25. Differenzierbarkeit im Mehrdimensionalen

Wie im eindimensionalen Fall in Kapitel 10 wollen wir uns nach der Stetigkeit von Abbildungen jetzt mit der Differenzierbarkeit beschäftigen. Wir erinnern uns dazu zunächst einmal daran, wie wir differenzierbare Funktionen damals definiert hatten: Hat D keine isolierten Punkte, ist $f\colon D \to \mathbb{K}$ eine Funktion und $a \in D$, so heißt f differenzierbar in a, wenn der Grenzwert des Differenzenquotienten

$$f'(a) := \lim_{x\to a} \frac{f(x)-f(a)}{x-a} \quad \in \mathbb{K} \tag{$*$}$$

(der dann die Ableitung von f in a genannt wird) existiert. Die anschauliche Idee hinter dieser Definition war, dass dann für x in der Nähe von a die Approximation

$$f'(a) \approx \frac{f(x)-f(a)}{x-a}, \quad \text{und damit} \quad f(x) \approx f(a) + f'(a)\cdot(x-a)$$

gilt, dass wir f also in der Nähe von a durch eine Gerade $x \mapsto f(a) + f'(a)\cdot(x-a)$ mit Steigung $f'(a)$ annähern. Wollen wir diese Idee der linearen Approximierbarkeit nun auf andere Start- und Zielbereiche verallgemeinern, stellen wir zunächst fest:

- Um lineare Approximierbarkeit formulieren zu können, müssen wir offensichtlich wissen, was lineare Abbildungen zwischen den betrachteten Räumen überhaupt sind. In metrischen Räumen wäre dies z. B. nicht der Fall, da dort im Allgemeinen ja nicht einmal eine Addition bzw. Multiplikation definiert ist. Wir werden uns daher für die Untersuchung der Differenzierbarkeit auf normierte Räume (bzw. Teilmengen davon) beschränken müssen.
- Betrachten wir als Beispiel einmal als Start- und Zielraum die normierten Räume $\mathbb{K}^n$ bzw. $\mathbb{K}^m$, so ist nun die Idee der linearen Approximierbarkeit $f(x) \approx f(a) + f'(a)\cdot(x-a)$ für $x \in \mathbb{K}^n$ in der Nähe eines gegebenen Punktes $a \in \mathbb{K}^n$ sinnvoll formulierbar, wenn wir dies als Gleichung im Zielraum $\mathbb{K}^m$ auffassen, $f'(a) \in \mathrm{Mat}(m\times n, \mathbb{K})$ eine Matrix ist und „$\cdot$" das Matrixprodukt mit dem Vektor $x-a \in \mathbb{K}^n$ darstellt. Ableitungen werden im Mehrdimensionalen also zu Matrizen werden.
- Die Ableitungsmatrix $f'(a)$ wird uns dann von a aus für jeden kleinen Schritt $x-a$ in einer beliebigen Richtung angeben, um welchen Wert $f'(a)\cdot(x-a)$ sich f dabei ändert. Damit dies eindeutig bestimmt ist, muss man sich dem Punkt a aber offensichtlich innerhalb des Definitionsbereichs überhaupt erst einmal aus jeder Richtung nähern können. Im Gegensatz zum Eindimensionalen werden wir im Folgenden daher nur offene Definitionsbereiche betrachten, also nicht über Differenzierbarkeit in Randpunkten sprechen.

Mit diesen Vorüberlegungen werden wir im folgenden Abschnitt nun die Ableitung im Mehrdimensionalen definieren und angeben, wie man sie berechnen kann.

25.A Differenzierbare Abbildungen

Wollen wir die Definition der Ableitung ins Mehrdimensionale übertragen, haben wir zunächst ein weiteres Problem: Wir können keinen Differenzenquotienten wie in $(*)$ in der Einleitung oben bilden, da wir hierzu durch den Vektor $x-a$ teilen müssten — was natürlich nicht möglich ist. Daher werden wir als Erstes nun die eindimensionale Definition so umschreiben, dass sie sich besser auf die neue Situation übertragen lässt.

Lemma 25.1 (Differenzierbarkeit als lineare Approximierbarkeit)**.** *Es seien $D \subset \mathbb{K}$ offen, $f\colon D \to \mathbb{K}$ eine Abbildung und $a \in D$. Dann ist f genau dann in a differenzierbar mit Ableitung $c = f'(a) \in \mathbb{K}$, wenn es eine Funktion $r\colon D \to \mathbb{K}$ gibt mit*

(a) *$f(x) = f(a) + c\,(x-a) + r(x)$ für alle $x \in D$, und*

(b) $\lim\limits_{x \to a} \dfrac{r(x)}{x-a} = 0.$

Beweis. Die Bedingung (a) besagt offensichtlich, dass r die Funktion $r(x) = f(x) - f(a) - c\,(x-a)$ ist. Damit gelten die beiden Bedingungen (a) und (b) genau dann, wenn

$$\lim_{x \to a} \frac{f(x) - f(a) - c\,(x-a)}{x-a} = 0, \quad \text{also} \quad \lim_{x \to a} \left(\frac{f(x) - f(a)}{x-a} - c \right) = 0$$

ist, d. h. wenn f differenzierbar in a mit Ableitung c ist. □

Bemerkung 25.2. Man kann sich die Funktion r in Lemma 25.1 als „Restfunktion" vorstellen, also als Differenz zwischen der eigentlichen Funktion f und ihrer Näherung $x \mapsto f(a) + c\,(x-a)$ bei a. Natürlich erhalten wir aus (a) am gegebenen Punkt $x = a$ in jedem Fall schon einmal $r(a) = 0$. Die Bedingung (b) besagt nun gerade, dass dieser Rest auch „in der Nähe des Punktes a sehr klein", die Näherung also dort sehr gut ist: Der Quotient $\frac{r(x)}{x-a}$ ist ja bei a von der Form „$\frac{0}{0}$", und die gegebene Bedingung besagt nun gerade, dass sich hier für $x \to a$ der Zähler $r(x)$ gegen den Nenner $x-a$ durchsetzt. Man sagt dafür auch, dass „die Restfunktion r für $x \to a$ schneller als linear gegen 0 konvergiert".

Auf den ersten Blick sieht es nun vielleicht so aus, als ob wir mit Lemma 25.1 im Hinblick auf eine Verallgemeinerung auf das Mehrdimensionale noch nicht viel erreicht haben, weil wir in (b) ja immer noch durch $x-a$ dividieren. Dies ist aber nicht so: Die gegebene Bedingung ist ja offensichtlich äquivalent zu $\lim_{x \to a} \frac{r(x)}{|x-a|} = 0$, und dort können wir nun einfach die Betragsstriche durch die Norm ersetzen. Wir erhalten so die folgende Definition der Differenzierbarkeit, die sich von der obigen Bedingung nur dadurch unterscheidet, dass c nun eine $m \times n$-Matrix und $x-a \mapsto c\,(x-a)$ damit eine allgemeine lineare Abbildung vom Startraum $\mathbb{K}^n$ in den Zielraum $\mathbb{K}^m$ (statt wie bisher von $\mathbb{K}$ nach $\mathbb{K}$) ist.

Definition 25.3 (Differenzierbarkeit in $\mathbb{K}^n$)**.** Es seien $D \subset \mathbb{K}^n$ offen und $a \in D$. Eine Abbildung $f \colon D \to \mathbb{K}^m$ heißt **(total) differenzierbar** in a, wenn es eine Matrix $A \in \mathrm{Mat}(m \times n, \mathbb{K})$ und eine Abbildung $r \colon D \to \mathbb{K}^m$ gibt, so dass

(a) $f(x) = f(a) + A\,(x-a) + r(x)$ für alle $x \in D$, und

(b) $\lim\limits_{x \to a} \dfrac{r(x)}{\|x-a\|} = 0.$

Wir werden in Folgerung 25.12 noch sehen, dass die Matrix A (und damit auch r) in diesem Fall eindeutig bestimmt ist. Wir nennen sie dann die **(totale) Ableitung** von f in a und bezeichnen sie wie üblich mit $f'(a) \in \mathrm{Mat}(m \times n, \mathbb{K})$ (in manchen Büchern ist hierfür auch die Bezeichnung $Df(a)$ üblich, da Matrizen oftmals mit großen Buchstaben bezeichnet werden).

