Kapitel 10

Ableitungen nach mehreren Variablen

10.1 Partielle Ableitungen

Wir wollen in diesem Kapitel die Differentialrechnung für auf $D \subseteq \mathbb{R}^n$ definierte Funktionen entwickeln. Dazu bezeichnen wir im folgenden mit e_i den i-ten kanonischen Basisvektor $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ mit der 1 an der i-ten Stelle.

10.1.1 Definition. Sei $(X, \|.\|)$ ein Banachraum[1], und $f : D \to X$ sei definiert auf einer offenen Menge $D \subseteq \mathbb{R}^n$. Existiert für ein $x \in D$ der Grenzwert

$$\frac{\partial f}{\partial x_i}(x) := \lim_{s \to 0} \frac{1}{s}(f(x + se_i) - f(x)), \tag{10.1}$$

so heißt dieser die *partielle Ableitung* von f nach der Variablen x_i an der Stelle x. Anstatt $\frac{\partial f}{\partial x_i}(x)$ schreibt man auch $D_i f(x)$ oder $f_{x_i}(x)$.
Ist allgemeiner $v = (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$ ein sogenannter *Richtungsvektor*[2], und existiert für $x \in D$ der Grenzwert

$$\frac{\partial f}{\partial v}(x) := \lim_{s \to 0} \frac{1}{s}(f(x + sv) - f(x)),$$

so heißt dieser die *Richtungsableitung* von f an der Stelle x in Richtung v.

Offensichtlich gilt für den Richtungsvektor e_i

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x).$$

10.1.2 Bemerkung. Für $x \in D$ und $v \in \mathbb{R}^n$ ist die Abbildung $t \mapsto x + tv$ von $\mathbb{R}$ nach $\mathbb{R}^n$ stetig, wobei 0 auf x abgebildet wird. Da $D \subseteq \mathbb{R}^n$ offen ist, gibt es infolge ein reelles Intervall $(-\delta, \delta)$ mit $\delta > 0$ um die reelle Null, sodass $x + tv \in D$ für alle $t \in (-\delta, \delta)$. Setzt man

$$g(t) := f(x + tv), \quad t \in (-\delta, \delta),$$

[1] Man sollte anfangs einfach an $\mathbb{R}$ oder $\mathbb{R}^m$ denken.
[2] $v \in \mathbb{R}^n$ muss nicht normiert sein und kann sogar der Nullvektor sein.

so bildet g das Intervall $(-\delta, \delta)$ nach X hinein ab. Wegen

$$\frac{1}{s}(f(x+(t+s)v) - f(x+tv)) = \frac{1}{s}(g(t+s) - g(t))$$

ist g genau dann bei t differenzierbar, wenn $\frac{\partial f}{\partial v}(x+tv)$ existiert. In dem Fall gilt

$$g'(t) = \frac{\partial f}{\partial v}(x+tv)\,.$$

Richtungsableitungen sind ihrer Natur nach in einem gewissen Sinn eindimensional. Denn man betrachtet ja f eingeschränkt auf die Strecke $x + (-\delta, \delta)v$. Insbesondere variiert bei der Bildung der partiellen Ableitung nach x_i wie in (10.1) beim Differenzenquotienten $\frac{f(x+se_i)-f(x)}{s}$ nur die i-te Koordinate.

Abbildung 10.1: Ableitung in Richtung der Koordinate x_1 bzw. des Vektors v

10.1.3 Definition. Sei $f : D \subseteq \mathbb{R}^n \to X$ definiert auf der offenen Menge $D \subseteq \mathbb{R}^n$. Existieren für alle $x \in D$ alle partiellen Ableitungen $\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)$, so heißt f *partiell differenzierbar* auf D.
Sind alle X-wertigen Funktionen $x \mapsto \frac{\partial f}{\partial x_j}(x),\ j = 1, \dots, n$, darüber hinaus stetig auf D, so heißt f *stetig partiell differenzierbar*. Wir schreiben $f \in C^1$ bzw. $f \in C^1(D)$ dafür.

10.1.4 Definition. Teilmengen A von Vektorräumen Y (über $\mathbb{R}$ oder $\mathbb{C}$) heißen *konvex*, falls für alle $x, y \in A$ folgt, dass auch alle Punkte auf ihrer Verbindungsgeraden in A liegen, also $tx + (1-t)y \in A,\ t \in [0, 1]$.

10.1.5 Bemerkung. Ist Y sogar mit einer Norm $\|.\|$ versehen und ist $A = U_\epsilon(z)$ oder $A = K_\epsilon(z)$ die offene bzw. abgeschlossene Kugel vom Radius ϵ, so ist A konvex, denn aus den aus den Ungleichungen $\|z - x\|, \|z - y\| < (\leq)\epsilon$ folgt

$$\begin{aligned}\|z - (tx + (1-t)y)\| &= \|t(z-x) + (1-t)(z-y)\| \\ &\leq t\|z-x\| + (1-t)\|z-y\| < (\leq)\ t\epsilon + (1-t)\epsilon = \epsilon\,.\end{aligned}$$

10.1.6 Lemma. *Ist $f : D \subseteq \mathbb{R}^n \to X$ stetig partiell differenzierbar auf D, so existiert für jedes $x \in D$ und jeden Richtungsvektor $v = (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^n$ die Richtungsableitung $\frac{\partial f}{\partial v}(x)$, wobei*

$$\frac{\partial f}{\partial v}(x) = \sum_{i=1}^{n} \mu_i \frac{\partial f}{\partial x_i}(x)\,. \tag{10.2}$$

Insbesondere ist die X-wertige Funktion $x \mapsto \frac{\partial f}{\partial v}(x)$ stetig auf D.

Beweis. Wähle $\delta > 0$, sodass $x + tv \in D$ für $t \in (-\delta, \delta)$. Nach Bemerkung 10.1.2 müssen wir zeigen, dass die Funktion $g : (-\delta, \delta) \to X$ definiert durch $g(t) := f(x + tv)$ bei 0 differenzierbar ist, wobei

$$g'(0) = \sum_{i=1}^{n} \mu_i \frac{\partial f}{\partial x_i}(x) .$$

Für $v = 0$ ist g konstant und infolge die Aussage richtig. Anderenfalls wählen wir $\rho > 0$ so klein, dass[3] $U_{\rho \cdot \|v\|_\infty}(x) \subseteq D$. Betrachte die Vektoren

$$v_0 := 0,\ v_i := \mu_1 e_1 + \ldots + \mu_i e_i,\ \ i = 1, \ldots, n .$$

Offensichtlich gilt $v_n = v$ sowie $v_i = v_{i-1} + \mu_i e_i$ für $i = 1, \ldots, n$. Da die ganze Kugel mit Radius $\rho \cdot \|v\|_\infty$ und Mittelpunkt x in D enthalten ist, liegen die Punkte $x + sv_i$ für $i = 1, \ldots, n$ und $|s| < \rho$ in D. Gemäß Bemerkung 10.1.5 sind Kugeln konvex. Also liegen auch alle Verbindungsstrecken zwischen diesen Punkten in D, womit für jedes $i = 1, \ldots, n$ die Funktion $\tau \mapsto f(x + sv_{i-1} + \tau e_i)$ für $\tau \in [\min(0, s\mu_i), \max(0, s\mu_i)]$ wohldefiniert ist. Ihre Ableitung stimmt mit $\frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i)$ überein, und mit dem Hauptsatz der Differential- und Integralrechnung (vgl. (9.15)) folgt

$$f(x + sv_i) - f(x + sv_{i-1}) = \int_0^{s\mu_i} \frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i)\, d\tau .$$

Für $|s| < \rho$ erhalten wir

$$\begin{aligned} \frac{g(s) - g(0)}{s} - \sum_{i=1}^{n} \mu_i \frac{\partial f}{\partial x_i}(x) &= \frac{1}{s}\Big(f(x + sv) - f(x) - \sum_{i=1}^{n} s\mu_i \frac{\partial f}{\partial x_i}(x)\Big) \\ &= \frac{1}{s} \sum_{i=1}^{n} \big(f(x + sv_i) - f(x + sv_{i-1}) - s\mu_i \frac{\partial f}{\partial x_i}(x)\big) \\ &= \frac{1}{s} \sum_{i=1}^{n} \int_0^{s\mu_i} \left(\frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i) - \frac{\partial f}{\partial x_i}(x)\right) d\tau . \end{aligned} \tag{10.3}$$

Wegen (9.12) ist die Norm dieses Ausdruckes kleiner oder gleich

$$\sum_{i=1}^{n} |\mu_i| \sup_{\tau \in [-|s\mu_i|, |s\mu_i|]} \left\| \frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i) - \frac{\partial f}{\partial x_i}(x) \right\| .$$

Wegen der Stetigkeit aller partiellen Ableitungen in x konvergiert dieser Ausdruck gegen 0, wenn s gegen 0 läuft.
Die Stetigkeit von $x \mapsto \frac{\partial f}{\partial v}(x)$ folgt wegen der vorausgesetzten Stetigkeit der Abbildungen $x \mapsto \frac{\partial f}{\partial x_i}(x)$ sofort aus (10.2). ❑

[3] Wir versehen hier praktischerweise $\mathbb{R}^n$ mit $\|.\|_\infty$.

10.1.7 Bemerkung. Wir haben im Beweis von Lemma 10.1.6 die Stetigkeit der partiellen Ableitungen (in allen Variablen gleichzeitig) benützt, um zu zeigen, dass (10.3) gegen Null strebt, wenn $s \to 0$.
Beachte, dass etwa die Funktion

$$f\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{cases} \frac{\xi\eta}{\xi^2+\eta^2}\,, & \text{falls } (\xi,\eta)^T \neq (0,0)^T\,,\\ 0\,, & \text{falls } (\xi,\eta)^T = (0,0)^T\,,\end{cases}$$

aus Beispiel 6.1.11 stets Richtungsableitungen bei 0 in Richtung der Koordinatenachsen besitzt, die Richtungsableitung in Richtung $v = e_1 + e_2$ jedoch nicht existiert. Man kann also in Lemma 10.1.6 die Voraussetzung der Stetigkeit der $\frac{\partial f}{\partial x_i}$ nicht weglassen.

Die Tatsache, dass für festes x die Richtungsableitung $\frac{\partial f}{\partial v}(x)$ in (10.2) linear von v abhängt, legt folgende Begriffsbildung nahe.

10.1.8 Definition. Sei $f : D \subseteq \mathbb{R}^n \to X$ stetig partiell differenzierbar, und $x \in D$. Dann bezeichne $df(x) \in L(\mathbb{R}^n, X)$ die lineare Abbildung

$$\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix} \mapsto \sum_{i=1}^{n} \mu_i \frac{\partial f}{\partial x_i}(x)\,.$$

Sie heißt die *Ableitung* von f im Punkt x.

Nach Lemma 10.1.6 ist $df(x)v$ nichts anderes als die Richtungsableitung $\frac{\partial f}{\partial v}(x)$ von f in x in Richtung v. Insbesondere gilt $df(x)e_i = \frac{\partial f}{\partial x_i}(x)$.

10.1.9 Fakta.

1. Gemäß Beispiel 9.2.10, (ii), sind alle linearen $A : \mathbb{R}^n \to X$ auch beschränkt, also $L(\mathbb{R}^n, X) = L_b(\mathbb{R}^n, X)$. Somit ist $df(x)$ beschränkt und infolge auch stetig; vgl. Satz 9.2.6.

2. Ist $f : D \subseteq \mathbb{R}^n \to X$ eine affine Abbildung, also von der Form $x \mapsto x_0 + Ax$ mit $x_0 \in X$ und einer linearen Abbildung $A : \mathbb{R}^n \to X$, so gilt

$$df(x)e_i = \lim_{s\to 0} \frac{x_0 + A(x + se_i) - (x_0 + A(x))}{s} = \lim_{s\to 0} \frac{A(te_i)}{s} = A(e_i)\,,$$

   und damit $df(x) = A$ für alle $x \in D$. Das ist eine Verallgemeinerung der Tatsache, dass im eindimensionalen die Ableitung einer linearen Funktion $kx + d$ konstant gleich der Steigung k ist.

3. Für $D \subseteq \mathbb{R}$, also $n = 1$, ist $df(x)$ gerade jene lineare Abbildung $\mathbb{R} \to X$, die 1 auf $f'(x)$ abbildet.

