In einem normierten Raum $(V, \|\cdot\|)$ ist für $x, y \in V$ die durch $t \mapsto (1-t)x + ty$ gegebene Funktion $\gamma : [0,1] \to V$ stetig, also ein Weg von x nach y. Damit ist die Strecke $[x,y] = \gamma([0,1])$ das Bild eines Weges. Auch Streckenzüge sind Bilder von Wegen.

Satz 6.6.4 *Sei $(V, \|\cdot\|)$ normierter Raum und $M \subset V$ offen und zusammenhängend. Dann existiert für alle $x, y \in M$ ein in M verlaufender Streckenzug von x nach y, d. h., es existieren $x_1, x_2, \ldots, x_n \in M$ mit $x_1 = x$, $x_n = y$ und $[x_1, x_2, \ldots, x_n] \subset M$. Insbesondere ist M wegzusammenhängend.*

Beweis. Sei $x \in M$. Wir betrachten die Menge U aller $y \in M$, für die ein in M verlaufender Streckenzug von x nach y existiert. Zu zeigen ist, dass $U = M$ gilt.

Wir zeigen zunächst, dass U offen ist. Sei dazu $y \in U$ und $[x_1, x_2, \ldots, x_n]$ ein in M verlaufender Streckenzug von x nach y. Da M offen und $y \in U \subset M$, existiert $\varepsilon > 0$ mit $U(y, \varepsilon) \subset M$. Für $z \in U(y, \varepsilon)$ ist dann $[x_1, x_2, \ldots, x_n, z]$ ein in M verlaufender Streckenzug von x nach z. Es folgt $U(y, \varepsilon) \subset U$ und damit ist U offen.

Ein analoges Argument zeigt, dass auch $V := M \backslash U$ offen ist. Wegen $U \neq \emptyset$, $U \cup V = M$ und $U \cap V = \emptyset$ folgt nun aus dem Zusammenhang von M, dass $V = \emptyset$, also $U = M$. $\square$

7 Differentialrechnung in mehreren Veränderlichen

7.1 Partielle Ableitungen

Definition 7.1.1 Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f : U \to \mathbb{R}$. Für $j \in \{1, \ldots, p\}$ heißt dann f *partiell differenzierbar* in ξ nach der j-ten Variablen (oder nach x_j), falls

$$\begin{aligned} & \partial_j f(\xi) \\ := & \lim_{x_j \to \xi_j} \frac{f(\xi_1, \ldots, \xi_{j-1}, x_j, \xi_{j+1}, \ldots, \xi_p) - f(\xi_1, \ldots, \xi_{j-1}, \xi_j, \xi_{j+1}, \ldots, \xi_p)}{x_j - \xi_j} \\ = & \lim_{h \to 0} \frac{f(\xi_1, \ldots, \xi_{j-1}, \xi_j + h, \xi_{j+1}, \ldots, \xi_p) - f(\xi_1, \ldots, \xi_{j-1}, \xi_j, \xi_{j+1}, \ldots, \xi_p)}{h} \end{aligned}$$

existiert. Der Grenzwert heißt dann *partielle Ableitung* (von f nach der j-ten Variablen an der Stelle ξ) und wird auch mit $\frac{\partial f}{\partial x_j}(\xi)$ oder $f_{x_j}(\xi)$ bezeichnet.

Definiert man für ein geeignetes Intervall I, welches ξ_j als inneren Punkt enthält, $g : I \to \mathbb{R}$ durch $g(x) = f(\xi_1, \ldots, \xi_{j-1}, x, \xi_{j+1}, \ldots, \xi_p)$, so existiert also $\partial_j f(\xi)$ genau dann, wenn $g'(\xi_j)$ existiert, und es gilt dann $\partial_j f(\xi) = g'(\xi_j)$. Man berechnet also $\partial_j f$ durch Differenzieren nach x_j bei festgehaltenem $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p$.

Ist $p = 2$ bzw. $p = 3$, so bezeichnet man die Variablen statt mit x_1, x_2 (und x_3) in der Regel mit x, y bzw. x, y, z. Dementsprechend schreibt man auch zum Beispiel $\frac{\partial f}{\partial y}$ und f_y statt $\frac{\partial f}{\partial x_2}$ und f_{x_2}. Analog verfährt man bei anderen Bezeichnungen für die Variablen.

Statt $\partial_j f(\xi)$ findet man in der Literatur auch die Bezeichnung $D_j f(\xi)$. Üblich (aber gelegentlich problematisch) sind auch Schreibweisen wie $\frac{\partial f(x,y)}{\partial x}$ oder $\frac{\partial f}{\partial x}(x,y)$ für $\partial_1 f(x,y)$.

Beispiele. 1. Es sei $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) = xe^y + \sin(xy)$. Dann ist $\partial_1 f(x,y) = e^y + y\cos(xy)$, $\partial_2 f(x,y) = xe^y + x\cos(xy)$.