Die Abbildung f heißt **differenzierbar**, wenn sie in jedem Punkt $a \in D$ differenzierbar ist.

Bemerkung 25.4.

(a) Mit dem gleichen Argument wie z. B. bei der Beschränktheit in Bemerkung 23.21 (b) sieht man auch hier, dass Definition 25.3 nicht von der verwendeten Norm abhängt.

(b) Im eindimensionalen Fall $m = n = 1$ stimmt Definition 25.3 aufgrund von Lemma 25.1 offensichtlich mit der alten Definition 10.3 der Differenzierbarkeit überein, wenn wir die 1×1-Matrix $f'(a)$ als Element von $\mathbb{K}$ auffassen.

(c) Ist $f \colon D \to \mathbb{K}^m$ differenzierbar, so können wir die Ableitung ihrerseits wieder als Funktion

$$f' \colon D \to \mathrm{Mat}(m \times n, \mathbb{K}), \; x \mapsto f'(x)$$

auffassen. Beachte allerdings, dass diese Ableitungsfunktion im Gegensatz zum Eindimensionalen zwar die gleiche Startmenge D, aber nicht die gleiche Zielmenge wie die ursprüngliche Funktion $f \colon D \to \mathbb{K}^m$ hat!

Beispiel 25.5 (Differenzierbarkeit linearer Abbildungen)**.** Ist $f\colon \mathbb{R}^n \to \mathbb{R}^m$ selbst bereits eine lineare Abbildung, also $f(x) = Ax$ für eine Matrix $A \in \operatorname{Mat}(m \times n, \mathbb{R})$, so gilt $f(x) = f(a) + A(x-a)$ für alle $x \in \mathbb{R}^n$. Die Bedingungen aus Definition 25.3 sind dann also mit der Nullfunktion als Restfunktion r erfüllt. Also ist f in diesem Fall differenzierbar mit (von x unabhängiger) Ableitung $f'(x) = A$.

Bemerkung 25.6 (Differenzierbarkeit in normierten Räumen)**.** Genau wie bei unserer Untersuchung linearer Abbildungen in Kapitel 16 haben wir den Begriff der Differenzierbarkeit hier zunächst einmal nur für Abbildungen zwischen (Teilmengen von) $\mathbb{K}^n$ und $\mathbb{K}^m$ definiert. Wie ihr euch vielleicht schon denken könnt, ist eine analoge Definition aber genauso in beliebigen normierten Räumen möglich, wobei die Ableitung dann statt einer Matrix eine lineare Abbildung zwischen den gegebenen Räumen ist: Sind V und W normierte Räume und $f\colon D \to W$ eine Abbildung auf einer offenen Teilmenge $D \subset V$, so nennt man f in einem Punkt $a \in D$ *differenzierbar*, wenn es eine *stetige* lineare Abbildung $\Phi \in \operatorname{Hom}(V,W)$ und eine Abbildung $r\colon D \to W$ gibt mit

(a) $f(x) = f(a) + \Phi(x-a) + r(x)$ für alle $x \in D$, und

(b) $\displaystyle\lim_{x\to a} \frac{r(x)}{\|x-a\|} = 0$.

Auch in diesem Fall heißt Φ dann die *Ableitung* $f'(a) \in \operatorname{Hom}(V,W)$ von f in a.

Für $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$ ist dies aufgrund des Isomorphismus $\operatorname{Mat}(m \times n, \mathbb{K}) \cong \operatorname{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ aus Satz 16.11 offensichtlich äquivalent zu Definition 25.3. Auch für beliebige endlich erzeugte normierte Räume können wir diese neue Definition nach Wahl von Basen von V und W noch auf Definition 25.3 zurückführen, da diese Basen ja Identifikationen von V und W mit $\mathbb{K}^n$ und $\mathbb{K}^m$ erzeugen. Beachte aber, dass dieser verallgemeinerte Begriff der Differenzierbarkeit in nicht endlich erzeugten normierten Räumen in der Regel von den gewählten Normen abhängen wird, und dass die Stetigkeit der linearen Abbildung $f'(a) \in \operatorname{Hom}(V,W)$ dann auch eine echte Bedingung ist (siehe Aufgabe 24.12).

Für den Rest dieser Vorlesung werden wir der Einfachheit halber aber fast ausschließlich mit Abbildungen zwischen Teilmengen von $\mathbb{K}^n$ und $\mathbb{K}^m$ arbeiten, und können die Ableitung solcher Funktionen daher wie in Definition 25.3 als Matrizen schreiben.

Als Erstes wollen wir nun wie im Eindimensionalen (siehe Folgerung 10.7) als einfache Folgerung aus Definition 25.3 sehen, dass jede differenzierbare Funktion auch stetig ist.

Lemma 25.7. *Sind $D \subset \mathbb{K}^n$ offen und $f\colon D \to \mathbb{K}^m$ eine in einem Punkt $a \in D$ differenzierbare Funktion, so ist f auch stetig in a.*

Beweis. Erfüllt f die Differenzierbarkeitsbedingung aus Definition 25.3, so gilt

$$\lim_{\substack{x\to a\\ x\neq a}} f(x) = \lim_{\substack{x\to a\\ x\neq a}} \Big(f(a) + f'(a)\, \underbrace{(x-a)}_{\to 0} + \underbrace{\frac{r(x)}{\|x-a\|}}_{\to 0}\, \underbrace{\|x-a\|}_{\to 0} \Big) = f(a)$$

und damit auch $\lim_{x\to a} f(x) = f(a)$, d. h. f ist stetig in a. □

Um Beispiele von differenzierbaren Funktionen und Ableitungen angeben zu können, müssen wir uns nun als Nächstes fragen, wie man Ableitungen überhaupt berechnen kann. Beachte, dass dies nicht offensichtlich ist, da Definition 25.3 die Ableitung im Gegensatz zum Eindimensionalen ja nicht über eine konkrete Formel, sondern als eine Matrix mit einer bestimmten Eigenschaft definiert — aber nichts darüber aussagt, wie man eine solche Matrix finden kann. Über das folgende Konzept der Richtungsableitungen können wir dieses Problem auf den uns bekannten eindimensionalen Fall zurückführen.

Definition 25.8 (Richtungsableitungen und partielle Ableitungen)**.** Es seien $D \subset \mathbb{K}^n$ offen, $a \in D$ und $f\colon D \to \mathbb{K}^m$ eine Abbildung. Ferner wählen wir einen Vektor $v \in \mathbb{K}^n \backslash \{0\}$.

(a) Existiert der Grenzwert

$$\partial_v f(a) := \lim_{t\to 0} \frac{f(a+t\,v)-f(a)}{t} \quad \in \mathbb{K}^m,$$

so nennen wir ihn die **Richtungsableitung** von f in a in Richtung v. In der Literatur schreibt man hierfür manchmal auch $\frac{\partial f}{\partial v}(a)$.

(b) Ist speziell $v = e_i$ der i-te Einheitsvektor für ein $i \in \{1,\dots,n\}$, so heißt die entsprechende Richtungsableitung

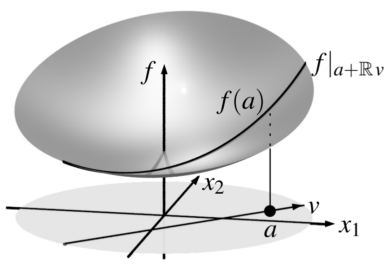
$$\partial_i f(a) := \partial_{e_i} f(a) = \lim_{t\to 0} \frac{f(a+t\,e_i)-f(a)}{t} \quad \in \mathbb{K}^m$$

die i-te **partielle Ableitung** von f in a; hierfür ist auch die Schreibweise $\frac{\partial f}{\partial x_i}(a)$ üblich. Existieren alle diese partiellen Ableitungen (aber nicht notwendig alle Richtungsableitungen) von f in a, so nennt man f in a **partiell differenzierbar**.