4. Wir wissen, dass Netze in $\mathbb{R}^m$ genau dann konvergieren, wenn sie komponentenweise konvergieren. Für $X = \mathbb{R}^m$ (versehen mit zB. $\|.\|_\infty$) ist somit die Tatsache, dass eine Funktion $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \dots, f_m)^T$, (stetig) partiell differenzierbar ist, äquivalent dazu, dass jede ihrer Komponenten f_k (stetig) partiell differenzierbar ist.

5. Ist $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^m$ stetig partiell differenzierbar, und $x \in D$, und zerlegen wir $f(x)$ in seine Komponenten, $f(x) = (f_1(x), \dots, f_m(x))^T$, so hat $df(x) \in L(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{m\times n}$ als Matrix die Darstellung
$$df(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{\substack{i=1,\dots,m\\ j=1,\dots,n}} .$$
Ist dabei $m = 1$, also f $\mathbb{R}$-wertig, dann nennt man den Vektor $df(x)^T$ auch den *Gradienten* und schreibt auch $\operatorname{grad} f(x)$ dafür.

10.1.10 Beispiel. Sei $D = \mathbb{R}^+ \times \mathbb{R} \subseteq \mathbb{R}^2$ und $T : D \to \mathbb{R}^2$ definiert durch
$$T\begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} r\ \cos t \\ r\ \sin t \end{pmatrix} .$$
Diese Polarkoordinatenfunktion kennen wir schon aus dem ersten Semester; siehe Definition 6.9.11 und Bemerkung 6.9.12. Nun gilt offensichtlich
$$\frac{\partial T}{\partial r}\begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \quad \text{und} \quad \frac{\partial T}{\partial t}\begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} -r\ \sin t \\ r\ \cos t \end{pmatrix} .$$
Also sind $\frac{\partial T}{\partial r}$ und $\frac{\partial T}{\partial t}$ stetig, und damit T stetig partiell differenzierbar. Die Ableitung $dT((r,t)^T) \in L(\mathbb{R}^2, \mathbb{R}^2)$ ist damit
$$dT\begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} \cos t & -r\ \sin t \\ \sin t & r\ \cos t \end{pmatrix} , \tag{10.4}$$
und die Richtungsableitung von T im Punkt $(r,t)^T$ in Richtung $v = (\mu_1, \mu_2)^T$ ist nach Lemma 10.1.6
$$\frac{\partial T}{\partial v}\begin{pmatrix} r \\ t \end{pmatrix} = dT\begin{pmatrix} r \\ t \end{pmatrix}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \cos t - \mu_2\ r\ \sin t \\ \mu_1 \sin t + \mu_2\ r\ \cos t \end{pmatrix} .$$
Berechnen wir diese Richtungsableitung gemäß Bemerkung 10.1.2 direkt dadurch, dass wir $g'(0)$ mit $g(s) = T((r,t)^T + s(\mu_1, \mu_2)^T)$ bestimmen, so erhalten wir auch
$$g'(0) = \begin{pmatrix} \frac{d}{ds}((r + s\mu_1)\ \cos(t + s\mu_2))|_{s=0} \\ \frac{d}{ds}((r + s\mu_1)\ \sin(t + s\mu_2))|_{s=0} \end{pmatrix} =$$
$$\begin{pmatrix} (\mu_1 \cos(t + s\mu_2) - (r + s\mu_1)\mu_2 \sin(t + s\mu_2))|_{s=0} \\ (\mu_1 \sin(t + s\mu_2) + (r + s\mu_1)\mu_2 \cos(t + s\mu_2))|_{s=0} \end{pmatrix} = \begin{pmatrix} \mu_1 \cos t - \mu_2\ r\ \sin t \\ \mu_1 \sin t + \mu_2\ r\ \cos t \end{pmatrix} .$$

10.1.11 Proposition. *Sei $f : D \subseteq \mathbb{R}^n \to X$ stetig partiell differenzierbar.*

(*i*) *Versieht man $L(\mathbb{R}^n, X) = L_b(\mathbb{R}^n, X)$ mit der Abbildungsnorm, wobei $\mathbb{R}^n$ mit[4] $\|.\|_\infty$ und X mit der darauf gegebenen Norm $\|.\|$ versehen wird, so ist die Abbildung*

$$df : D \subseteq \mathbb{R}^n \to L_b(\mathbb{R}^n, X),\ x \mapsto df(x)$$

stetig.

(*ii*) *Für festes $x \in D$ und $x + z \in D \setminus \{x\}$ gilt*

$$f(x+z) = f(x) + df(x)z + \|z\|_\infty \epsilon(z)\,, \tag{10.5}$$

mit einer Funktion[5] $\epsilon : (D \setminus \{x\} - x) \to X$, sodass $\lim_{z\to 0} \epsilon(z) = 0$.

Beweis.

(*i*) Für $x, y \in D$ gilt

$$\begin{aligned} \|\, df(x) - df(y) \,\| &= \sup_{\|v\|_\infty \le 1} \|\, df(x)v - df(y)v \,\| \\ &= \sup_{|\mu_1|,\dots,|\mu_n| \le 1} \left\| \sum_{j=1}^n \mu_j (\, df(x)e_j - df(y)e_j \,) \right\| \\ &\le \sum_{j=1}^n \left\| \frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(y) \right\| . \end{aligned}$$

Wegen der vorausgesetzten Stetigkeit der partiellen Ableitungen folgt aus dieser Abschätzung $df(y) \to df(x)$ für $y \to x$.

(*ii*) Mit festem $x \in D$ und beliebigem $z \neq 0$, sodass $x + z \in D$, setzen wir

$$\epsilon(z) := \frac{1}{\|z\|_\infty} \left(f(x+z) - f(x) - df(x)z \right) .$$

Ist $\rho > 0$, sodass $U_\rho(x) = x + U_\rho(0) \subseteq D$ und $0 < \|z\| < \rho$, so ist $\epsilon(z)$ definiert, und wir erhalten wegen (10.3) mit $s = \|z\|_\infty$, $v = (\mu_1, \dots, \mu_n)^T := \frac{z}{\|z\|_\infty}$, wobei offensichtlich $\max_{j=1,\dots,n} |\mu_j| = \|v\|_\infty = 1$,

$$\begin{aligned} \|\, \epsilon(z) \,\| &= \left\| \sum_{i=1}^n \frac{1}{s} \int_0^{s\mu_i} \left(\frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i) - \frac{\partial f}{\partial x_i}(x) \right) d\tau \right\| \\ &\le \sum_{i=1}^n |\mu_i| \sup_{\tau \in [-|s\mu_i|, |s\mu_i|]} \left\| \frac{\partial f}{\partial x_i}(x + sv_{i-1} + \tau e_i) - \frac{\partial f}{\partial x_i}(x) \right\| \\ &\le \sum_{i=1}^n \sup_{\|y\|_\infty \le \|z\|_\infty} \left\| \frac{\partial f}{\partial x_i}(x+y) - \frac{\partial f}{\partial x_i}(x) \right\| . \end{aligned}$$

[4] Versieht man $\mathbb{R}^n$ dagegen mit $\|.\|_1$, $\|.\|_2$ oder einer anderen äquivalenten Norm, so ist die resultierende Abbildungsnorm auf $L_b(\mathbb{R}^n, X)$ zwar eine andere aber zur eingangs definierten Abbildungsnorm auf $L_b(\mathbb{R}^n, X)$ äquivalent.

[5] Man beachte, dass $D \setminus \{x\} - x$ eine Menge der Bauart $U_\delta(0) \setminus \{0\}$ enthält.

Dieser Ausdruck konvergiert wegen der Stetigkeit der partiellen Ableitungen für $z \to 0$ gegen 0. ❑

10.1.12 Bemerkung. Wir sehen also, dass sich f lokal bei x durch die affine Abbildung $x + z \mapsto f(x) + df(x)z$ approximieren lässt, wobei der Fehler mit $\|z\|_\infty \epsilon(z)$ verhältnismäßig klein ist.
Ist $X = \mathbb{R}^m$, und stellt man sich $f(D)$ als n-dimensionale Fläche im $\mathbb{R}^m$ vor, so ist der affine Teilraum

$$\{f(x) + df(x)z : z \in \mathbb{R}^n\}$$

der Anschauung nach gerade die *Tangentialebene* an $f(D)$ im Punkt $f(x)$.

Man kann die Differentialrechnung in mehreren Variablen auch ausgehend von den in Proposition 10.1.11 genannten Eigenschaften hochziehen.

10.1.13 Definition. Sei $f : D \subseteq \mathbb{R}^n \to X$ und $x \in D$. Gibt es ein $df(x) \in L(\mathbb{R}^n, X)$, ein $\rho > 0$ mit $U_\rho(x) \subseteq D$ und eine Funktion $\epsilon : U_\rho(0) \setminus \{0\} \to X$ mit $\lim_{z\to 0} \epsilon(z) = 0$, sodass (10.5) für alle $z \in U_\rho(0) \setminus \{0\}$ gilt, dann heißt f bei x *differenzierbar*.
f heißt *stetig differenzierbar*, wenn f bei allen $x \in D$ differenzierbar und die $L_b(\mathbb{R}^n, X)$-wertige Abbildung $x \mapsto df(x)$ stetig auf D ist.

10.1.14 Fakta.

1. Wir haben in Proposition 10.1.11 gesehen, dass jede stetig partiell differenzierbare Funktion stetig differenzierbar ist.

2. Aus (10.5) erkennt man unmittelbar, dass die Differenzierbarkeit von f in x im Sinne von Definition 10.1.13 die Stetigkeit von f bei $x \in D$ nach sich zieht.

3. Ist f bei x differenzierbar, so gilt für $0 \neq v \in \mathbb{R}^n,\ s \in \mathbb{R},\ sv \in U_\rho(0)$,

$$\frac{f(x + sv) - f(x)}{s} = df(x)v + \operatorname{sgn}(s)\,\|v\|_\infty\,\epsilon(sv) \overset{s\to 0}{\longrightarrow} df(x)v\,.$$

   Also existieren alle Richtungsableitungen, insbesondere alle partiellen Ableitungen, wobei $\frac{\partial f}{\partial v}(x) = df(x)v$.

   Man sieht damit auch, dass $df(x)$ eindeutig durch f bestimmt ist, es also kein weiteres $\tilde{d}f(x) \in L(\mathbb{R}^n, X)$ mit $\tilde{d}f(x) \neq df(x)$ geben kann, das auch (10.5) erfüllt.

4. Ist f stetig differenzierbar, so sind auch die partiellen Ableitungen stetig, da ja

$$\left\|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(y)\right\| = \|\, df(x)e_i - df(y)e_i \,\| \leq \|\, df(x) - df(y) \,\| \cdot \|e_i\|_\infty\,.$$

   Also folgt aus der stetigen Differenzierbarkeit die stetige partielle Differenzierbarkeit.

Wir erhalten folgenden Satz.

10.1.15 Satz. *Eine Funktion $f : D \subseteq \mathbb{R}^n \to X$ ist genau dann stetig differenzierbar, wenn sie stetig partiell differenzierbar ist.*

10.1.16 Bemerkung. Wenn X mit einer zweiten, zu $\|.\|$ äquivalenten Norm versehen ist, bleiben die partiellen Ableitungen und die Ableitung unverändert, da für äquivalente Normen Netze genau denn bezüglich der einen Norm konvergieren, wenn sie bezüglich der anderen konvergieren, und in dem Fall auch die Grenzwerte übereinstimmen.

10.1.17 Proposition. *Sei $D \subseteq \mathbb{R}^n$ offen. Sind $f, g : D \to X$ zwei stetig differenzierbare Funktionen und $\lambda, \mu \in \mathbb{R}$, so ist es auch $\lambda f + \mu g$ und es gilt für alle $x \in D$*

$$d(\lambda f + \mu g)(x) = \lambda d(f)(x) + \mu dg(x) .$$

Beweis. Der Beweis folgt wegen Definition 10.1.13 unmittelbar durch eine entsprechende Linearkombination der Gleichungen

$$f(x + z) = f(x) + df(x)z + \|z\|_\infty \epsilon_f(z)$$

und

$$g(x + z) = g(x) + dg(x)z + \|z\|_\infty \epsilon_g(z) .$$

❑

Wie im Eindimensionalen lassen sich auch im Mehrdimensionalen viele Funktionen als Zusammensetzung von einfacheren Funktionen schreiben. Zur Berechnung der Ableitung dient dann folgender Satz, der die bekannte Kettenregel verallgemeinert.

