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.

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

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

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

8 ▷ A. Für eine affine Abbildung erhält man

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

B. Für eine quadratische Form gilt

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

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

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

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

10 ⊳ Betrachte

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

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.

10 **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. ⋊

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

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

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

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

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.

11 **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*

$$Jf(a) \coloneqq (\partial_j f_i(a))_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}.$$

11 ⟨⟨⟨⟨⟨ 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

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

11 ▹ 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*

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

12 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

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

▪ 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.* ⋊

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

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

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

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

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

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

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

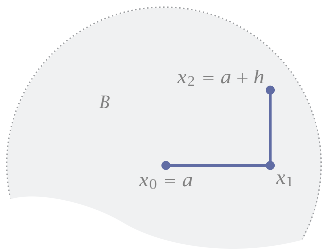
$$f(x_i) - f(x_{i-1}) = \partial_i f(a) h_i + o(h_i).$$

Insgesamt erhalten wir

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

Abb 3

Zum Beweis des Differenzierbarkeitskriteriums



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].

14.4 Das Lemma von Hadamard

Der Mittelwertsatz der eindimensionalen Differenzialrechnung [8.10] 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

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

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*

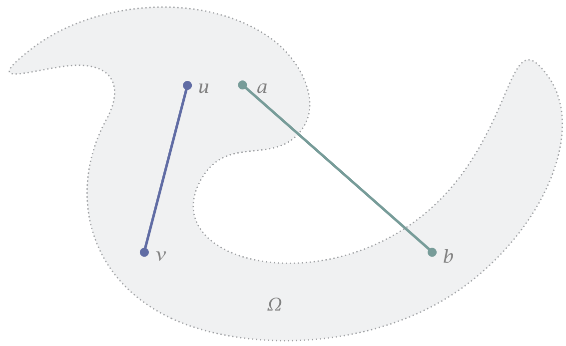
$$[u, v] \coloneqq \{(1-t)u + tv : 0 \leqslant t \leqslant 1\}$$

zwischen zwei gegebenen Punkten u und v ihres Definitionsbereiches.

14.19

Abb 4

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



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

$$f(v) - f(u) = A(v-u)$$

mit der linearen Abbildung

$$A = \int_0^1 Df((1-t)u + tv)\,\mathrm{d}t. \qquad \rtimes$$

Hierbei ist $t \mapsto Df((1-t)u+tv)$ eine 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.

14 ⟨⟨⟨⟨⟨ Betrachte die Streckenparametrisierung

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

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 $_{A\text{-}10.29}$ [2]. 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.

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

Oft benötigt man den Mittelwertsatz jedoch nur als Grundlage des *Schrankensatzes* $_{8.11}$. Dieser gilt in folgender Form auch in höheren Dimensionen.

15 **Schrankensatz** *Sei $f: V \hookrightarrow W$ 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 $\|\cdot\|$. ⋊

15 ⟨⟨⟨⟨⟨ Aufgrund des Hadamardschen Lemmas ist

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

mit

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

wobei $\varphi(t) = (1-t)u + tv$ wie zuvor. ⟩⟩⟩⟩⟩

Somit gilt auch in höheren Dimensionen das folgende

15 **Korollar** *Ist $f: V \hookrightarrow W$ 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.

15 ⟨⟨⟨⟨⟨ Nach Voraussetzung ist $Df: V \hookrightarrow L(V,W)$ 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 $_{15}$ gilt

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

Somit ist f auf B M-lipschitz. ⟩⟩⟩⟩⟩