Die Funktion f heißt *partiell differenzierbar*, wenn sie in jedem Punkt $a \in D$ partiell differenzierbar ist.

Bemerkung 25.9.

(a) Im Fall $m = 1$ und $\mathbb{K} = \mathbb{R}$ ist die Richtungsableitung leicht geometrisch zu interpretieren: Nach Definition 25.8 (a) ist sie einfach die gewöhnliche Ableitung der Funktion $t \mapsto f(a+tv)$ in einer Variablen t im Punkt 0, also wie im Bild rechts die Steigung der Funktion f im Punkt a, wenn man sie auf die Gerade $a+\mathbb{R}\,v$ durch a mit Richtungsvektor v einschränkt (und t als Koordinate auf dieser Geraden wählt). Die Richtungsableitung $\partial_v f(a)$ gibt also an, wie stark sich f ändert, wenn man sich von a aus ein kleines Stück in Richtung v bewegt.



(b) Im Gegensatz zur (totalen) Ableitung haben die Richtungsableitungen (und damit auch die partiellen Ableitungen) einer Funktion $f\colon D \to \mathbb{K}^m$ den Vorteil, dass sie im Fall ihrer Existenz in jedem Punkt selbst wieder Funktionen $\partial_v f\colon D \to \mathbb{K}^m$ mit der gleichen Start- und Zielmenge wie f sind (vergleiche Bemerkung 25.4 (c)).

(c) Im Fall $m = 1$ einer Abbildung mit Zielraum $\mathbb{K}$ lassen sich die partiellen Ableitungen von f leicht mit unseren Rechenregeln aus Kapitel 10 berechnen: Es ist dann nämlich

$$\partial_i f(a) = \lim_{t\to 0} \frac{1}{t} \cdot \left(f\begin{pmatrix} a_1 \\ \vdots \\ a_i+t \\ \vdots \\ a_n \end{pmatrix} - f\begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} \right)$$

genau die gewöhnliche Ableitung von f nach der Variablen x_i im Punkt a_i, wenn man die übrigen Variablen als konstant ansieht. Aber auch im allgemeinen Fall $m > 1$ ist die Berechnung der partiellen Ableitungen einfach: Da sich der Grenzwert aus Definition 25.8 dann nach Lemma 24.7 koordinatenweise berechnen lässt, ist $\partial_i f(a)$ einfach der Vektor mit Koordinaten $\partial_i f_j(a)$ für $j = 1,\dots,m$, wenn $f_1,\dots,f_m\colon D \to \mathbb{K}$ die Komponentenfunktionen von $f\colon D \to \mathbb{K}^m$ sind.

Beispiel 25.10. Um die partiellen Ableitungen der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 x_2 \\ x_1 \sin x_2 \end{pmatrix}$$

zu berechnen, müssen wir nach Bemerkung 25.9 (c) einfach nur die Komponentenfunktionen von f nach x_1 bzw. x_2 differenzieren und dabei jeweils die andere Variable als Konstante ansehen. Wir erhalten also mit den üblichen Rechenregeln aus Kapitel 10

$$\partial_1 f(x) = \begin{pmatrix} x_2 \\ \sin x_2 \end{pmatrix} \quad \text{und} \quad \partial_2 f(x) = \begin{pmatrix} x_1 \\ x_1 \cos x_2 \end{pmatrix}.$$

Insbesondere ist f also partiell differenzierbar.

Die partiellen Ableitungen einer Funktion können in der Regel also sehr einfach berechnet werden. Um zu sehen, wie sie uns bei der Bestimmung der totalen Differenzierbarkeit bzw. der Ableitungsmatrix helfen, benötigen wir das folgende einfache Lemma.

Lemma 25.11 (Differenzierbare Funktionen haben Richtungsableitungen)**.** *Es seien $D \subset \mathbb{K}^n$ offen, $f\colon D \to \mathbb{K}^m$ eine Funktion, und $a \in D$. Ist f in a differenzierbar, so existiert für jedes $v \in \mathbb{K}^n \backslash \{0\}$ die Richtungsableitung $\partial_v f(a)$, und es gilt*

$$\partial_v f(a) = f'(a) \cdot v$$

(wobei $f'(a) \cdot v$ das Matrixprodukt von $f'(a) \in \mathrm{Mat}(m \times n, \mathbb{K})$ mit $v \in \mathbb{K}^n$ ist).

Beweis. Setzen wir die Darstellung $f(x) = f(a) + f'(a)\,(x-a) + r(x)$ aus Definition 25.3 an der Stelle $x = a + t\,v$ in die Formel für die Richtungsableitung aus Definition 25.8 (a) ein, so erhalten wir

$$\partial_v f(a) = \lim_{t \to 0} \frac{f(a) + f'(a)\,(a+t\,v-a) + r(a+t\,v) - f(a)}{t} = f'(a) \cdot v + \underbrace{\lim_{t \to 0} \frac{r(a+t\,v)}{t}}_{(*)}.$$

Es bleibt also nur noch zu zeigen, dass der Grenzwert $(*)$ gleich 0 ist. Dies können wir z. B. mit dem Folgenkriterium tun: Es sei dazu $(t_n)_n$ eine beliebige Nullfolge in $\mathbb{K}\backslash\{0\}$, so dass $x_n := a + t_n\,v$ für alle n in D liegt. Dann gilt natürlich $x_n \to a$ für $n \to \infty$, und wegen $|t_n| = \frac{\|x_n - a\|}{\|v\|}$ haben wir

$$\lim_{n \to \infty} \frac{r(a+t_n\,v)}{|t_n|} = \|v\| \cdot \lim_{n \to \infty} \frac{r(x_n)}{\|x_n - a\|} = 0$$

nach (dem Folgenkriterium aus Satz 24.4 angewendet auf) Bedingung (b) aus Definition 25.3. Wiederum mit dem Folgenkriterium bedeutet dies nun aber $\lim_{t \to 0} \frac{r(a+t\,v)}{|t|} = 0$ und damit auch $(*) = 0$, was zu zeigen war. □

Folgerung 25.12 (Differenzierbare Funktionen sind partiell differenzierbar)**.** *Es sei $D \subset \mathbb{K}^n$ offen. Ist eine Funktion $f\colon D \to \mathbb{K}^m$ in einem Punkt $a \in D$ (total) differenzierbar, so ist f dort auch partiell differenzierbar, und es gilt*

$$f'(a) = (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,),$$

d. h. die Spalten der Ableitungsmatrix $f'(a)$ sind gerade die partiellen Ableitungen von f. Insbesondere ist die Ableitung $f'(a)$ also eindeutig bestimmt.

Beweis. Da die partiellen Ableitungen Spezialfälle der Richtungsableitungen sind, folgt die partielle Differenzierbarkeit sofort aus Lemma 25.11. Darüber hinaus ist die i-te Spalte der Matrix $f'(a)$ gerade $f'(a) \cdot e_i$ für alle $i = 1, \dots, n$, nach Lemma 25.11 also gleich $\partial_{e_i} f(a) = \partial_i f(a)$. □

64

Definition 25.13 (Jacobi-Matrix)**.** Ist die Funktion $f\colon D \to \mathbb{K}^m$ mit Komponentenfunktionen $f_1, \dots, f_m\colon D \to \mathbb{K}$ in einem Punkt $a \in D$ partiell differenzierbar, so nennt man die Matrix

$$Jf(a) := (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,) \;=\; (\partial_j f_i(a))_{i,j} \quad \in \mathrm{Mat}(m \times n, \mathbb{K})$$

aus Folgerung 25.12 die **Jacobi-Matrix** von f im Punkt a. Folgerung 25.12 besagt also gerade, dass die Ableitung einer differenzierbaren Funktion gleich ihrer Jacobi-Matrix sein muss.

Beispiel 25.14.