10.1.18 Proposition (*Kettenregel*)**.** *Seien $f : D_f \subseteq \mathbb{R}^m \to X$, und $g : D_g \subseteq \mathbb{R}^n \to \mathbb{R}^m$ beide stetig differenzierbar, wobei $g(D_g) \subseteq D_f$.*
Dann ist auch $f \circ g : D_g \subseteq \mathbb{R}^n \to X$ stetig differenzierbar, und es gilt für $x \in D_g$[6]

$$d(f \circ g)(x) = df(g(x))\, dg(x) ,$$

wobei die rechte Seite die Hintereinanderausführung von zuerst $dg(x)$ und dann $df(g(x))$ ist.

Beweis. Laut Voraussetzung haben wir für $a, a + b \in D_f,\ b \neq 0$ gemäß (10.5)

$$f(a + b) = f(a) + df(a)b + \|b\|_\infty \epsilon_f(b) ,$$

und falls $x, x + z \in D_g,\ z \neq 0$

$$g(x + z) = g(x) + dg(x)z + \|z\|_\infty \epsilon_g(z) .$$

Wegen $\lim_{b\to 0} \epsilon_f(b) = 0$ und $\lim_{z\to 0} \epsilon_g(z) = 0$ können wir durch $\epsilon_g(0) = 0$ und $\epsilon_f(0) = 0$ diese beiden Funktion so fortsetzen, dass sie bei $z = 0$ bzw. $b = 0$ stetig sind.

[6] Man beachte, dass wegen $dg(x) \in L(\mathbb{R}^n, \mathbb{R}^m)$ und $df(g(x)) \in L(\mathbb{R}^m, X)$ diese Hintereinanderausführung linearer Abbildungen Sinn macht.

Ist nun $a = g(x)$ und b, sodass $a + b = g(x + z)$, so folgt aus der ersten Gleichung durch Einsetzen der zweiten

$$f(g(x + z)) =$$

$$f(g(x)) + df(g(x))\,(g(x + z) - g(x)) + \|\, g(x + z) - g(x)\,\|_\infty\, \epsilon_f(g(x + z) - g(x)) =$$

$$f(g(x)) + df(g(x))\,(dg(x)z + \|z\|_\infty\, \epsilon_g(z)) + \\ \Big\|\, dg(x)z + \|z\|_\infty \epsilon_g(z)\,\Big\|_\infty\, \epsilon_f(g(x + z) - g(x)) =$$

$$f(g(x)) + df(g(x))\, dg(x)\, z + \|z\|_\infty\, \gamma(z)\,,$$

wobei

$$\gamma(z) = df(g(x))(\epsilon_g(z)) + \|\, dg(x)(\frac{1}{\|z\|_\infty} z) + \epsilon_g(z)\,\|_\infty\, \epsilon_f(g(x + z) - g(x))$$

der Abschätzung

$$\|\gamma(z)\| \le \|df(g(x))\| \cdot \|\epsilon_g(z)\|_\infty + (\|dg(x)\| + \|\epsilon_g(z)\|_\infty)\, \|\epsilon_f(g(x + z) - g(x))\|$$

genügt. Wegen $\epsilon_f(0) = 0$ und der Stetigkeit von ϵ_f bei 0 folgt aus $\lim_{z\to 0}(g(x+z)-g(x)) = 0$, dass $\lim_{z\to 0} \epsilon_f(g(x+z)-g(x)) = 0$ (vgl. Proposition 6.1.4), und somit auch $\gamma(z) \to 0,\ z \to 0$. Wir haben also $d(f \circ g)(x) = df(g(x))\, dg(x)$ gezeigt. Um mit Hilfe von Satz 10.1.15 sicher zu gehen, dass $f \circ g \in C^1$, brauchen wir noch die Stetigkeit von $x \mapsto df(g(x))\, dg(x)$. Diese gilt aber wegen Korollar 9.2.9. ❑

10.1.19 Bemerkung. Ist $v \in \mathbb{R}^n$, so folgt unmittelbar aus Proposition 10.1.18 zusammen mit Lemma 10.1.6 und Definition 10.1.8, dass

$$\frac{\partial}{\partial v} f \circ g(x) = df(g(x))\, \frac{\partial}{\partial v} g(x)\,,$$

wobei rechts die Anwendung der linearen Abbildung $df(g(x)) \in L(\mathbb{R}^m, X)$ auf den Vektor $\frac{\partial}{\partial v} g(x) \in \mathbb{R}^m$ steht. Man beachte, dass die rechte Seite nicht dasselbe ist wie $\frac{\partial}{\partial v} f(g(x))\, \frac{\partial}{\partial v} g(x)$. Letzterer Ausdruck macht überhaupt keinen Sinn.

10.1.20 Bemerkung. Hat man $n = 1$ in Proposition 10.1.18, also $D_g \subseteq \mathbb{R}$, dann gilt $dg(t) = g'(t)$, wenn man $g'(t)$ mit der linearen Abbildung $\xi \mapsto \xi \cdot g'(t)$ aus $L(\mathbb{R}, \mathbb{R}^m)$ identifiziert, siehe Fakta 10.1.9, 3. Entsprechendes gilt für $(f \circ g)'(t)$. Also liest sich die Kettenregel in dem Fall als

$$(f \circ g)'(t) = df(g(t))\, g'(t), \quad t \in D_g\,, \tag{10.6}$$

wobei rechts die Anwendung der linearen Abbildung $df(g(t))$ auf den Vektor $g'(t)$ steht.

10.1.21 Bemerkung (*). Wir nehmen im Fall $n = 1$ nun sogar an, dass $g \in C^1(I)$ für irgendein, nicht notwendigerweise offenes Intervall $I \subseteq \mathbb{R}$ mit $g(I) \subseteq D_f$. Dann gilt (10.6) sogar für in I enthaltene Randpunkte t von I.
Dazu setze man g auf ein etwas größeres Intervall $J := I \cup (t - \epsilon, t + \epsilon)$ durch $g(s) := (s - t)g'(t) + g(t)$ für $s \in J \setminus I$ fort. Man überprüft unmittelbar, dass $g \in C^1(J)$, und dass für eine hinreichend kleine Wahl von $\epsilon > 0$ auch $g(J) \subseteq D_f$. Die Kettenregel angewandt auf $g|_{(t-\epsilon,t+\epsilon)}$ ergibt dann (10.6).

10.1.22 Beispiel. Als Beispiel wollen wir sogenannte *Parameterintegrale* betrachten. Sei $h : (a, b) \times (c, d) \to \mathbb{R}$ stetig, sodass die Ableitung nach der ersten Variablen für alle $(s, t) \in (a, b) \times (c, d)$ existiert, und dass

$$(s, t) \mapsto \frac{\partial}{\partial s} h(s, t)\,,$$

ebenfalls stetig ist. Weiters seien $\alpha, \beta : (a, b) \to (c, d)$ zwei stetig differenzierbare Funktionen. Wir wollen die Ableitung der Funktion

$$I(s) := \int_{\alpha(s)}^{\beta(s)} h(s, t)\, dt$$

berechnen. Dazu schreiben wir $I(s)$ als $f \circ g(s)$, wobei

$$f : (c, d) \times (c, d) \times (a, b) \subseteq \mathbb{R}^3 \to \mathbb{R},\ (\xi, \eta, \zeta)^T \mapsto \int_{\xi}^{\eta} h(\zeta, t)\, dt$$

und

$$g : (a, b) \subseteq \mathbb{R} \to \mathbb{R}^3,\ s \mapsto (\alpha(s), \beta(s), s)^T\,,$$

und wenden die Kettenregel an. Die partiellen Ableitungen von f sind nach dem Hauptsatz und nach Korollar 8.7.12 stetig, und die Matrixdarstellung von $df(x)$ ist ($x = (\xi, \eta, \zeta)^T$)

$$df(x) = \Big(\frac{\partial f}{\partial \xi}(x),\ \frac{\partial f}{\partial \eta}(x),\ \frac{\partial f}{\partial \zeta}(x)\Big) = (-h(\zeta, \xi),\ h(\zeta, \eta),\ \int_{\xi}^{\eta} \frac{\partial h}{\partial \zeta}(\zeta, t)\, dt)\,.$$

Auch $dg(s) = (\alpha'(s), \beta'(s), 1)^T$ ist stetig, und die Kettenregel ergibt (siehe auch Fakta 10.1.9, 3)

$$\begin{aligned}
\Big(I'(s)\Big) &= d(f \circ g)(s) = df(g(s))dg(s) \\
&= \Big(-h(s, \alpha(s)) \quad h(s, \beta(s)) \quad \textstyle\int_{\alpha(s)}^{\beta(s)} \frac{\partial h}{\partial s}(s, t)\, dt\Big) \begin{pmatrix} \alpha'(s) \\ \beta'(s) \\ 1 \end{pmatrix} \\
&= \Big(h(s, \beta(s))\beta'(s) - h(s, \alpha(s))\alpha'(s) + \int_{\alpha(s)}^{\beta(s)} \frac{\partial h}{\partial s}(s, t)\, dt\Big)\,.
\end{aligned}$$

10.1.23 Lemma. *Sei $D \subseteq \mathbb{R}^n$ offen und $f : D \to X$ stetig differenzierbar. Sind mit x, y auch alle Punkte auf ihrer Verbindungsgeraden in D, also $tx + (1-t)y \in D$ für $t \in [0, 1]$, so folgt*

$$\|f(x) - f(y)\| \leq \max_{t\in[0,1]} \|\, df(tx + (1-t)y)\,\| \cdot \|x - y\|_\infty .$$

Beweis. Wegen der Stetigkeit von $t \mapsto tx + (1-t)y$ ist das Bild der kompakten Menge $[0, 1]$ ebenfalls kompakt. Somit hat die stetige Funktion $z \mapsto \|df(z)\|$ darauf ein Maximum. Die Funktion $g(t) := f(x + t(y - x))$ ist nach Lemma 10.1.6 stetig differenzierbar mit $g'(t) = df(x + t(y - x))(y - x)$. Nach dem Hauptsatz (9.15) gilt

$$\begin{aligned}\|f(y) - f(x)\| = \|g(1) - g(0)\| &= \left\| \int_0^1 df(x + t(y - x))(y - x)\, dt \right\| \\ &\leq \sup_{t\in[0,1]} \|\, df(x + t(y - x))(y - x)\,\| \\ &\leq \max_{t\in[0,1]} \|\, df(x + t(y - x))\,\| \cdot \|y - x\|_\infty .\end{aligned}$$

❑

10.1.24 Korollar. *Gilt mit der Notation aus Lemma 10.1.23 für alle x aus einem konvexen $M \subseteq D$, dass $\|df(x)\| \leq C$, so folgt*

$$\|f(x) - f(y)\| \leq C\|x - y\|_\infty \quad \textit{für alle} \quad x, y \in M .$$

Insbesondere ist dann f auf M gleichmäßig stetig.

10.1.25 Bemerkung. Man kann auch Funktionen $f : D \to X$ betrachten, wo D offene Teilmenge eines allgemeinen Banachraumes Y ist, und sich fragen, ob und in welcher Hinsicht f differenzierbar ist.
Es stellt sich heraus, dass der Zugang aus Definition 10.1.13 der zweckmäßigste ist. Man spricht dann von der sogenannten *Fréchet-Differenzierbarkeit*. Mit fast denselben Beweisen wie hier angegeben bleiben alle Behauptungen ab Definition 10.1.13 bis zum Ende dieses Abschnitt richtig.
Eine Ausnahme stellt Satz 10.1.15 dar. Es gilt nur, dass aus der Differenzierbarkeit folgt, dass alle Richtungsableitungen existieren. Die Umkehrung gilt nicht, da es in dieser allgemeinen Situation kein Analogon zu Lemma 10.1.6 gibt.

10.2 Höhere Ableitungen

Sei f definiert auf einer offenen Menge $D \subseteq \mathbb{R}^n$ mit Werten in einem Banachraum X, und existiere für alle $x \in D$ die partielle Ableitung $\frac{\partial f}{\partial x_i}(x)$. Dann ist $\frac{\partial f}{\partial x_i}$ selbst eine Funktion

$$\frac{\partial f}{\partial x_i} : D \to X .$$

Also macht es Sinn, von der Differenzierbarkeit dieser Funktion zu sprechen. Existiert an einer Stelle $x \in D$ die Ableitung

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x) ,$$

so spricht man von einer zweiten Ableitung von f oder von der Ableitung zweiter Ordnung und schreibt kürzer $\frac{\partial^2 f}{\partial x_j \partial x_i}$. Höhere partielle Ableitungen definiert man induktiv:
Ist schon definiert, was partielle Ableitungen k-ter Ordnung sind, und ist $(i_1, i_2, \dots, i_{k+1})$ ein $(k+1)$-Tupel von Indizes $i_l \in \{1, \dots, n\}$, sodass $\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}$ auf ganz D existiert, dann setzten wir

$$\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}(x) := \frac{\partial}{\partial x_{i_{k+1}}} \left(\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}} \right)(x) ,$$

falls diese Ableitung existiert, und sprechen von Ableitungen $(k+1)$-ter Ordnung.

