb{R}^1 \to \mathbb{R}^3$ mit $f(t) = (t^2 \cos t, t^2 \sin t, \sinh t)^\top$

c. $f \colon \mathbb{R}^4 \to \mathbb{R}^2$ mit $f(a,b,c,d) = (ab + cd, a^2c^2 - b^2d^2)^\top$

d. $f \colon \mathbb{R}^3 \to \mathbb{R}^3$ mit $f(x,y,z) = \begin{pmatrix} x \sin y \sin z \\ x \sin y \cos z \\ x \cos y \end{pmatrix}$

19 Seien $f, g \colon \mathbb{R}^n \to \mathbb{R}$ differenzierbare Funktionen und $\varphi = fg$.

a. Bestimmen sie $D\varphi(x)h$ für $h \in \mathbb{R}^n$.

b. Bestimmen sie $\nabla\varphi(x)$ bezüglich des Standardskalarprodukts.

20 Bestimmen sie zu

$$f : \; \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \int_{(x^2+y^2)/2}^{xy} \mathrm{e}^{-t} \, \mathrm{d}t$$

die Tangentialebene an den Graphen über dem Punkt $(1,1)$.

21 Sei I ein kompaktes Intervall und $C(I)$ mit der Supremumsnorm versehen. Dann ist

$$\Phi : \; C(I) \to \mathbb{R}, \qquad \Phi(f) = \frac{1}{2} \int_I f^2(t) \, \mathrm{d}t$$

differenzierbar mit

$$D\Phi(f)\eta = \int_I f(t)\eta(t) \, \mathrm{d}t.$$

22 *Rayleighquotient* Für eine symmetrische $n \times n$-Matrix A heißt

$$\varphi : \; \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad \varphi(x) = \frac{1}{2} \frac{\langle Ax, x \rangle}{\langle x, x \rangle}$$

der *Rayleighquotient*. Bestimmen sie die kritischen Punkte x_0 von φ und die zugehörigen kritischen Werte $\varphi(x_0)$.

23 *Eine Variante des Lemmas von Hadamard* Sei $f \in C^2(\Omega)$ und $0 \in \Omega$. Dann gibt es Funktionen $g_1, .., g_n \in C^1(\Omega)$ so dass

$$f(x) = f(0) + \sum_{i=1}^n x_i g_i(x).$$

24 Ist $f \colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und L-lipschitz, so ist Df gleichmäßig beschränkt in dem Sinne, dass in jedem Punkt

$$\|Df(x)\| \coloneqq \sup_{0 \neq h \in \mathbb{R}^n} \frac{|Df(x)h|}{|h|} \leqslant L.$$

25 Die Funktion $u \in C^1(\mathbb{R}^2)$ erfülle die Differenzialgleichung $yu_x = xu_y$. Dann ist $Du(0,0) = 0$, und es existiert ein $\varphi \in C^1([0,\infty))$ derart, dass $u(x,y) = \varphi(x^2 + y^2)$. Gilt dies auch, wenn nur $u \in C^1(\mathbb{R}^2)$?

Springer

http://www.springer.com/978-3-658-05859-3

Etwas mehr Analysis
Eine Einführung in die mehrdimensionale Analysis
Pöschel, J.
2014, X, 292 S. 93 Abb. in Farbe., Softcover
ISBN: 978-3-658-05859-3