(a) Die Jacobi-Matrix der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 x_2 \\ x_1 \sin x_2 \end{pmatrix}$$

aus Beispiel 25.10 existiert in jedem Punkt und ist gleich

$$Jf(x) = (\,\partial_1 f(x) \mid \partial_2 f(x)\,) = \begin{pmatrix} x_2 & x_1 \\ \sin x_2 & x_1 \cos x_2 \end{pmatrix}.$$

Beachte jedoch, dass dies noch nicht zeigt, dass f auch differenzierbar ist, also dass $f'(a)$ existiert! In der Tat zeigt das folgende Beispiel, dass aus der partiellen Differenzierbarkeit im Allgemeinen noch nicht die totale Differenzierbarkeit folgt.

(b) (Partiell differenzierbare Funktionen müssen nicht differenzierbar sein) Wir betrachten noch einmal die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{cases} \frac{2x_1x_2}{x_1^2+x_2^2} & \text{für } (x_1,x_2) \neq (0,0), \\ 0 & \text{für } (x_1,x_2) = (0,0) \end{cases}$$

aus Beispiel 24.8. Wir haben dort gesehen, dass f im Nullpunkt nicht stetig, nach Lemma 25.7 also insbesondere auch nicht differenzierbar ist. Dennoch ist f dort partiell differenzierbar: Da f auf den Koordinatenachsen gleich 0 ist, ist natürlich

$$\partial_1 f(0) = \lim_{t\to 0} \frac{f(t\,e_1) - f(0)}{t} = 0,$$

und genauso $\partial_2 f(0) = 0$. Obwohl also die Jacobi-Matrix $Jf(0) = (0\ \ 0) \in \operatorname{Mat}(1\times 2, \mathbb{R})$ existiert, existiert die Ableitung $f'(0)$ in diesem Fall nicht.

Algorithmus 25.15 (Differenzierbarkeit einer Abbildung)**.** Mit Folgerung 25.12 können wir nun ein explizites Verfahren angeben, um zu bestimmen, ob eine gegebene Funktion $f\colon D \to \mathbb{K}^m$ in einem Punkt $a \in D$ differenzierbar ist, und in diesem Fall die Ableitung zu berechnen:

(a) Berechne mit den eindimensionalen Methoden aus Abschnitt 10.A die partiellen Ableitungen von f in a, und damit die Jacobi-Matrix $Jf(a) = (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,)$. Existiert eine dieser partiellen Ableitungen nicht, so sind wir fertig: Dann ist f nach Folgerung 25.12 nicht differenzierbar in a.

(b) Existiert die Jacobi-Matrix $Jf(a)$, so kommt nach Folgerung 25.12 nur sie als Ableitungsmatrix $f'(a)$ in Frage. Damit muss die Restfunktion in Definition 25.3 (a) gleich $r(x) = f(x) - f(a) - Jf(a)\cdot(x-a)$ sein, und gemäß Definition 25.3 (b) ist f genau dann in a differenzierbar, wenn

$$\lim_{x\to a} \frac{f(x) - f(a) - Jf(a)\cdot(x-a)}{\|x-a\|} = 0.$$

Diesen Grenzwert können wir genau wie in Abschnitt 24.A bestimmen.

Bemerkung 25.16 (Koordinatenweise Differenzierbarkeit im Zielraum)**.** Ist $f\colon D \to \mathbb{K}^m$ eine in einem Punkt $a \in D$ partiell differenzierbare Funktion mit Jacobi-Matrix $Jf(a)$ und Komponentenfunktionen $f_1,\dots,f_m\colon D \to \mathbb{K}$, so kann die Differenzierbarkeit von f in a in diesen Komponenten separat überprüft werden: Nach Algorithmus 25.15 ist f nämlich genau dann in a differenzierbar, wenn die dort in (b) angegebene Grenzwertbedingung gilt. Dieser Grenzwert kann nach Lemma 24.7 nun aber koordinatenweise in $\mathbb{K}^m$ überprüft werden — d. h. wir erhalten die äquivalenten Bedingungen

$$\lim_{x\to a} \frac{f_i(x) - f_i(a) - Jf_i(a)\cdot(x-a)}{\|x-a\|} = 0$$

für alle $i = 1,\dots,m$, die genau besagen, dass alle Komponentenfunktionen in a differenzierbar sind. Wie im Fall der Stetigkeit kann die Differenzierbarkeit jedoch nicht koordinatenweise im Startraum überprüft werden.

Auch wenn wir mit Algorithmus 25.15 jetzt eine Möglichkeit haben, eine Abbildung in mehreren Variablen auf Differenzierbarkeit zu prüfen, können die dafür nötigen Rechnungen in der Grenzwertbetrachtung (b) schnell sehr aufwändig werden. Wünschenswert wäre es daher, wenn wir genau wie im Eindimensionalen ein allgemeines Kriterium hätten, das uns auch ohne Rechnung sagt, dass „aus schönen Funktionen zusammengesetzte Abbildungen" immer differenzierbar sind. Ein solches Kriterium wollen wir jetzt beweisen. Da wir hierfür den nur im Reellen gültigen Mittelwertsatz 10.24 brauchen, beschränken wir uns dabei auf den Fall $\mathbb{K} = \mathbb{R}$.

Satz 25.17 (Stetig partiell differenzierbare Funktionen sind differenzierbar)**.** *Es seien $D \subset \mathbb{R}^n$ offen, $f\colon D \to \mathbb{R}^m$ eine Abbildung, und $a \in D$. Wir setzen voraus, dass alle partiellen Ableitungen $\partial_1 f, \dots, \partial_n f$ in einer Umgebung von a existieren und in a stetig sind — man sagt dann auch, dass f in a* ***stetig partiell differenzierbar*** *ist.*

Dann ist f in a auch total differenzierbar (mit Ableitung $f'(a) = Jf(a)$).

Beweis. Da sowohl die Stetigkeit als auch die Differenzierbarkeit nach Lemma 24.7 bzw. Bemerkung 25.16 koordinatenweise im Zielraum überprüft werden können, genügt es, den Fall $m = 1$ zu betrachten. Nach Algorithmus 25.15 müssen wir dann die Restfunktion

$$r(x) := f(x) - f(a) - Jf(a) \cdot (x-a) = f(x) - f(a) - \sum_{i=1}^{n} \partial_i f(a) \cdot (x_i - a_i) \tag{1}$$

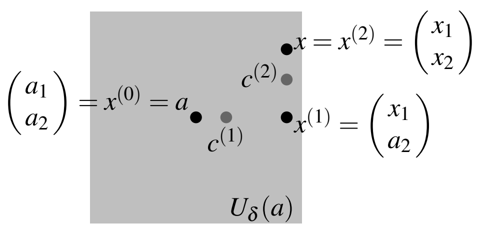
bilden und beweisen, dass $\lim\limits_{x \to a} \frac{r(x)}{\|x-a\|} = 0$ ist.

Wir zeigen dies direkt mit der Definition des Grenzwerts und verwenden dabei die Maximumsnorm auf $\mathbb{R}^n$. Es sei also $\varepsilon > 0$ gegeben. Da D offen und die partiellen Ableitungen nach Voraussetzung stetig sind, gibt es dann ein $\delta > 0$, so dass $U_\delta(a) \subset D$ gilt und

$$|\partial_i f(x) - \partial_i f(a)| < \frac{\varepsilon}{n} \quad \text{für alle } x \in U_\delta(a) \text{ und } i = 1, \dots, n. \tag{2}$$

Es sei nun $x \in U_\delta(a)$ beliebig. Wie im Bild rechts definieren wir für $i = 0, \dots, n$ dann die Punkte

$$x^{(i)} := \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} \in U_\delta(a),$$



so dass also $x^{(0)} = a$ und $x^{(n)} = x$ gilt, und jedes $x^{(i)}$ aus $x^{(i-1)}$ entsteht, indem die i-te Koordinate von a_i auf x_i abgeändert wird.