10.2.1 Definition. Sei $f : D \subseteq \mathbb{R}^n \to X$ definiert auf der offenen Menge D und sei $k \in \mathbb{N}$. Wir sagen, f ist *k-mal stetig differenzierbar* auf D, und schreiben $f \in C^k$ oder $f \in C^k(D)$, falls alle partiellen Ableitungen k-ter Ordnung von f auf ganz D existieren und stetig sind. Falls $f \in C^k(D)$ für alle $k \in \mathbb{N}$, so schreibt man $f \in C^\infty$ oder $f \in C^\infty(D)$ für diesen Sachverhalt.

10.2.2 Bemerkung. Bezeichnet man mit $C(D)$ die Menge aller stetigen Funktionen auf D, dann gilt

$$C(D) \supseteq C^1(D) \supseteq \dots \supseteq C^k(D) \supseteq C^{k+1}(D) \supseteq \dots .$$

Die erste Inklusion gilt wegen Fakta 10.1.14, 2, und die anderen zeigt man rekursiv:
Ist $f \in C^{k+1}(D)$, dann existieren alle partiellen Ableitungen $(k+1)$-ter Ordnung. Nach der induktiven Definition der $(k+1)$-ten partiellen Ableitungen müssen auch alle partiellen Ableitungen k-ter Ordnung existieren. Sei $g = \frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}$ eine solche. Dann existieren alle partiellen Ableitungen von g und sind stetig, also $g \in C^1$. Wegen $C^1(D) \subseteq C(D)$ folgt, dass g selbst stetig ist; vgl. Fakta 10.1.14, 2. Also gilt $f \in C^k(D)$.

10.2.3 Satz (*Satz von Schwarz*). *Sei $f \in C^2(D)$, $D \subseteq \mathbb{R}^n$, und seien $i, j \in \{1, \dots, n\}$, sodass $\frac{\partial^2 f}{\partial x_i \partial x_j}$ und $\frac{\partial^2 f}{\partial x_j \partial x_i}$ auf ganz D existieren und stetig sind. Dann gilt*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} .$$

Allgemeiner gilt für $f \in C^k(D)$, dass es bei der Bildung einer partiellen Ableitung höchstens k-ter Ordnung nicht darauf ankommt, in welcher Reihenfolge differenziert wird.

Beweis. Sei $x \in D$ und $\delta > 0$, so dass $U_\delta(x) \subseteq D$, wobei wir $\mathbb{R}^n$ mit $\|.\|_\infty$ versehen. Nach dem Hauptsatz der Differential- und Integralrechnung (siehe (9.15)) gilt für $|\xi|, |\eta| < \delta$

$$f(x + \xi e_i + \eta e_j) - f(x + \eta e_j) = \int_0^\xi \frac{\partial f}{\partial x_i}(x + \zeta e_i + \eta e_j)\, d\zeta .$$

Die Ableitung des Integranden nach η ist gerade $\frac{\partial^2 f}{\partial x_j \partial x_i}(x + \zeta e_i + \eta e_j)$, und daher stetig in ζ und η. Somit können wir (9.18) – siehe auch Korollar 8.7.12 – anwenden und erhalten bei der Ableitung obiger Gleichung nach η

$$\frac{\partial f}{\partial x_j}(x + \xi e_i + \eta e_j) - \frac{\partial f}{\partial x_j}(x + \eta e_j) = \int_0^\xi \frac{\partial^2 f}{\partial x_j \partial x_i}(x + \zeta e_i + \eta e_j)\, d\zeta .$$

Leiten wir nun nach ξ ab, so folgt mit dem Hauptsatz

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x + \xi e_i + \eta e_j) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x + \xi e_i + \eta e_j)\,.$$

Die Behauptung für Funktionen $f \in C^k(D)$ folgt aus dem Bewiesenen leicht durch vollständiger Induktion, wenn man sich vor Augen hält, dass sich jede Permutation als Hintereinanderausführung von Transpositionen schreiben lässt. ❑

10.2.4 Bemerkung. Ist $(i_1, i_2, \ldots, i_k) \in \{1, \ldots, n\}^k$ ein k-Tupel, und bezeichnet α_j ($\in \mathbb{N} \cup \{0\}$) für $j = 1, \ldots, n$ die Anzahl der $m \in \{1, \ldots, k\}$ mit $i_m = j$, so gilt für ein $f \in C^k(D)$ wegen Satz 10.2.3

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(x) = \frac{\partial^k f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x)\,, \tag{10.7}$$

wobei $\partial x_j^{\alpha_j}$ für α_j mal nach x_j abgeleitet steht. Offenbar gilt $k = \alpha_1 + \cdots + \alpha_n$.
Man überzeugt sich durch Induktion nach n davon, dass es dabei genau

$$\frac{k!}{\alpha_1! \cdot \cdots \cdot \alpha_n!}$$

viele k-Tupel $(i_1, i_2, \ldots, i_k)$ gibt, die im obigen Sinne dasselbe n-Tupel $(\alpha_1, \ldots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n$ erzeugen. Für den Ausdruck in (10.7) schreibt man auch

$$D^{(\alpha_1, \ldots, \alpha_n)} f\,(x)\,.$$

Die n-Tupeln $(\alpha_1, \ldots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n$ nennt man auch *Multiindizes*.

10.2.5 Beispiel. Wir betrachten abermals die Funktion $T : D \to \mathbb{R}^2$ definiert durch

$$T\begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} r\ \cos t \\ r\ \sin t \end{pmatrix}.$$

mit $D = \mathbb{R}^+ \times \mathbb{R} \subseteq \mathbb{R}^2$. Um die höheren Ableitungen dieser speziellen Funktion zu studieren, bietet es sich an, die Bildmenge $\mathbb{R}^2$ als $\mathbb{C}$ zu betrachten, da man dann

$$T\begin{pmatrix} r \\ t \end{pmatrix} = r \cdot \exp(it)$$

schreiben kann. Die höheren partiellen Ableitungen sind dann

$$\frac{\partial^l}{\partial t^l} T\begin{pmatrix} r \\ t \end{pmatrix} = r \cdot i^l \exp(it) = r \cdot \exp\left(i(t + l\frac{\pi}{2})\right) \cong \begin{pmatrix} r\ \cos\left(t + l\frac{\pi}{2}\right) \\ r\ \sin\left(t + l\frac{\pi}{2}\right) \end{pmatrix},$$

$$\frac{\partial^{1+l}}{\partial r\, \partial t^l} T\begin{pmatrix} r \\ t \end{pmatrix} = i^l \exp(it) \cong \begin{pmatrix} \cos\left(t + l\frac{\pi}{2}\right) \\ \sin\left(t + l\frac{\pi}{2}\right) \end{pmatrix},$$

und

$$\frac{\partial^{k+l}}{\partial r^k\, \partial t^l} T\begin{pmatrix} r \\ t \end{pmatrix} = 0, \quad k \geq 2\,.$$

10.2.6 Bemerkung. Für $f \in C^1(D)$ haben wir die Ableitung $df : D \to L(\mathbb{R}^n, X)$ so definiert, dass $df(x)v$ genau die Richtungsableitung von f in Richtung v ist; siehe Lemma 10.1.6. Ist nun $f \in C^k(D)$ und sind $v_1 = (\mu_{1,1} \dots, \mu_{1,n})^T, \dots, v_k = (\mu_{k,1} \dots, \mu_{k,n})^T$ Richtungsvektoren, so erhält man mit Lemma 10.1.6

$$\frac{\partial f}{\partial v_1}(x) = \sum_{l=1}^{n} \mu_{1,l} \frac{\partial f}{\partial x_l}(x) .$$

Wendet man Lemma 10.1.6 nochmals an, so folgt

$$\frac{\partial}{\partial v_2} \frac{\partial}{\partial v_1} f(x) = \sum_{l=1}^{n} \mu_{1,l} \frac{\partial}{\partial v_2} \frac{\partial f}{\partial x_l}(x) = \sum_{l_1=1}^{n} \mu_{1,l_1} \sum_{l_2=1}^{n} \mu_{2,l_2} \frac{\partial^2 f}{\partial x_{l_1} \partial x_{l_2}}(x) .$$

Macht man das nun insgesamt k-mal, so sieht man, dass die k-malige Hintereinanderausführung der Richtungsableitungen nach v_1 bis v_k existiert und mit

$$\frac{\partial}{\partial v_k} \dots \frac{\partial}{\partial v_1} f(x) = \sum_{l_1, \dots, l_k=1}^{n} \mu_{1,l_1} \cdot \dots \cdot \mu_{k,l_k} \frac{\partial^k f}{\partial x_{l_1} \dots \partial x_{l_k}}(x) \tag{10.8}$$

übereinstimmt.

10.2.7 Definition. Für eine mindestens k-mal stetig differenzierbare Funktion $f : D \to X$ mit einem offenen $D \subseteq \mathbb{R}^n$ und einem Banachraum X und für $x \in D$ sowie $v_1, \dots, v_k \in \mathbb{R}^n$ bezeichnen wir mit $d^k f(x)(v_1, \dots, v_k)$ den Ausdruck in (10.8).

10.2.8 Fakta.

1. Im Falle $k = 1$ stimmt (10.8) mit $df(x)v_1$ überein.

2. Für allgemeines $k \in \mathbb{N}$ ist $d^k f(x)(v_1, \dots, v_k)$ linear in jedem Argument:

$$d^k f(x)(v_1, \dots, \alpha v_l + \beta v_l', \dots, v_k) = \\ \alpha d^k f(x)(v_1, \dots, v_l, \dots, v_k) + \beta d^k f(x)(v_1, \dots, v_l', \dots, v_k) .$$

   Außerdem ist $d^k f(x)(v_1, \dots, v_k)$ symmetrisch in $v_1, \dots, v_k$, hängt daher nicht von der Reihenfolge der Richtungsvektoren ab. Das folgt unmittelbar aus Satz 10.2.3 und (10.8).

   Somit ist $d^k f(x) : \mathbb{R}^{k \times n} \to X$ eine symmetrische, multilineare Abbildung, sodass $d^k f(x)(v_1, \dots, v_k)$ gerade $\frac{\partial}{\partial v_k} \dots \frac{\partial}{\partial v_1} f(x)$ ist. Wir nennen $d^k f(x)$ die *k-te Ableitung* von f an der Stelle x.

3. Im Fall $n = 1$, also $D \subseteq \mathbb{R}$, gilt $d^k f(x)(v_1, \dots, v_k) = v_1 \cdot \dots \cdot v_k\, f^{(k)}(x)$.