Damit können wir die Restfunktion (1) nun als

$$r(x) = \sum_{i=1}^{n} \Big(f(x^{(i)}) - f(x^{(i-1)}) - \partial_i f(a) \cdot (x_i - a_i) \Big) \tag{3}$$

umschreiben, da sich in der Summe der ersten beiden Terme alle Funktionswerte außer $f(x)$ und $f(a)$ wegheben. Wenden wir nun den Mittelwertsatz 10.24 (a) für $i = 1, \dots, n$ auf die zwischen a_i und x_i definierten reellwertigen Funktionen einer reellen Variablen

$$g_i\colon t \mapsto f \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ t \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} \quad \text{mit Ableitung} \quad g_i'\colon t \mapsto \partial_i f \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ t \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}$$

an, so erhalten wir t_i zwischen a_i und x_i und damit wie im Bild oben rechts $c^{(i)}$ auf den Verbindungsstrecken von a_i nach x_i mit

$$f(x^{(i)}) - f(x^{(i-1)}) = g(x_i) - g(a_i) = g'(t_i)\,(x_i - a_i) = \partial_i f(c^{(i)}) \cdot (x_i - a_i).$$

Setzen wir dies in (3) ein und beachten dabei, dass die Punkte $c^{(1)}, \dots, c^{(n)}$ in $U_\delta(a)$ liegen, so folgt daraus nun

$$\begin{aligned} |r(x)| &= \left| \sum_{i=1}^n (\partial_i f(c^{(i)}) - \partial_i f(a)) \cdot (x_i - a_i) \right| \le \sum_{i=1}^n |\partial_i f(c^{(i)}) - \partial_i f(a)| \cdot |x_i - a_i| \\ &\overset{(2)}{<} \sum_{i=1}^n \frac{\varepsilon}{n} \cdot \|x-a\|_\infty = \varepsilon \cdot \|x-a\|_\infty \end{aligned}$$

und damit $\frac{|r(x)|}{\|x-a\|_\infty} < \varepsilon$, woraus der behauptete Grenzwert $\lim_{x\to a} \frac{r(x)}{\|x-a\|_\infty} = 0$ folgt. □

Die Umkehrung von Satz 25.17 gilt natürlich nicht — sie ist ja schon im Eindimensionalen falsch, da wir in Aufgabe 10.36 bereits gesehen haben, dass eine differenzierbare Funktion auf $\mathbb{R}$ nicht stetig differenzierbar sein muss. Dieses Beispiel lässt sich auch problemlos auf den höherdimensionalen Fall übertragen:

Aufgabe 25.18 (Differenzierbare Funktionen müssen nicht stetig partiell differenzierbar sein)**.** Zeige, dass die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto \begin{cases} (x_1^2 + x_2^2)\,\sin \frac{1}{\sqrt{x_1^2+x_2^2}} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0 \end{cases}$$

(total) differenzierbar ist, die partiellen Ableitungen von f aber nicht stetig sind. Berechne auch die Ableitung f'!

Bemerkung 25.19. Insbesondere folgt aus Satz 25.17 also, dass alle Abbildungen differenzierbar sind, deren Komponentenfunktionen mit Hilfe der vier Grundrechenarten und Verkettungen aus stetig differenzierbaren Funktionen der Koordinaten des Startraums zusammengesetzt sind — denn für solche Abbildungen existieren ja nach den Rechenregeln aus Kapitel 10 dann alle partiellen Ableitungen und sind stetig. So ist z. B. die Funktion aus den Beispielen 25.10 und 25.14 (a) in jedem Punkt differenzierbar, und genauso die Abbildung aus den Beispielen 24.8 und 25.14 (b) in jedem Punkt mit Ausnahme des Nullpunkts.

Bemerkung 25.20 (Stetig differenzierbar = stetig partiell differenzierbar)**.** In Satz 25.17 haben wir eine Funktion $f\colon D \to \mathbb{R}^m$ auf einer offenen Teilmenge $D \subset \mathbb{R}^n$ aus naheliegenden Gründen *stetig partiell differenzierbar* genannt, wenn ihre partiellen Ableitungen existieren und stetig sind.

Genauso natürlich ist allerdings der schon aus Definition 11.8 (b) bekannte Begriff einer **stetig differenzierbaren** Funktion, der besagt, dass f (total) differenzierbar und die Ableitungsfunktion $f'\colon D \to \mathrm{Mat}(m \times n, \mathbb{R})$ stetig ist. Erfreulicherweise sind diese beiden Begriffe aufgrund unserer bisherigen Resultate äquivalent zueinander: Sind $f_1, \dots, f_m\colon D \to \mathbb{R}$ die Komponentenfunktionen von f, so gilt

$$\begin{aligned} f \text{ ist stetig differenzierbar } &\Leftrightarrow f \text{ ist differenzierbar und } f'\colon D \to \mathrm{Mat}(m\times n, \mathbb{R}) \text{ ist stetig} \\ &\overset{24.7}{\Leftrightarrow} f \text{ ist differenzierbar und alle } \partial_j f_i\colon D \to \mathbb{R} \text{ sind stetig} \\ &\overset{(*)}{\Leftrightarrow} f \text{ ist partiell differenzierbar und alle } \partial_j f_i\colon D \to \mathbb{R} \text{ sind stetig} \\ &\Leftrightarrow f \text{ ist stetig partiell differenzierbar,} \end{aligned}$$

wobei in $(*)$ die Richtung „$\Leftarrow$“ gerade Satz 25.17 ist. In der Regel verwendet man daher nur die einfachere der beiden Formulierungen und spricht von *stetig differenzierbaren* Funktionen.

Zusammenfassend haben wir damit nun also die folgenden Äquivalenzen bzw. Implikationen gezeigt (und gesehen, dass andere Implikationen zwischen diesen Begriffen nicht gelten):

$$\begin{array}{ccccc}
\text{stetig differenzierbar} & & & \overset{25.7}{\nearrow} & \text{stetig} \\
\text{25.20} \Updownarrow & \overset{25.17}{\searrow\!\!\nearrow} & \text{(total) differenzierbar} & & \\
\text{stetig partiell differenzierbar} & & & \underset{25.12}{\searrow} & \text{partiell differenzierbar}
\end{array}$$

Aufgabe 25.21. Untersuche, welche Richtungsableitungen der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto \begin{cases} \left(1-\cos\frac{x_1^2}{x_2}\right)\cdot\sqrt{x_1^2+x_2^2} & \text{für } x_2 \neq 0, \\ 0 & \text{für } x_2 = 0. \end{cases}$$

im Nullpunkt existieren, und ob f dort differenzierbar ist.

Aufgabe 25.22.

(a) Es sei $A \in \mathrm{Mat}(n\times n,\mathbb{R})$ eine quadratische Matrix. Zeige, dass die Abbildung

$$f\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto x^{\mathsf{T}}Ax$$

differenzierbar ist, und berechne ihre Ableitung.

(b) Zeige, dass die Abbildung $f\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \|x\|$ für *keine* Norm $\|\cdot\|$ auf $\mathbb{R}^n$ mit $n>0$ differenzierbar ist.

(c) Zeige, dass die Abbildung $f\colon \mathrm{Mat}(n\times n,\mathbb{R}) \mapsto \mathrm{Mat}(n\times n,\mathbb{R}),\ A \mapsto A^2$ differenzierbar ist, und berechne ihre Ableitung im Sinne von Bemerkung 25.6.

Aufgabe 25.23. Es sei $g\colon \mathbb{R}^2 \to \mathbb{R}$ eine stetige Abbildung. Zeige, dass $f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_1\,g(x)$ dann im Nullpunkt differenzierbar ist.

Aufgabe 25.24. Es sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ eine differenzierbare Abbildung mit $f(\lambda x) = \lambda f(x)$ für alle $\lambda \in \mathbb{R}$ und $x \in \mathbb{R}^n$. Zeige, dass f dann bereits eine lineare Abbildung ist.

Aufgabe 25.25 (*Parameterintegrale*)**.** Es seien $f\colon D \to \mathbb{R}$ eine stetige Abbildung auf einer offenen Teilmenge $D \subset \mathbb{R}^2$ sowie $[a,b]\times[c,d]$ ein in D enthaltenes Rechteck. Man zeige:

(a) Die Integralfunktion $F\colon [a,b] \to \mathbb{R},\ x_1 \mapsto \int_c^d f(x_1,x_2)\,dx_2$ ist stetig.

(b) Ist f stetig partiell nach x_1 differenzierbar, so ist F auf (a,b) differenzierbar mit Ableitung $F'(x_1) = \int_c^d \partial_1 f(x_1,x_2)\,dx_2$ (d. h. Differentiation und Integration nach verschiedenen Variablen können vertauscht werden).

(c) $\int_a^b \left(\int_c^d f(x_1,x_2)\,dx_2\right) dx_1 = \int_c^d \left(\int_a^b f(x_1,x_2)\,dx_1\right) dx_2$ (siehe auch Folgerung 28.17).

25.B Eigenschaften differenzierbarer Abbildungen

In diesem Abschnitt wollen wir differenzierbare Abbildungen im Mehrdimensionalen etwas genauer untersuchen und beginnen dazu mit einer geometrischen Deutung für den Fall, dass $\mathbb{K} = \mathbb{R}$ gilt und der Start- oder Zielraum eindimensional ist.

Bemerkung 25.26 (Geometrische Interpretation der Ableitung bei eindimensionalem Startraum)**.** Ist $f\colon D \to \mathbb{R}^m$ mit $D \subset \mathbb{R}$ offen, so können wir f als „Kurve in $\mathbb{R}^m$" ansehen, da f dann durch nur eine reelle Variable parametrisiert wird. Ist f nun differenzierbar in einem Punkt $a \in D$, so ist

$$f(x) \approx h(x) := f(a) + f'(a)\,(x-a)$$

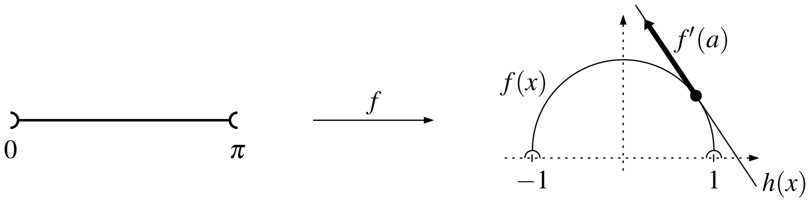
für x in der Nähe von a. Offensichtlich ist h die Parameterdarstellung einer Geraden in $\mathbb{R}^m$ mit Aufpunkt $f(a)$ und Richtungsvektor $f'(a) \in \mathrm{Mat}(m\times 1,\mathbb{R}) = \mathbb{R}^m$, und wir können uns diese Gerade als diejenige vorstellen, die die Kurve f beim Parameterwert a am besten approximiert. Mit anderen Worten ist h gerade die *Tangente* an die Kurve f im Punkt a. Ihr Richtungsvektor $f'(a)$, also die Ableitung von f in a, wird daher als **Tangentialvektor** von f in a bezeichnet.

Als konkretes Beispiel können wir wie im Bild unten den Halbkreisbogen

$$f\colon (0,\pi) \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \quad \text{mit Ableitung} \quad f'\colon (0,\pi) \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}$$

betrachten: Hier ist die Tangente an f im Punkt $a \in (0,\pi)$ wie im folgenden Bild gleich

$$h(x) = \begin{pmatrix} \cos a \\ \sin a \end{pmatrix} + \begin{pmatrix} -\sin a \\ \cos a \end{pmatrix} \cdot (x-a).$$



Bemerkung 25.27 (Geometrische Interpretation der Ableitung bei eindimensionalem Zielraum)**.** Auch für eine differenzierbare Abbildung $f\colon D \to \mathbb{R}$ mit offenem $D \subset \mathbb{R}^n$ können wir wieder mit unserer linearen Näherungsgleichung

$$f(x) \approx h(x) := f(a) + f'(a)\,(x-a)$$

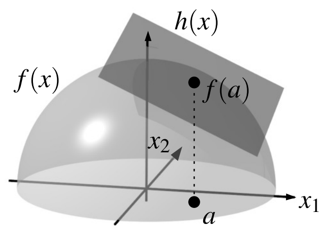
starten. In diesem Fall ist $f'(a)\,(x-a)$ jedoch ein Matrixprodukt einer $1 \times n$-Matrix mit einer $n \times 1$-Matrix, und der Graph $\{(x,h(x)) : x \in \mathbb{R}^n\}$ von h ist ein n-dimensionaler verschobener Unterraum in $\mathbb{R}^n \times \mathbb{R}$, der den Funktionsgraphen von f im Punkt a möglichst gut annähert.

Im Bild rechts ist dies für die Abbildung

$$f\colon D \to \mathbb{R},\ x \mapsto \sqrt{1-x_1^2-x_2^2}$$

mit Ableitung

$$f'(x) = \left(-\frac{x_1}{\sqrt{1-x_1^2-x_2^2}}, -\frac{x_2}{\sqrt{1-x_1^2-x_2^2}} \right)$$



auf dem offenen Einheitskreis $D = \{x \in \mathbb{R}^2 : \|x\|_2 < 1\}$ dargestellt, die offensichtlich eine Halbkugeloberfläche beschreibt. Wir können die Näherungsfunktion h hier als die *Tangentialebene* an den Graphen von f auffassen; in höheren Dimensionen $n > 2$ nennt man den durch h parametrisierten verschobenen Unterraum von $\mathbb{R}^n$ den **Tangentialraum** an den Graphen von f im Punkt a.

Beachte, dass die Ableitung in diesem Fall ein Zeilenvektor $f'(a) \in \mathrm{Mat}(1 \times n, \mathbb{R})$ ist, der damit zunächst einmal nicht wieder im Startraum $\mathbb{R}^n$ liegt. Man definiert daher im Fall eines eindimensionalen Zielraums in der Regel den **Gradienten** von f als den transponierten Vektor

$$\operatorname{grad} f(a) := f'(a)^{\mathsf{T}} \quad \in \mathbb{R}^n,$$

so dass sich die obige Näherungsgleichung mit Hilfe des Standardskalarprodukts auf $\mathbb{R}^n$ auch als

$$f(x) \approx f(a) + \langle \operatorname{grad} f(a), x-a \rangle$$

schreiben lässt.

In der Tat ermöglicht dies auch noch eine weitere geometrische Interpretation der Ableitung bzw. des Gradienten bei eindimensionalem Zielraum: Mit der Konstruktion 21.23 des Winkels zwischen zwei Vektoren können wir auch

$$f(x) \approx f(a) + \|\operatorname{grad} f(a)\|_2 \cdot \|x-a\|_2 \cdot \cos\alpha$$

schreiben, wobei α den Winkel zwischen den Vektoren $x-a$ und $\operatorname{grad} f(a)$ bezeichnet. Bewegt man sich also von a einen kleinen Schritt der Länge ε weg, d. h. geht man zu einem $x \in D$ mit $\|x-a\|_2 = \varepsilon$, so erzeugt dies in f näherungsweise eine Änderung von

$$\|\operatorname{grad} f(a)\|_2 \cdot \varepsilon \cdot \cos\alpha,$$

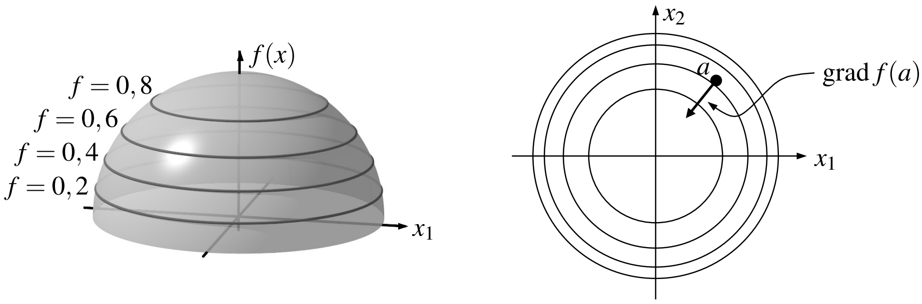
was (für festes ε) maximal wird für $\cos\alpha = 1$, d. h. für $\alpha = 0$, also wenn der gemachte Schritt $x-a$ und $\operatorname{grad} f(a)$ in dieselbe Richtung zeigen. Da die Änderung von f in diesem Fall in etwa gleich $\|\operatorname{grad} f(a)\|_2 \cdot \varepsilon$ ist, können wir den Gradienten also wie folgt geometrisch interpretieren:

> Die *Richtung* von $\operatorname{grad} f(a)$ ist die Richtung des stärksten Anstiegs von f im Punkt a.
> Der *Betrag* von $\operatorname{grad} f(a)$ ist die Stärke des Anstiegs von f in dieser Richtung.