4. Für einen späteren Gebrauch stellen wir noch heraus, dass wenn $y, v \in \mathbb{R}^n$ mit $y + tv \in D$ für alle $t \in [0,1]$, die k-te Ableitung[7] der Funktion $\phi : [0,1] \to X$,
$$\phi(t) = f(y + tv)\,,$$
nichts anderes als $\frac{\partial}{\partial v} \dots \frac{\partial}{\partial v} f(y + tv) = d^k f(y + tv)(v, \dots, v)$ ist. Ordnen wir für $v = v_1 = \dots = v_k$ die Ableitungen ∂x_l in (10.8) nach aufsteigendem l wie in (10.7), so erhalten wir mit $v = (\mu_1, \dots, \mu_n)^T$
$$\begin{aligned} \phi^{(k)}(t) &= d^k f(y + tv)(v, \dots, v) \\ &= \sum_{\substack{(\alpha_1,\dots,\alpha_n)\in(\mathbb{N}\cup\{0\})^n \\ \alpha_1+\dots+\alpha_n=k}} \mu_1^{\alpha_1} \cdot \dots \cdot \mu_n^{\alpha_n} \, \frac{k!}{\alpha_1! \cdot \dots \cdot \alpha_n!} \, \frac{\partial^k f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(y + tv)\,. \end{aligned} \tag{10.9}$$

10.2.9 Beispiel. Wir betrachten die Funktion T aus Beispiel 10.2.5. Die lineare Abbildung $dT((r,t)^T)$ haben wir in (10.4) berechnet.
Für $m \in \mathbb{N},\ m \geq 2$ folgt aus Beispiel 10.2.5 mit $x_1 = r, x_2 = t,\ \mu_{j,1} = \rho_j, \mu_{j,2} = \tau_j$
$$d^m T\begin{pmatrix} r \\ t \end{pmatrix}\left(\begin{pmatrix} \rho_1 \\ \tau_1 \end{pmatrix}, \dots, \begin{pmatrix} \rho_m \\ \tau_m \end{pmatrix}\right) = \sum_{l_1,\dots,l_m=1}^{2} \mu_{1,l_1} \cdot \dots \cdot \mu_{m,l_m} \frac{\partial^m T}{\partial x_{l_1} \dots \partial x_{l_k}} \begin{pmatrix} r \\ t \end{pmatrix} =$$
$$(\rho_1\tau_2 \dots \tau_m + \dots + \tau_1 \dots \tau_{m-1}\rho_m)\, \frac{\partial^m}{\partial r\, \partial t^{m-1}} T\begin{pmatrix} r \\ t \end{pmatrix} + \tau_1 \cdot \dots \cdot \tau_m \frac{\partial^m}{\partial t^m} T\begin{pmatrix} r \\ t \end{pmatrix} =$$
$$(\rho_1\tau_2 \dots \tau_m + \dots + \tau_1 \dots \tau_{m-1}\rho_m) \begin{pmatrix} \cos\left(t + (m-1)\frac{\pi}{2}\right) \\ \sin\left(t + (m-1)\frac{\pi}{2}\right) \end{pmatrix} + \tau_1 \cdot \dots \cdot \tau_m \begin{pmatrix} r\ \cos\left(t + m\frac{\pi}{2}\right) \\ r\ \sin\left(t + m\frac{\pi}{2}\right) \end{pmatrix}.$$
Man erkennt unmittelbar aus (10.4), dass diese Beziehung auch für $m = 1$ richtig ist.

Wir können nun die mehrdimensionale Taylorsche Formel herleiten.

10.2.10 Satz (*Taylorsche Formel*)**.** *Sei $f : D\ (\subseteq \mathbb{R}^n) \to X$ und $q \in \mathbb{N} \cup \{0\}$, sodass f mindestens $q+1$-mal stetig differenzierbar ist. Weiters sei $y \in D$ fest und $x \in D$, sodass die gesamte Strecke von y bis x in D liegt*[8]*. Dann gilt*
$$f(x) = f(y) + \sum_{l=1}^{q} \frac{1}{l!} d^l f(y)(\underbrace{x - y, \dots, x - y}_{l-mal}) + R_q(x)\,,$$
wobei sich das Restglied R_q schreiben lässt als
$$R_q(x) = \frac{1}{q!} \int_0^1 (1-t)^q \cdot d^{q+1} f((1-t)y + tx)(\underbrace{x - y, \dots, x - y}_{(q+1)-mal})\, dt\,. \tag{10.10}$$

[7] An den Randpunkten ist die Ableitung einseitig zu verstehen.
[8] Das stimmt immer, wenn D konvex ist.

Beweis. Wir entwickeln die Funktion $\phi(t) = f(y + th)$, $t \in [0, 1]$ nach Taylor wie in (9.19). Also gilt

$$\phi(1) = \phi(0) + \sum_{l=1}^{q} \frac{\phi^{(l)}(0)}{l!} + \int_0^1 \frac{(1-t)^q}{q!} \phi^{(q+1)}(t)\, dt .$$

Wegen (10.9) gilt $\phi^{(l)}(t) = d^l f(y + th)(h, \dots, h)$, woraus unmittelbar die behauptete Entwicklung folgt. ❑

10.2.11 Bemerkung. Den Ausdruck

$$T_q(x) = f(y) + \sum_{l=1}^{q} \frac{1}{l!} d^l f(y)(\underbrace{x-y, \dots, x-y}_{l-\text{mal}})$$

nennt man das q-te *mehrdimensionale Taylorsche Polynom* der Funktion f an der Anschlussstelle y. Mit (10.9) folgt die Darstellung

$$T_q(x) = \sum_{\substack{(\alpha_1,\dots,\alpha_n)\in(\mathbb{N}\cup\{0\})^n \\ \alpha_1+\dots+\alpha_n\le q}} \frac{1}{\alpha_1! \cdot \dots \cdot \alpha_n!} (x_1 - y_1)^{\alpha_1} \cdot \dots \cdot (x_n - y_n)^{\alpha_n} \frac{\partial^{\alpha_1+\dots+\alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(y) ,$$

wenn wir für $\alpha_1 + \dots + \alpha_n = 0$ den entsprechenden Summanden als $f(y)$ interpretieren. Wir sehen insbesondere, dass $T_q(x)$ tatsächlich ein Polynom in den n Variablen $x_1, \dots, x_n$ mit Werten in X vom Grad $\le q$ ist; vgl. Beispiel 10.2.12.

10.2.12 Beispiel. Wir setzen für $\alpha \in (\mathbb{N} \cup \{0\})^n$ und $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$

$$|\alpha| = \alpha_1 + \dots + \alpha_n, \quad \alpha! = \alpha_1! \cdot \dots \cdot \alpha_n!, \quad x^\alpha = x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n} .$$

Sind c_α für $|\alpha| \le m$ Koeffizienten aus einem Banachraum X, so heißt die Funktion $f : \mathbb{R}^n \to X$

$$f(x) = \sum_{\substack{\alpha\in(\mathbb{N}\cup\{0\})^n \\ |\alpha|\le m}} x^\alpha c_\alpha$$

Polynom in den Variablen $x_1, \dots, x_n$ vom Grad $\le m$ mit Werten in X.
Durch Ausmultiplizieren erkennt man leicht, dass für jedes feste $y \in \mathbb{R}^n$ mit f auch $x \mapsto f(x - y)$ ein Polynom vom Grad $\le m$ ist.
Ebenfalls unschwer überprüft man, dass für $\alpha,\beta \in (\mathbb{N} \cup \{0\})^n$ mit $\beta \le \alpha$, also $\beta_1 \le \alpha_1, \dots, \beta_n \le \alpha_n$, [9]

$$\frac{\partial^{|\beta|}(x \mapsto x^\alpha)}{\partial x^\beta}(y) = \frac{\alpha!}{(\alpha-\beta)!} y^{\alpha-\beta} , \tag{10.11}$$

gilt, wobei $\alpha - \beta \in (\mathbb{N} \cup \{0\})^n$ komponentenweise berechnet wird. Falls $\beta_j > \alpha_j$ für mindestens ein $j \in \{1, \dots, n\}$, so gilt $\frac{\partial^{|\beta|}(x\mapsto x^\alpha)}{\partial x^\beta} \equiv 0$.

[9] Hier und im Rest des Beispiels ist $\frac{\partial^{|\beta|} f}{\partial x^\beta}$ eine Kurzschreibweise für $\frac{\partial^{\beta_1+\dots+\beta_n} f}{\partial x_1^{\beta_1} \cdots \partial x_n^{\beta_n}}$.

Wenden wir für ein festes $y \in \mathbb{R}^n$ Satz 10.2.10 auf ein Polynom f vom Grad $\leq m$ und $q \geq m$ an, so folgt

$$f(x) = \sum_{\substack{\alpha \in (\mathbb{N} \cup \{0\})^n \\ |\alpha| \leq m}} (x-y)^\alpha \, \frac{1}{\alpha!} \, \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(y)\,,$$

da $\frac{\partial^{|\alpha|} f}{\partial x^\alpha} = 0$ für $|\alpha| > m$ und damit insbesondere wegen

$$d^{q+1} f((1-t)y + tx)(\underbrace{x-y, \dots, x-y}_{(q+1)-\mathrm{mal}}) = \sum_{\substack{\alpha \in (\mathbb{N} \cup \{0\})^n \\ |\alpha| = q+1}} (x-y)^\alpha \, \frac{(q+1)!}{\alpha!} \, \frac{\partial^{q+1} f}{\partial x^\alpha}((1-t)y + tx) = 0$$

auch das Restglied verschwindet.
Für $y = 0$ folgt aus (10.11), dass $c_\alpha = \frac{1}{\alpha!} \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(0)$, womit die Koeffizienten c_α eindeutig durch die Funktionswerte von f auf jeder beliebig kleinen offenen Kugel um die Null bestimmt sind.

10.2.13 Proposition. *Mit der Notation aus Satz 10.2.10 gilt:*

(*i*) *Ist $C \geq 0$, sodass sich die partiellen Ableitungen $q+1$-ter Ordnung abschätzen lassen durch*

$$\left\| \frac{\partial^{q+1} f}{\partial x_{l_1} \cdots \partial x_{l_{q+1}}}(u) \right\| \leq C\,, \tag{10.12}$$

für alle $l_1, \dots, l_{q+1} \in \{1, \dots, n\}$ und u auf der Strecke von y bis x, dann gilt

$$\|R_q(x)\| \leq C \frac{n^{q+1}}{(q+1)!} \|x-y\|_\infty^{q+1}\,.$$

(*ii*) *Ist D konvex und f sogar in $C^\infty(D)$ und gilt* (10.12) *für alle q und alle $u \in D$, so konvergiert die Reihe*

$$f(y) + \sum_{l=1}^{\infty} \frac{1}{l!} d^l f(y)(\underbrace{x-y, \dots, x-y}_{l-mal})$$

gleichmäßig auf beschränkten Teilmengen von D gegen $f(x)$.

Beweis.

(*i*) Aus (10.12) folgt wegen (10.8) für $u = (1-t)y + tx$ und $h = (h_1, \dots, h_n)^T \in \mathbb{R}^n$

$$\|\, d^{q+1} f(u)(h, \dots, h)\,\| \leq \sum_{l_1, \dots, l_{q+1}=1}^{n} |h_{l_1}| \dots |h_{l_{q+1}}| \cdot \left\| \frac{\partial^{q+1} f}{\partial x_{l_1} \dots \partial x_{l_{q+1}}}(u) \right\| \leq C n^{q+1} \|h\|_\infty^{q+1}\,.$$

Für $h = x - y$ erhalten wir daraus zusammen mit (10.10)

$$\|R_q(x)\| \leq \frac{1}{q!} \int_0^1 (1-t)^q \cdot \|\, d f^{q+1}((1-t)y + tx)(x-y, \dots, x-y)\,\| \, dt$$

$$\leq C \frac{n^{q+1}}{q!} \|h\|_\infty^{q+1} \int_0^1 (1-t)^q \, dt = C \frac{n^{q+1}}{(q+1)!} \|h\|_\infty^{q+1}\,.$$

(ii) Bei fest gewähltem y konvergiert $\frac{n^{q+1}}{(q+1)!}\|x-y\|^{q+1}$ und damit $R_q(x)$ für jedes x gegen 0. In der Tat geht diese Konvergenz für x in beschränkten Teilmengen von D gleichmäßig von statten. ❑

10.2.14 Beispiel. Sei $T : D \to \mathbb{R}^2$ wie in Beispiel 10.2.5, wobei aber jetzt $D : (0,\alpha)\times\mathbb{R}$ für irgend ein festes $\alpha \in \mathbb{R}^+$, $\alpha > 1$. Wendet man Satz 10.2.10 für den Punkt $y = (r,t)^T = (1,0)$ an, so erhalten wir mit $x - y = (\rho,\tau)^T$ aus Beispiel 10.2.9

$$\begin{aligned} T(x) &= T(y) + \sum_{l=1}^{q} \frac{1}{l!} d^l T(y)(\underbrace{x-y,\dots,x-y}_{l-\text{mal}}) + R_q(x) \\ &= \begin{pmatrix}1\\0\end{pmatrix} + \sum_{l=1}^{q} \frac{1}{l!}\left(l\rho\tau^{l-1}\begin{pmatrix}\cos(l-1)\frac{\pi}{2}\\ \sin(l-1)\frac{\pi}{2}\end{pmatrix} + \tau^l \begin{pmatrix}\cos l\frac{\pi}{2}\\ \sin l\frac{\pi}{2}\end{pmatrix}\right) + R_q(x)\,. \end{aligned} \tag{10.13}$$

In Beispiel 10.2.5 haben wir insbesondere gesehen, dass (10.12) für $C = \alpha$ und $\|.\| = \|.\|_\infty$ erfüllt ist. Also konvergiert der entsprechende Ausdruck in (10.13) gegen $T(x)$.