Das Bild unten zeigt diesen Sachverhalt für die eben als Beispiel betrachtete Funktion. Sowohl im dreidimensionalen Bild links als auch im zweidimensionalen Bild rechts sind wie auf einer Landkarte die sogenannten *Höhenlinien* von f eingezeichnet — also die Kurven, auf denen f konstant ist. Der dort eingezeichnete Gradient ist ein Vektor, der in unserem Fall mit

$$\operatorname{grad} f(x) = -\frac{1}{\sqrt{1-x_1^2-x_2^2}} \cdot x$$

direkt zum Nullpunkt hin zeigt — also in die Richtung, in der f an dieser Stelle am stärksten ansteigt.



65

Wir wollen nun noch die Rechenregeln für Ableitungen aus Kapitel 10 so weit wie möglich auf den mehrdimensionalen Fall übertragen. Auch wenn wir unsere Berechnungen durch koordinatenweise Betrachtungen mit Hilfe der partiellen Ableitungen in den meisten Fällen bereits auf den eindimensionalen Fall zurückführen können, sind in der Praxis auch Rechenregeln nützlich, die direkt mit den Ableitungsmatrizen arbeiten und die koordinatenweise Berechnung der Ableitungen damit umgehen können.

Die folgende Aussage entspricht nahezu wörtlich der von Satz 10.8. Beachte aber, dass die Ableitungen jetzt *Matrizen* sind und keine einfachen Zahlen mehr. Wegen der neuen Definition der Differenzierbarkeit können die Beweise aus dem eindimensionalen Fall auch nicht einfach übernommen werden.

Satz 25.28 (Rechenregeln für Ableitungen)**.** *Es seien $D \subset \mathbb{K}^n$ offen und $f,g\colon D \to \mathbb{K}^m$ zwei Abbildungen, die in einem Punkt $a \in D$ differenzierbar sind. Dann gilt:*

(a) *$f \pm g$ ist differenzierbar in a mit $(f\pm g)'(a) = f'(a) \pm g'(a)$.*

(b) *Für $\lambda \in \mathbb{K}$ ist λf differenzierbar in a mit $(\lambda f)'(a) = \lambda f'(a)$.*

Im Fall $m = 1$, wenn Multiplikation und Division der Funktionen ebenfalls möglich sind, gilt zusätzlich:

(c) *(**Produktregel**) fg ist differenzierbar in a mit $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.*

(d) *(**Quotientenregel**) Ist $g(a) \neq 0$, so ist $\frac{f}{g}$ differenzierbar in a mit $(\frac{f}{g})'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}$.*

Beweis. Nach Definition 25.3 der Differenzierbarkeit können wir $f(x) = f(a) + f'(a)(x-a) + r(x)$ und $g(x) = g(a) + g'(a)(x-a) + s(x)$ mit $\frac{r(x)}{\|x-a\|} \to 0$ und $\frac{s(x)}{\|x-a\|} \to 0$ für $x \to a$ schreiben.

(a) Addition der beiden Ausdrücke für $f(x)$ und $g(x)$ liefert

$$(f+g)(x) = (f+g)(a) + (f'(a)+g'(a))(x-a) + r(x) + s(x).$$

Wegen

$$\lim_{x\to a} \frac{r(x)+s(x)}{\|x-a\|} = \lim_{x\to a} \frac{r(x)}{\|x-a\|} + \lim_{x\to a} \frac{s(x)}{\|x-a\|} = 0$$

ergibt sich also wiederum mit Definition 25.3, dass $f'(a)+g'(a)$ die Ableitung von $f+g$ in a ist. Die Aussagen für die Differenz $f-g$ sowie für λf in (b) zeigt man natürlich genauso.

(c) Hier ergibt die Multiplikation

$$\begin{aligned}(fg)(x) &= (fg)(a) + (f'(a)g(a) + f(a)g'(a))(x-a) \\ &\quad + \underbrace{(f'(a)(x-a))(g'(a)(x-a))}_{\text{(A)}} + \underbrace{(\text{Terme mit } r(x) \text{ und / oder } s(x))}_{\text{(B)}}.\end{aligned}$$

Beachte, dass der Ausdruck $f'(a)(x-a)$ in (A) die Multiplikation einer $(1\times n)$-Matrix mit einer $(n\times 1)$-Matrix ist, die wir daher auch mit Hilfe des Gradienten und des Standardskalarprodukts als $\langle \operatorname{grad} f(a), x-a\rangle$ schreiben können (siehe Bemerkung 25.27). Analog gilt dies für $g'(a)(x-a)$, und damit folgt nach der Cauchy-Schwarz-Ungleichung aus Satz 21.20

$$\|(\mathrm{A})\|_2 \le \|\operatorname{grad} f(a)\|_2 \cdot \|\operatorname{grad} g(a)\|_2 \cdot \|x-a\|_2^2, \quad \text{d. h.} \quad \frac{(\mathrm{A})}{\|x-a\|_2} \to 0 \text{ für } x \to a.$$

Genauso gilt wegen der enthaltenen Faktoren $r(x)$ bzw. $s(x)$ natürlich $\frac{(\mathrm{B})}{\|x-a\|} \to 0$ für $x \to a$. Also können wir hier (A) + (B) als Restterm auffassen, so dass mit Definition 25.3 folgt, dass $f'(a)g(a) + f(a)g'(a)$ die Ableitung von fg in a ist.

Der Beweis von (d) verläuft ganz analog und soll daher hier nicht gegeben werden. □

Auch die Kettenregel (siehe Satz 10.10) überträgt sich wie erwartet ins Mehrdimensionale.

Satz 25.29 (Kettenregel). *Es seien $D \subset \mathbb{K}^n$ und $D' \subset \mathbb{K}^m$ offen. Ferner seien $f\colon D \to \mathbb{K}^m$ und $g\colon D' \to \mathbb{K}^p$ Abbildungen mit $f(D) \subset D'$. Ist dann f differenzierbar in $a \in D$ und g differenzierbar in $f(a)$, so ist auch die Verkettung $g \circ f$ differenzierbar in a, und es gilt*

$$(g\circ f)'(a) = g'(f(a)) \cdot f'(a),$$

d. h. „die Ableitung einer Verkettung ist das Produkt der beiden Ableitungen". (Beachte, dass es sich hierbei um ein Matrixprodukt einer $p \times m$-Matrix mit einer $m \times n$-Matrix handelt, das wie gewünscht eine $p \times n$-Matrix als Ableitung von $g \circ f$ liefert — insbesondere ist die Reihenfolge der beiden Faktoren hier also wichtig!)

Beweis. Wegen der Differenzierbarkeit von f in a und g in $f(a)$ gibt es nach Definition 25.3 Funktionen $r\colon D \to \mathbb{K}^m$ und $s\colon D' \to \mathbb{K}^p$ mit

$$f(x) = f(a) + f'(a)\,(x-a) + r(x) \tag{1}$$

$$\text{und} \qquad g(y) = g(f(a)) + g'(f(a))\,(y-f(a)) + s(y) \tag{2}$$

sowie $\lim_{x\to a} \frac{r(x)}{\|x-a\|} = 0$ und $\lim_{y\to f(a)} \frac{s(y)}{\|y-f(a)\|} = 0$. Insbesondere können wir also die Funktion

$$\tilde{s}\colon D'\backslash\{f(a)\} \to \mathbb{K}^p,\ y \mapsto \frac{s(y)}{\|y-f(a)\|}$$

durch $\tilde{s}(f(a)) := 0$ zu einer stetigen Funktion auf ganz D' fortsetzen und damit statt (2) auch

$$g(y) = g(f(a)) + g'(f(a))\,(y-f(a)) + \tilde{s}(y) \cdot \|y-f(a)\|$$

für alle $y \in D'$ schreiben. Setzen wir hier nun $y = f(x)$ ein, so ergibt sich mit (1)

$$\begin{aligned} g(f(x)) &= g(f(a)) + g'(f(a))\,(f(x)-f(a)) + \tilde{s}(f(x)) \cdot \|f(x)-f(a)\| \\ &= g(f(a)) + g'(f(a))\,f'(a)\,(x-a) + \underbrace{g'(f(a))\,r(x) + \tilde{s}(f(x)) \cdot \|f'(a)\,(x-a) + r(x)\|}_{=:R(x)} \end{aligned}$$

für alle $x \in D$. Verwenden wir für die auftretenden Matrizen die von der gewählten Vektornorm induzierte Matrixnorm, so können wir den Restterm R mit Folgerung 24.38 nun abschätzen durch

$$\frac{\|R(x)\|}{\|x-a\|} \le \|g'(f(a))\| \cdot \underbrace{\frac{\|r(x)\|}{\|x-a\|}}_{\to 0} + \underbrace{\|\tilde{s}(f(x))\|}_{\to 0} \cdot \|f'(a)\| \cdot \frac{\|x-a\|}{\|x-a\|} + \underbrace{\|\tilde{s}(f(x))\|}_{\to 0} \cdot \underbrace{\frac{\|r(x)\|}{\|x-a\|}}_{\to 0},$$

wobei die markierten Terme für $x \to a$ gegen 0 konvergieren (beachte dabei für den Term $\tilde{s}(f(x))$, dass f stetig in a und $\tilde{s}$ stetig in $f(a)$ ist, so dass $\tilde{s}(f(x))$ mit $x \to a$ gegen $\tilde{s}(f(a)) = 0$ konvergiert). Also gilt $\lim_{x\to a} \frac{\|R(x)\|}{\|x-a\|} = 0$ und damit auch $\lim_{x\to a} \frac{R(x)}{\|x-a\|} = 0$. Die Aussage des Satzes folgt damit aus Definition 25.3. □

Bemerkung 25.30. Auch ohne sich den genauen Beweis von Satz 25.29 anzusehen, ist einfach zu verstehen, warum bei der Formel für die Ableitung einer Verkettung Matrixprodukte auftreten: Die Ableitungen können wir ja gerade als lineare Näherungen der gegebenen Funktionen betrachten — und lineare Funktionen werden nach Lemma 16.12 verkettet, indem man die zugehörigen Matrizen miteinander multipliziert. In Formeln bedeutet dies unter Vernachlässigung der Restterme folgendes: Können wir als lineare Näherungen

$$f(x) \approx f(a) + f'(a)(x-a) \quad \text{und} \quad g(f(x)) \approx g(f(a)) + g'(f(a))(f(x) - f(a))$$

schreiben, so ergibt sich daraus durch Einsetzen

$$g(f(x)) \approx g(f(a)) + g'(f(a))f'(a)(x-a),$$

d. h. $g'(f(a))f'(a)$ ist die Ableitung von $g \circ f$ in a. (Die genaue Behandlung der Restterme, also die Präzisierung des Zeichens „$\approx$“, ist dann natürlich die eigentliche Arbeit beim Beweis.)

Beispiel 25.31.

(a) Wir betrachten noch einmal die geometrische Deutung der Ableitung bei eindimensionalem Zielraum aus Bemerkung 25.27. Es seien dazu $D \subset \mathbb{R}^n$ offen, $f\colon D \to \mathbb{R}$ eine differenzierbare Funktion, und $\gamma\colon (a,b) \to D$ eine differenzierbare Höhenlinie von f, d. h. so dass $f \circ \gamma\colon (a,b) \to \mathbb{R}$ konstant ist. Dann ist die Ableitung dieser Verkettung natürlich in jedem Punkt $t \in (a,b)$ gleich 0, und wir erhalten somit nach Satz 25.29

$$0 = (f \circ \gamma)'(t) = f'(\gamma(t)) \cdot \gamma'(t) = \langle \operatorname{grad} f(\gamma(t)), \gamma'(t) \rangle.$$

Also ist $\operatorname{grad} f(\gamma(t)) \perp \gamma'(t)$, d. h. der Gradient von f steht in jedem Punkt senkrecht auf (dem Tangentialvektor) der Höhenlinie durch diesen Punkt. Diese Tatsache, die wir im Beispiel von Bemerkung 25.27 auch sofort im Höhenlinienbild erkennen können, ist natürlich jedem bekannt, der schon einmal eine Landkarte mit Höhenlinien benutzt hat: Um möglichst schnell nach oben zu kommen, also in Richtung des Gradienten zu laufen, geht man am besten senkrecht zu den Höhenlinien.

(b) Es sei f die Funktion

$$f\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto \int_1^x \frac{\sin(xt)}{t}\, dt.$$

Offensichtlich handelt es sich hierbei um eine Funktion mit eindimensionaler Start- und Zielmenge, wie wir sie schon vor langer Zeit betrachtet haben. Man kann zeigen, dass dieses Integral nicht mit den uns bekannten speziellen Funktionen aus Kapitel 9 berechenbar ist, so dass wir f nicht auf einfachere Art hinschreiben können. Dennoch können wir aber die Ableitung von f mit der *mehrdimensionalen* Kettenregel konkret bestimmen: Dazu schreiben wir $f = g \circ h$ mit

$$g\colon \mathbb{R}^2_{>0} \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \int_1^{x_1} \frac{\sin(x_2 t)}{t}\, dt \quad \text{und} \quad h\colon \mathbb{R}_{>0} \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} x \\ x \end{pmatrix}.$$

Dann lässt sich g (und natürlich auch h) problemlos differenzieren: Nach dem Hauptsatz 12.21 der Differential- und Integralrechnung liefert die Ableitung des Integrals nach seiner

Obergrenze gerade

$$\partial_1 g \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{\sin(x_2 x_1)}{x_1}.$$

Die partielle Ableitung $\partial_2 g$ dagegen lässt sich nach Aufgabe 25.25 unter dem Integral berechnen, so dass wir dafür

$$\partial_2 g = \int_1^{x_1} \partial_2 \left(\frac{\sin(x_2 t)}{t} \right) dt = \int_1^{x_1} \cos(x_2 t)\, dt = \frac{1}{x_2} \left(\sin(x_2 x_1) - \sin x_2\right)$$

erhalten. Beide partiellen Ableitungen sind offensichtlich stetig, so dass g nach Satz 25.17 differenzierbar ist. Damit ergibt die Kettenregel aus Satz 25.29

$$\begin{aligned} f'(x) = g'(h(x))\, h'(x) &= \left(\frac{\sin(x^2)}{x}, \frac{\sin(x^2) - \sin x}{x} \right) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{2\sin(x^2)}{x} - \frac{\sin x}{x}. \end{aligned}$$

Aufgabe 25.32. Es sei $f\colon \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$ mit $n \geq 2$ eine differenzierbare Funktion. Man zeige:

(a) Hängt f nur von $\|x\|_2$ ab, ist also $f(x) = g(\|x\|_2)$ für eine differenzierbare Funktion $g\colon \mathbb{R}_{>0} \to \mathbb{R}$ (man sagt auch: f ist *kugelsymmetrisch*), so ist

$$\operatorname{grad} f(x) = \frac{g'(\|x\|_2)}{\|x\|_2} \cdot x \quad \text{für alle } x \in \mathbb{R}^n \backslash \{0\}.$$

(b) Gibt es umgekehrt eine Funktion $h\colon \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$ mit $\operatorname{grad} f(x) = h(x) \cdot x$ für alle x, so ist f kugelsymmetrisch. (Hinweis: Zeige, dass f entlang eines beliebigen Weges auf einer Kugeloberfläche konstant ist.)