10.3 Extremwerte

Wir haben im ersten Semester in Definition 7.2.1 definiert, was es bedeutet, dass eine reellwertige Funktion $f : D \to \mathbb{R}$ – D ist Teilmenge eines metrischen Raumes – in einem Punkt $x \in D$ ein *lokales Maximum* hat:

$$\exists \delta > 0 : \ \forall t \in U_\delta(x) : \ f(t) \leq f(x)\,.$$

Entsprechendes gilt für *lokale Minima*. Ist D offene Teilmenge von $\mathbb{R}$, und ist f differenzierbar in einem lokalen Extremum, also einem lokalen Maximum oder lokalen Minimum x, dann gilt $f'(x) = 0$; siehe Lemma 7.2.2. Wir wollen diese Tatsache nun für auf $D \subseteq \mathbb{R}^n$ definierte Funktionen herleiten.

10.3.1 Satz. *Sei x ein lokales Extremum, also lokales Maximum oder lokales Minimum, von $f : D \to \mathbb{R}$ mit einem offenen $D \subseteq \mathbb{R}^n$. Falls für $v \in \mathbb{R}^n$ die Richtungsableitung*

$$\frac{\partial f}{\partial v}(x) = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

bei x existiert, so ist diese gleich Null. Für $f \in C^1$ folgt $df(x) = 0$.

Beweis. Die Funktion $\phi(t) = f(x+tv)$ ist für $t \in (-\delta,\delta)$ mit hinreichend kleinem $\delta > 0$ definiert. Wegen unserer Voraussetzung ist ϕ bei 0 differenzierbar und hat die Ableitung $\phi'(0) = \frac{\partial f}{\partial v}(x)$. Da x ein lokales Extremum von f ist, hat ϕ ein lokales Extremum bei 0. Somit gilt $\phi'(0) = 0$.
Für $f \in C^1$ ist $df(x)$ derart definiert, dass $df(x)v = \frac{\partial f}{\partial v}(x)$ für alle $v \in \mathbb{R}^n$. Es folgt also $df(x) = 0$. ❑

Für $f \in C^1(D)$ ist also $df(x) = 0$ eine notwendige Bedingung für ein Extremum. Punkte mit $df(x) = 0$ heißen *stationäre Punkte*. Wie im eindimensionalen ist $df(x) = 0$ aber nicht hinreichend. Um dieses Problem genauer studieren zu können, wollen wir folgende Sprechweise einführen.

10.3.2 Definition. Sei $D \subseteq \mathbb{R}^n$ offen und liege $f : D \to \mathbb{R}$ in $C^k(D)$. Für gerades k heißt die k-te Ableitung $d^k f(x) : \mathbb{R}^{k\times n} \to \mathbb{R}$ von f im Punkt x *positiv definit* bzw. *positiv semidefinit*, wenn $d^k f(x)(h, \dots, h) > 0$ bzw. $d^k f(x)(h, \dots, h) \geq 0$ für alle $0 \neq h \in \mathbb{R}^n$.
Entsprechend definiert man *negativ definit* bzw. *negativ semidefinit*.

10.3.3 Bemerkung. Ist $n = 1$, also $f : D \subseteq \mathbb{R} \to \mathbb{R}$, dann gilt $d^k f(x)(h, \dots, h) = h^k \cdot f^{(k)}(x)$, und man sieht, dass $d^k f(x)$ genau dann positiv definit (semidefinit) ist, wenn $f^{(k)}(x) > (\geq)\, 0$. Entsprechendes gilt für den negativ definiten (semidefiniten) Fall.

10.3.4 Bemerkung. Der für die meisten Anwendungen relevante Fall ist $k = 2$. Hier lassen sich die Definitheitseigenschaften von $d^2 f(x)$ mit Mitteln der Linearen Algebra betrachten. In diesem Fall ist $d^2 f(x) : \mathbb{R}^{2\times n} \to \mathbb{R}$ eine symmetrische *Bilinearform*:

$$d^2 f(x)(u, v) = \sum_{l_1, l_2 = 1}^{n} u_{l_1} v_{l_2} \frac{\partial^2 f}{\partial x_{l_1} \partial x_{l_2}}(x) =$$

$$\begin{pmatrix} u_1 & \dots & u_n \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Diese symmetrische $n \times n$-Matrix $H_f(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ heißt auch *Hesse-Matrix*.
Aus der Linearen Algebra ist bekannt, dass die Bilinearform positiv definit (semidefinit) ist, wenn alle Eigenwerte $\lambda_1, \dots, \lambda_n$ – diese sind nicht notwendigerweise paarweise verschieden – von $H_f(x)$ größer (größer gleich) Null sind. Entsprechendes gilt für den negativen Fall.
Weiters ist diese Bilinearform positiv definit genau dann, wenn alle Hauptminoren[10] von $H_f(x)$ größer Null sind. Entsprechend ist sie genau dann negativ definit, wenn die Hauptminoren nicht verschwinden und abwechselndes Vorzeichen beginnend mit einem negativen Vorzeichen haben.

Für den Beweis des folgenden Satzes benötigen wir ein Lemma.

10.3.5 Lemma. *Für $k, n \in \mathbb{N}$ sei $\omega : \mathbb{R}^{k\times n} \to \mathbb{R}$ eine symmetrische, multilineare Abbildung. Gilt $\omega(v, \dots, v) = 0$ für alle $v \in \mathbb{R}^n$, so verschwindet $\omega(v_1, \dots, v_k)$ für alle $v_1, \dots, v_k \in \mathbb{R}^n$.*

Beweis. Für $k = 1$ ist nichts zu beweisen. Für größere k's zeigen wir die Aussage durch vollständige Induktion nach k. Ist $k = 2$, so gilt für beliebige $v, w \in \mathbb{R}^n$

$$0 = \omega(v + w, v + w) = \omega(v, v) + \omega(v, w) + \omega(w, v) + \omega(w, w) = \omega(v, w) + \omega(w, v),$$

[10] Hauptminoren einer quadratischen Matrix $(\mu_{ij})_{i,j=1,\dots,n}$ sind die Determinanten aller Untermatrizen $(\mu_{ij})_{i,j=1,\dots,k}$ für $k = 1, \dots, n$.

woraus $\omega(v, w) = -\omega(w, v)$ folgt. Zusammen mit $\omega(v, w) = \omega(w, v)$ erhält man $\omega(v, w) = 0$. Ist $k > 2$, so seien wieder $v, w \in \mathbb{R}^n$ beliebig. Weiters sei $\lambda \in \mathbb{R}$. Wegen der Symmetrie gilt

$$\omega(v + \lambda w, \dots, v + \lambda w) =$$

$$\omega(v, \dots, v) + \sum_{j=1}^{k-1} \lambda^j \binom{k}{j} \omega(\underbrace{v, \dots, v}_{j-\text{mal}}, \underbrace{w, \dots, w}_{(k-j)-\text{mal}}) + \lambda^k \omega(w, \dots, w)\,.$$

Für feste $v, w \in \mathbb{R}^n$ und variablem λ steht links immer Null und rechts ein Polynom in λ, dessen Koeffizienten alle verschwinden müssen. Für $j = k - 1$ gilt daher insbesondere

$$\omega(\underbrace{v, \dots, v}_{(k-1)-\text{mal}}, w) = 0\,,$$

und zwar für alle $v, w \in \mathbb{R}^n$. Somit erfüllt für festes $w \in \mathbb{R}^n$ die offensichtlich symmetrische und multilineare Abbildung

$$(v_1, \dots, v_{k-1}) \mapsto \omega(v_1, \dots, v_{k-1}, w)$$

die Voraussetzungen unseres Lemmas für $k-1$ anstatt für k. Nach Induktionsvoraussetzung verschwindet diese multilineare Abbildung für alle $v_1, \dots, v_{k-1} \in \mathbb{R}^n$. Da w zwar fest, aber beliebig war, verschwindet auch ω. ❑

10.3.6 Satz. *Für ein offenes $D \subseteq \mathbb{R}^n$ sei $f : D \to \mathbb{R}$, $f \in C^r(D)$ mit $r \in \mathbb{N} \cup \{\infty\}$, und sei $x \in D$ mit $df(x) = 0, \dots, d^{q-1}f(x) = 0, d^q f(x) \neq 0$, wobei $q \leq r$. Dann gilt:*

(*i*) *Hat f ein lokales Maximum bei x, so ist q gerade und $d^q f(x)$ ist negativ semidefinit.*

(*ii*) *Ist q gerade und $d^q f(x)$ negativ definit, so hat f ein lokales Maximum bei x.*

(*iii*) *Hat f ein lokales Minimum bei x, so ist q gerade und $d^q f(x)$ ist positiv semidefinit.*

(*iv*) *Ist q gerade und $d^q f(x)$ positiv definit, so hat f ein lokales Minimum bei x.*

(*v*) *Ist q ungerade oder ist $d^q f(x)$ indefinit, also weder positiv semidefinit noch negativ semidefinit, dann ist x kein lokales Extremum.*

Beweis.

(*i*) Sei x ein relatives Maximum von f. Dann existiert ein $\delta > 0$, sodass

$$f(y) \leq f(x) \text{ für alle } y \in U_\delta(x)\,.$$

Für ein $v \in \mathbb{R}^n \setminus \{0\}$ definieren wir $\phi_v(t) = f(x + tv)$, $|t| < \frac{\delta}{\|v\|}$, und sehen wie in Satz 10.3.1, dass ϕ_v bei Null ein lokales Maximum hat. Außerdem ist $\phi_v^{(k)}(0) = df^k(x)(v, \dots, v)$, $k = 1, \dots, q$. Also $\phi_v'(0) = \cdots = \phi_v^{(q-1)}(0) = 0$.

Nach Voraussetzung ist $d^q f(x) \neq 0$, und wegen Lemma 10.3.5 muss $\phi_v^{(q)}(0) = d^q f(x)(v, \dots, v) \neq 0$ für zumindest eine Richtung v. Wäre q ungerade, so wäre nach Korollar 7.4.7 die Zahl 0 kein lokales Extremum von ϕ_v.

Damit ist q gerade. Wäre nun $d^q f(x)$ nicht negativ semidefinit, also $\phi_v^{(q)}(0) = d^q f(x)(v, \dots, v) > 0$ für zumindest eine Richtung v, so folgt wieder aus Korollar 7.4.7, dass 0 ein lokales Minimum von ϕ_v ist. Somit ist 0 lokales Minimum und lokales Maximum. Also muss ϕ_v auf einem hinreichend kleinen Intervall um 0 konstant sein, was aber $\phi_v^{(q)}(0) = 0$ implizieren würde.

(*ii*) Sei q gerade und $d^q f(x)$ negativ definit. Aus Stetigkeitsgründen gibt es ein $\delta > 0$, sodass $d^q f(w)$ ebenfalls negativ definit ist für alle $w \in U_\delta(x)$.

Um das einzusehen, nehme man das Gegenteil an. Dann gibt es zu jedem $m \in \mathbb{N}$ ein x_m, $\|x_m - x\|_\infty < \frac{1}{m}$, sodass $d^q f(x_m)$ nicht negativ definit ist. Es gibt daher ein $0 \neq h_m \in \mathbb{R}^n$ mit $d^q f(x_m)(h_m, \dots, h_m) \geq 0$. Wegen

$$d^q f(x_m)\Big(\frac{h_m}{\|h_m\|_\infty}, \dots, \frac{h_m}{\|h_m\|_\infty}\Big) = \frac{1}{\|h_m\|_\infty^q} d^q f(x_m)(h_m, \dots, h_m) \geq 0$$

können wir annehmen, dass die h_m immer $\|h_m\|_\infty = 1$ erfüllen – also in der abgeschlossenen und beschränkten Teilmenge

$$K_1(0) \setminus U_1(0) = \{h : \|h\|_\infty = 1\} \subseteq \mathbb{R}^n$$

liegen. Da diese Menge nach Korollar 5.2.9 kompakt ist, gilt $h_{m(j)} \to h$, $j \to \infty$, für eine Teilfolge und für ein $h \in K_1(0) \setminus U_1(0)$.

Somit konvergieren die Komponenten von $h_{m(j)}$ gegen die entsprechenden Komponenten von h und alle partiellen Ableitungen $\frac{\partial^k f}{\partial x_{l_1} \dots \partial x_{l_k}}(x_{m(j)})$ gegen $\frac{\partial^k f}{\partial x_{l_1} \dots \partial x_{l_k}}(x)$. Wegen (10.8) konvergiert dann auch $d^q f(x_{m(j)})(h_{m(j)}, \dots, h_{m(j)})$ gegen $d^q f(x)(h, \dots, h)$, und wir erhielten $d^q f(x)(h, \dots, h) \geq 0$ im Widerspruch zur Voraussetzung.

Nach Satz 10.2.10 hat man für $w \in U_\delta(x)$

$$f(w) = f(x) + \sum_{l=1}^{q-1} \frac{1}{l!} df^l(x)(\underbrace{w - x, \dots, w - x}_{l-\text{mal}}) + R_{q-1}(w) = f(x) + R_{q-1}(w)\,,$$

$$R_{q-1}(w) = \frac{1}{(q-1)!} \int_0^1 df^q((1-t)x + tw)(\underbrace{w - x, \dots, w - x}_{q-\text{mal}})(1-t)^{q-1}\, dt\,.$$

Wegen der Wahl von δ ist der Integrand aber negativ. Also gilt $f(w) = f(x) + R_q(w) < f(x)$ für $w \neq x$, und somit ist x ein lokales Maximum.

Die Aussagen (*iii*) und (*iv*) zeigt man genauso wie (*i*) und (*ii*), und (*v*) folgt sofort aus (*i*) und (*iii*). ❑

Zur Bestimmung aller lokalen Extrema einer überall differenzierbaren Funktion braucht man nur unter jenen Punkten x mit $df(x) = 0$ suchen. In der Praxis wird man die Gleichung $df(x) = 0$ aber meistens nicht explizit nach x auflösen können.

10.3.7 Bemerkung. Ob ein lokales Maximum (Minimum) einer Funktion $f : D \to \mathbb{R}$ mit offenem $D \subseteq \mathbb{R}^n$ ein globales Maximum (Minimum) ist, ist im Allgemeinen nicht leicht zu beantworten. Zum Beispiel hat $g : \mathbb{R}^2 \to \mathbb{R}$,

$$g((\xi, \eta)^T) = (\xi^2 + \eta^2)^2 - 2(\xi^2 + \eta^2)$$

bei $(0, 0)^T$ ein lokales Maximum. Wegen $\lim_{\|(\xi,\eta)^T\|\to+\infty} g((\xi, \eta)^T) = +\infty$ hat sie jedoch kein globales Maximum.
Ist aber f stetig fortsetzbar auf den Abschluss $c(D)$ von D und ist D beschränkt, so muss wegen der daraus resultierenden Kompaktheit von $c(D)$ die Funktion f auf $c(D)$ mindestens ein globales Maximum (Minimum) y haben. Diese Maximum (Minimum) kann nun in D oder im Rand $c(D) \setminus D$ liegen. Im ersten Fall ist dann y klarerweise auch Maximum (Minimum) von $f|_D$ und somit eine Lösung der Gleichung $df(x) = 0$. Im zweiten Fall ist y insbesondere ein globales Maximum (Minimum) von $f|_{c(D)\setminus D}$.
Um also von einer gegebenen Funktion $f : c(D) \to \mathbb{R}$ mit beschränktem und offenem D die globalen Maxima (Minima) auf $c(D)$ zu suchen, sind zunächst alle Lösungen der Gleichung $df(x) = 0$ auf D zu suchen. Dann sucht man alle Maxima (Minima) von $f|_{c(D)\setminus D}$ auf dem Rand $c(D) \setminus D$ und vergleicht schließlich die Funktionswerte an allen erhaltenen Punkten von $c(D)$.

10.3.8 Beispiel. Wir betrachten die Funktion

$$f : \begin{cases} \mathbb{R}^2 & \to \quad \mathbb{R}\,, \\ (\xi_1, \xi_2)^T & \mapsto \quad 1 - \xi_1^2 - \xi_2^2\,. \end{cases}$$

Der Graph $\{(\xi_1, \xi_2, \xi_3)^T \in \mathbb{R}^3 : \xi_3 = f((\xi_1, \xi_2)^T)\}$ dieser Funktion ist ein Paraboloid.

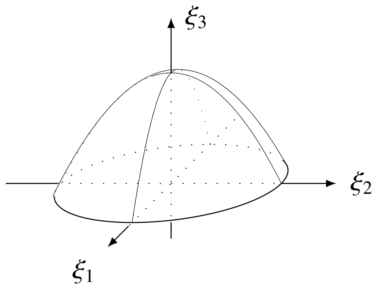

Abbildung 10.2: Graph der Funktion $\xi_3 = f\binom{\xi_1}{\xi_2}$ für $\xi_1^2 + \xi_2^2 \leq 1$

$\rightsquigarrow$ Die Ableitung von f berechnet sich als

$$df(x) = \Big(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\Big) = (-2\xi_1, -2\xi_2)\,.$$

Schreibt man $y \in \mathbb{R}^2$ in Koordinaten $(\eta_1, \eta_2)^T$ bezüglich der kanonischen Basis, so ergibt sich

$$df(x)y = -2\xi_1\eta_1 - 2\xi_2\eta_2\,.$$

Der einzige Punkt x mit $df(x) = (0,0)$ ist offenbar $x = (0,0)^T$. Die zweiten partiellen Ableitungen von f sind

$$\frac{\partial^2 f}{\partial x_1^2} = -2,\ \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 0,\ \frac{\partial^2 f}{\partial x_2^2} = -2\,.$$

Die zweite Ableitung ist daher gleich

$$d^2 f(x)(h,h) = -2h_1^2 - 2h_2^2\,,$$

und sie ist offenbar stets < 0, also negativ definit. Insbesondere ist $(0,0)^T$ ein lokales Maximum. Ist 0 nun auch ein globales Maximum? Das Bild, das man von der Fläche hat, legt das sicherlich nahe. Aber wie lässt sich das zeigen?

$\rightsquigarrow$ Eine Möglichkeit ist, festzustellen, dass für $(\xi_1, \xi_2)^T \neq (0,0)^T$ sicherlich $f((\xi_1, \xi_2)^T) = 1 - (\xi_1^2 + \xi_2^2) < 1 = f((0,0)^T)$.

$\rightsquigarrow$ Man kann auch nach einem (globalen, lokalen) Minimum von f fragen. Ein globales Minimum gibt es offensichtlich nicht, da z.B. für festes ξ_2 und für $|\xi_1| \to +\infty$ der Ausdruck $f((\xi_1, \xi_2)^T)$ nach $-\infty$ strebt.

Man kann aber z.B. fragen, ob $f|_{\mathbb{D}}$ mit $\mathbb{D} = \{(\xi_1, \xi_2)^T \in \mathbb{R}^2 : \|(\xi_1, \xi_2)^T\|_2 < 1\}$ ein lokales oder globales Minimum $y = (\eta_1, \eta_2)^T$ hat. Das ist aber auch nicht der Fall, denn dann wäre dort $df(y) = 0$, was aber nur für $y = 0$ der Fall ist.

$\rightsquigarrow$ Nun kann man z.B. auch fragen, ob $f|_{\overline{\mathbb{D}}}$ mit $\overline{\mathbb{D}} = \{(\xi_1, \xi_2)^T \in \mathbb{R}^2 : \|(\xi_1, \xi_2)^T\|_2 \leq 1\}$ ein Minimum $y = (\eta_1, \eta_2)^T$ hat. Man beachte, dass man hier nicht unmittelbar die Resultate dieses Abschnittes anwenden kann, da ja $\overline{\mathbb{D}}$ nicht offen ist. In der Tat ist $\overline{\mathbb{D}}$ kompakt, und somit muss f auf $\overline{\mathbb{D}}$ mindestens ein Minimum haben.

Ein solches kann aber nicht in $\mathbb{D}$ liegen, da es dann ein Minimum von $f|_{\mathbb{D}}$ wäre, von dem wir ja wissen, dass es diese nicht gibt.

Also muss $y \in \overline{\mathbb{D}} \setminus \mathbb{D} = \mathbb{T} = \{(\xi_1, \xi_2)^T \in \mathbb{R}^2 : \|(\xi_1, \xi_2)^T\|_2 = 1\}$. Aber f ist auf $\mathbb{T}$ konstant. Also sind alle $y \in \mathbb{T}$ Minima von $f|_{\overline{\mathbb{D}}}$.

$\rightsquigarrow$ Wollen wir die Extrema von f auf dem abgeschlossenen Dreieck Δ mit den Eckpunkten $(0,1)^T, (2,0)^T, (2,2)^T$ finden, so gehen wir folgendermaßen vor.

Zunächst existieren zumindest ein Minimum und zumindest ein Maximum von f auf Δ, da f stetig und Δ abgeschlossen, beschränkt und somit kompakt ist.

Das Innere Δ^o von Δ ist eine offene Teilmenge von $\mathbb{R}^2$, welche $(0,0)^T$ nicht enthält. Also kann ein Extremum von f nicht in Δ^o liegen, da df dort verschwinden müsste. Wir haben aber schon gesehen, dass $(0,0)^T$ der einzige Punkt ist, wo df verschwindet.

Also müssen die Extrema am Rand $\Delta \setminus \Delta^o$, also in der Vereinigung der drei Seiten

$$a = \{\begin{pmatrix} 2\alpha \\ 1-\alpha \end{pmatrix} : \alpha \in [0,1]\}, \quad b = \{\begin{pmatrix} 2\alpha \\ 1+\alpha \end{pmatrix} : \alpha \in [0,1]\}, \quad c = \{\begin{pmatrix} 2 \\ 2\alpha \end{pmatrix} : \alpha \in [0,1]\},$$

suchen. Die Extrema von f auf a, b und c sind aber die Extrema von

$$\begin{aligned} f_a(\alpha) &= f\begin{pmatrix} 2\alpha \\ 1-\alpha \end{pmatrix} = 1 - 4\alpha^2 - 1 + 2\alpha - \alpha^2 = -5\alpha^2 + 2\alpha, \\ f_b(\alpha) &= f\begin{pmatrix} 2\alpha \\ 1+\alpha \end{pmatrix} = 1 - 4\alpha^2 - 1 - 2\alpha - \alpha^2 = -5\alpha^2 - 2\alpha, \quad \text{bzw.} \\ f_c(\alpha) &= f\begin{pmatrix} 2 \\ 2\alpha \end{pmatrix} = 1 - 4 - 4\alpha^2 = -3 - 4\alpha^2, \end{aligned}$$

jeweils auf $[0,1]$.

Für f_a gilt $f_a'(\alpha) = -10\alpha + 2$ und $f_a''(\alpha) = -10$. Also ist $\alpha = \frac{1}{5}$ ein lokales Maximum mit $f_a(\frac{1}{5}) = \frac{1}{5}$. Für die Randpunkte von $[0,1]$ gilt $f_a(0) = 0$, $f_a(1) = -3$. Insgesamt hat f_a auf $[0,1]$ ein Maximum bei $\alpha = \frac{1}{5}$ und ein Minimum bei $\alpha = 1$.

Die Funktion f_b ist auf $[0,1]$ offensichtlich monoton fallend. Also ist $\alpha = 0$ ein Maximum mit $f_b(0) = 0$ und $\alpha = 1$ ein Minimum mit $f_b(1) = -7$.

Die Funktion f_c ist auf $[0,1]$ auch monoton fallend. Also ist $\alpha = 0$ ein Maximum mit $f_c(0) = -3$ und $\alpha = 1$ ein Minimum mit $f_c(1) = -7$.

Somit hat f auf $\Delta \setminus \Delta^o$ das Minimum bei $(2,2)^T$ mit Wert $f((2,2)^T) = -7$ und das Maximum bei $\begin{pmatrix} \frac{2}{5} \\ 1-\frac{2}{5} \end{pmatrix}$ mit Wert $\frac{1}{5}$.

10.4 Übungsaufgaben

10.1 Man betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}^2$, $f((\xi,\eta)^T) = (\xi^2\eta \sin \xi\eta, \frac{\xi}{\xi^2+\eta^2+1})^T$. Berechne alle partielle Ableitungen sowie $df(x)$, $x = (\xi,\eta)^T \in \mathbb{R}^2$. Schließlich berechne man die Richtungsableitung $\frac{\partial f}{\partial v}(x)$ für $v = (1,1)^T$ und $(1,-1)^T$.

10.2 Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch, also $A^T = A$, und betrachte die Abbildung $f(x) = x^T A x$ von $\mathbb{R}^n$ nach $\mathbb{R}$. Man zeige, dass $df(x) = 2(Ax)^T$, und berechne die Richtungsableitung entlang von $v = x$.

10.3 Sei $p \in \mathbb{N}$, $p \geq 2$ und $f : \mathbb{R}^p \setminus \{0\} \to \mathbb{R}$ definiert durch

$$f(x) = \ln \|x\|_2 \quad \text{im Falle} \quad p = 2 \quad \text{und} \quad f(x) = \frac{1}{(2-p)\|x\|_2^{p-2}} \quad \text{sonst}. \tag{10.14}$$

Man zeige, dass $\operatorname{grad} f(x) = (df(x))^T = \frac{1}{\|x\|_2^p} x$.

10.4 Man betrachte die Funktion f aus Beispiel 6.1.11 und zeige, dass in $(0,0)^T$ alle partiellen Ableitungen existieren, sie aber in $(0,0)^T$ nicht differenzierbar ist. Berechnen Sie von f auch alle partiellen Ableitungen zweiter Ordnung an allen Punkten der Ebene $\neq (0,0)^T$.

10.5 Man betrachte die Funktion $g : \mathbb{R}^2 \setminus \{(x,0) : x \leq 0\} \to \mathbb{R}^+ \times (-\pi, \pi)$,

$$g\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \sqrt{\xi^2 + \eta^2} \\ \arg(\xi + i\eta) \end{pmatrix},$$

wobei das Argument $\arg(\xi + i\eta) \in (-\pi, \pi)$ so definiert ist, dass $f((\sqrt{\xi^2 + \eta^2}, \arg(\xi + i\eta))^T) = (\xi, \eta)^T$, wobei $f((r, \phi)^T) = (r \cos\phi, r \sin\phi)^T$.

Berechne alle partiellen Ableitungen sowie $dg(x)$ und $\det dg(x)$, $x \in \mathbb{R}^2$.

Hinweis: Stellen Sie $\arg(x + iy)$ mit Hilfe des Arcustangens bzw. Arcuscotangens quadrantenweise dar!

10.6 Berechnen Sie $I'(\alpha)$, wobei $I : \mathbb{R} \to \mathbb{R}$ gegeben ist durch $I(\alpha) = \int_{-\exp(\alpha)}^{\alpha^2} \cos(\alpha t^2)\, dt$.

10.7 Sei $h : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ definiert durch $h(t) = (\cos t)^{\sin t}$ man berechne h' auf 2 Arten. Zuerst direkt und dann mittels Anwendung der Kettenregel auf $f(t) = (\cos t,\ \sin t)^T$ und $g((\xi, \eta)^T) = \xi^\eta$. Man gebe auch geeignete offene Definitionsbereiche von f und g an!

10.8 Funktionen $f : D \to \mathbb{C} \cong \mathbb{R}^2$ mit offenem $D \subseteq \mathbb{C} \cong \mathbb{R}^2$, die stetig differenzierbar sind, und die zusätzlich die sogenannten *Cauchy-Riemannschen Differentialgleichungen*

$$\frac{\partial u}{\partial x}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{\partial v}{\partial y}\begin{pmatrix} x \\ y \end{pmatrix}, \quad \frac{\partial u}{\partial y}\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{\partial v}{\partial x}\begin{pmatrix} x \\ y \end{pmatrix},$$

wobei $u((x,y)^T) = \operatorname{Re} f(x + iy)$ und $v((x,y)^T) = \operatorname{Im} f(x + iy)$, erfüllen, nennt man holomorph.

Man zeige, dass $z \mapsto \exp(z)$, $z \mapsto z$ holomorph auf $D = \mathbb{C}$ und $z \mapsto \frac{1}{z}$ holomorph auf $D = \mathbb{C} \setminus \{0\}$ sind, indem man die Cauchy-Riemannschen Differentialgleichungen nachprüft!

10.9 Für die Abbildung $\tau : \mathbb{R}^2 \to \mathbb{R}^3$,

$$\tau\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{1 + x^2 + y^2} \begin{pmatrix} 2x \\ 2y \\ x^2 + y^2 - 1 \end{pmatrix}$$

aus Übungsbeispiel 6.37, berechne man $d\tau((x,y)^T) \in \mathbb{R}^{3\times 2}$ und zeige, dass $d\tau((x,y)^T) = \frac{2}{1+x^2+y^2} M$ für eine von $(x,y)^T$ abhängige Matrix $M \in \mathbb{R}^{3\times 2}$, welche immer orthogonal ist, also $M^T M = I$ bzw. äquivalent dazu $(Mw, Mw) = (w, w)$ für alle $w \in \mathbb{R}^2$.

10.10 Man betrachte die Funktion $f((x,y)^T) = x^3 - 2x^2y^2 + 4xy^3 + y^4 + 10$. Berechne alle partiellen Ableitungen $\frac{\partial^{k+l}}{\partial x^k \partial y^l} f((x,y)^T)$ mit $k, l = 0, 1, 2$, sowie $d^m f((x,y)^T)(v_1, \ldots, v_m)$ für $m = 1, 2$. Schließlich berechne man das Taylorsche Polynom (in x, y) mit $q = 2$ gemäß Satz 10.2.10 an der Anschlussstelle $(0,0)^T$.

10.11 Man berechne alle partiellen Ableitungen 1., 2. und 3. Grades von $f((x,y,z)^T) = \sin(3x + yz)$, sowie $d^m f((x,y,z)^T)(v_1, \ldots, v_m)$ für $m = 1, 2$. Schließlich berechne man das Taylorsche Polynom (in x, y, z) mit $q = 3$ gemäß Satz 10.2.10 an der Anschlussstelle $(0,0)^T$.

10.12 Sei $p \in \mathbb{N},\ p \geq 2$ und $f : \mathbb{R}^p \setminus \{0\} \to \mathbb{R}$ definiert durch (10.14). Man zeige, dass f harmonisch ist!

Weiters zeige man, dass für ein festes $x_0 \in \mathbb{R}^p$ mit einem $h : B \to X$ auch $x \mapsto h(x - x_0),\ x \in x_0 + B$ harmonisch ist.

Dabei heißt eine zweimal stetig differenzierbare Funktion $h : B \to X$ harmonisch, wenn

$$\Delta h(x) := \frac{\partial^2}{\partial x_1 \partial x_1} h(x) + \cdots + \frac{\partial^2}{\partial x_m \partial x_m} h(x) = 0$$

für alle $x \in B$. Hier ist X ein Banachraum und $B \subseteq \mathbb{R}^m$ offen.

10.13 Bestimmen Sie für $f((x, y)^T) = x e^{\sqrt{y}}$ die Ableitung in Richtung des Vektors $v = (\cos\varphi, \sin\varphi)^T$. Für welchen Winkel φ wird $\frac{\partial f((2,3)^T)}{\partial v}$ maximal? Bestimmen Sie ebenfalls die Ableitung entlang der Parabel $p(t) := (x(t), y(t))^T = (t, t^2)^T$, also die Richtungsableitung von f in Richtung $p'(t)$ im Punkt $p(t)$.

10.14 Lineare Regression: Seien $(x_i, y_i) \in \mathbb{R}^2,\ i = 1, \ldots, n$ endlich viele, fest vorgegebene Messdaten, wobei zumindest zwei verschiedene x_i auftreten. Man bestimme eine linear Funktion $f(x) = kx + d$, sodass der quadratische Abstand

$$\sum_{i=1}^{n} (f(x_i) - y_i)^2$$

minimal wird! Man betrachte also $F(k, d) = \sum_{i=1}^{n} (f(x_i) - y_i)^2$ als Funktion von (k, d) und finden Sie die Kandidaten für lokale Extrema!

Begründen Sie auch, warum der erhaltene Kandidat Tatsächlich ein Minimum ist, indem Sie u.a. $\lim_{\|(k,d)^T\| \to +\infty} F(k, d) = +\infty$ zeigen.

Hinweis zum letzten Teil: Zunächst ist $\sqrt{F(k, d)} \geq \sqrt{(kx_1 + d - y_1)^2 + (kx_2 + d - y_2)^2}$, wobei ohne Beschränkung der Allgemeinheit $x_1 \neq x_2$. Zeigen Sie nun, dass

$$\|((kx_1 + d - y_1), (kx_2 + d - y_2))^T\|_2 \geq \|(kx_1 + d, kx_2 + d)^T\|_2 - \|(y_1, y_2)^T\|_2$$

und dann $\|A^{-1}\| \cdot \|((kx_1 + d), (kx_2 + d))^T\|_2 \geq \|(k, d)^T\|_2$, wobei $A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \end{pmatrix}$.

10.15 Sei $D \subseteq \mathbb{R}^n$ offen und beschränkt. $c(D)$ sei der Abschluss von D in $\mathbb{R}^n$. Weiters sei $f : c(D) \to \mathbb{R}$ stetig, sodass $f|_D : D \to \mathbb{R}$ aus C^2 ist mit $\Delta f(x) \geq 0$ für alle $x \in D$, wobei

$$\Delta f(x) := \frac{\partial^2}{\partial x_1 \partial x_1} f(x) + \cdots + \frac{\partial^2}{\partial x_n \partial x_n} f(x)\,.$$

Zeigen Sie, dass $f : c(D) \to \mathbb{R}$ in $c(D)$ und $f|_{c(D)\setminus D} : c(D) \setminus D \to \mathbb{R}$ in $c(D) \setminus D$ jeweils mindestens eine Maximalstelle haben und dass $\max_{t \in c(D)} f(t) = \max_{t \in c(D)\setminus D} f(t)$.

Hinweis: Angenommen $x_0 \in D$ wäre Maximalstelle mit $f(x_0) > \max_{t \in c(D)\setminus D} f(t) =: \eta$. Ohne Beschränkung der Allgemeinheit sei $f(x_0) > 0 > \eta$ angenommen. Betrachte eine Maximalstelle von $g(x) = f(x) + c(\|x - x_0\|_2^2 - d)$ für $x \in c(D)$ mit geeignet gewählten $c, d > 0$, sodass $\|x - x_0\|_2^2 - d < 0$ für $x \in c(D) \setminus D$ und sodass $g(x_0) > 0$! Warum geht das? Wo können die Maximalstellen x_1 von g nur liegen und was gilt für $\Delta g(x_1)$?

10.16 Wo besitzt $f : \mathbb{R}^2 \to \mathbb{R}$ ein globales bzw. lokales Extremum, wobei

$$f\begin{pmatrix} x \\ y \end{pmatrix} = x^3 e^{x-y} .$$

10.17 Zeigen Sie, dass $f : \mathbb{R}^2 \to \mathbb{R}$, definiert durch

$$f\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2(1-x)^3 ,$$

genau ein lokales Extremum, aber kein globales Extremum besitzt.

10.18 Wo besitzt $f : [0,1] \times [0,1] \to \mathbb{R}$ ein globales Extremum, wobei

$$f\begin{pmatrix} x \\ y \end{pmatrix} = 3x^2 - 2(y+1)x + 3y - 1 .$$

Hinweis: Kandidaten für Extrema in $(0,1) \times (0,1)$ erfüllen $df((x,y)^T) = 0$. Kandidaten in $(0,1) \times \{0\}$ erfüllen $\frac{\partial f}{\partial x}((x,0)^T) = 0$ usw. .

10.19 Wie im letzten Beispiel, aber für $f : \{(x,y)^T \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} \to \mathbb{R}$

$$f\begin{pmatrix} x \\ y \end{pmatrix} = 8x^2 - 2yx + 3y - 1 .$$

10.20 Bestimmen Sie für die Funktion

$$f(z) = \bar{z}(z-2) - 2\,\mathrm{Re}\, z , \qquad z = x + \mathrm{i}y \in \mathbb{C} ,$$

alle lokalen Extrema der Funktion $|f(z)|$, und geben Sie an, ob es sich um ein Minimum oder Maximum handelt!

10.21 Bestimmen Sie alle (lokalen) Extrema der Funktionen $f, g : \mathbb{R}^2 \to \mathbb{R}$. Sind diese (lokale) Minima bzw. Maxima?

$$f\begin{pmatrix} x \\ y \end{pmatrix} = (x^2 + y^2)e^{x^2 - y^2}, \quad g\begin{pmatrix} x \\ y \end{pmatrix} = (y^2 - x^2)(y^2 - 2x^2) .$$

10.22 Sei $K = \{(\cos t, \sin t)^T \in \mathbb{R}^2 : t \in \mathbb{R}\}$ und $A = \{(\xi, \eta)^T : 2\xi + 3\eta = 10\}$. Man zeige, dass K kompakt und A abgeschlossen ist. Weiters bestimme man $x \in K, y \in A$, sodass $d(x,y) = d(A,K)$; vgl. Übungsbeispiel 5.21. Schließlich zeige man, dass x normal auf die Gerade A steht.