2. Es sei $f : \mathbb{R}^p \backslash \{0\} \to \mathbb{R}$, $f(x) = \frac{1}{\|x\|_2^2}$, also

$$f(x_1, \ldots, x_p) = \frac{1}{x_1^2 + \cdots + x_p^2}.$$

Dann gilt

$$\partial_j f(x_1, \ldots, x_p) = -\frac{2x_j}{(x_1^2 + \cdots + x_p^2)^2}.$$

Ist $U \subset \mathbb{R}^p$ offen, $k \in \{1, \ldots, p\}$, $f : U \to \mathbb{R}$ und existiert $\partial_k f(x)$ für alle $x \in U$, so ist durch $x \mapsto \partial_k f(x)$ eine Funktion $\partial_k f : U \to \mathbb{R}$ gegeben. Man kann auch diese wieder auf partielle Differenzierbarkeit untersuchen. Für $\partial_j \partial_k f$ schreiben wir dabei kurz ∂_{jk} (oder $\partial_{j,k}$), bzw. $\frac{\partial^2 f}{\partial x_j \partial x_k}$ oder $f_{x_k x_j}$.

Entsprechend definiert man höhere partielle Ableitungen, etwa

$$\partial_1 \partial_3^2 \partial_2 f = \frac{\partial^4 f}{\partial x_1 \partial x_3^2 \partial x_2} = f_{x_2 x_3 x_3 x_1}.$$

Die Anzahl der auftretenden Ableitunge wird also *Ordnung* der Ableitung bezeichnet, im angegebenen Beispiel handelt es sich also um eine partielle Ableitung 4. Ordnung.

Beispiele. 1. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = xe^y + \sin(xy)$ wie oben, mit $\partial_1 f(x,y) = e^y + y\cos(xy)$, $\partial_2 f(x,y) = xe^y + x\cos(xy)$. Es ist

$$\begin{aligned}
\partial_{11} f(x,y) &= -y^2 \sin(xy), \\
\partial_{21} f(x,y) &= e^y + \cos(xy) - xy\sin(xy), \\
\partial_{12} f(x,y) &= e^y + \cos(xy) - xy\sin(xy) = \partial_{21} f(x,y), \\
\partial_{22} f(x,y) &= xe^y - x^2 \sin(xy).
\end{aligned}$$

2. Sei $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{falls } (x,y) \neq (0,0), \\ 0 & \text{falls } (x,y) = (0,0), \end{cases}$$

Dann gilt

$$\partial_1 f(0,y) = \lim_{x \to 0} \frac{f(x,y)}{x} = \left\{ \begin{array}{ll} 0 & \text{falls } y = 0, \\ \lim\limits_{x \to 0} y\dfrac{x^2 - y^2}{x^2 + y^2} & \text{falls } y \neq 0 \end{array} \right\} = -y$$

und analog $\partial_2 f(x,0) = x$. Es folgt $\partial_{12} f(0,0) = 1 \neq -1 = \partial_{21} f(0,0)$.

Satz 7.1.1 (Schwarz) *Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$, $f : U \to \mathbb{R}$ und $j, k \in \{1, \ldots, p\}$. Es sei weiter vorausgesetzt, dass die zweiten partiellen Ableitungen ∂_{jk} und ∂_{kj} in U existieren und in ξ stetig sind. Dann gilt $\partial_{jk}(\xi) = \partial_{kj}(\xi)$.*

Beweis. Ohne Beschränkung der Allgemeinheit sei $p = 2$, $j = 1$, $k = 2$. Zu zeigen ist also $\partial_{12}(\xi) = \partial_{21}(\xi)$.

Sei $\varepsilon > 0$ mit $U(\xi, \varepsilon) \subset U$ (wobei $U(\xi, \varepsilon) \subset U$ mit der Norm $\|\cdot\|_\infty$ gebildet sei). Es sei $\xi = (a, b)$ und

$$F : U(\xi, \varepsilon) \to \mathbb{R},\ F(x, y) = f(x, y) - f(x, b) - f(a, y) + f(a, b).$$

Für festes $y \in (b - \varepsilon, b + \varepsilon)$, $y \neq b$, betrachten wir

$$\phi : (a - \varepsilon, a + \varepsilon) \to \mathbb{R},\ \phi(x) = f(x, y) - f(x, b).$$

Es ist dann $F(x, y) = \phi(x) - \phi(a)$. Dann ist ϕ differenzierbar und nach Mittelwertsatz existiert für $x \neq a$ dann α_1 zwischen a und x mit

$$F(x, y) = \phi(x) - \phi(a) = (x - a)\phi'(\alpha_1) = (x - a)\left(\partial_1 f(\alpha_1, y) - \partial_1 f(\alpha_1, b)\right).$$

Erneute Anwendung des Mittelwertsatzes liefert β_1 zwischen b und y mit

$$F(x, y) = (x - a)(y - b)\partial_{21} f(\alpha_1, \beta_1).$$

Jetzt betrachten wir für festes $x \in (a - \varepsilon, a + \varepsilon)$, $x \neq a$, die Funktion

$$\psi : (b - \varepsilon, b + \varepsilon) \to \mathbb{R},\ \psi(x) = f(x, y) - f(a, y).$$

Man findet für $y \neq b$ dann β_2 zwischen b und y mit

$$F(x, y) = \psi(y) - \psi(b) = (y - b)\psi'(\beta_2) = (y - b)\left(\partial_2 f(x, \beta_2) - \partial_2 f(a, \beta_2)\right)$$

und anschließend α_2 zwischen a und x mit

$$F(x, y) = (x - a)(y - b)\partial_{12} f(\alpha_2, \beta_2).$$

Es folgt

$$\partial_{21} f(\alpha_1, \beta_1) = \partial_{12} f(\alpha_2, \beta_2).$$

Da für $(x, y) \to (a, b)$ auch $(\alpha_1, \beta_1) \to (a, b)$ und $(\alpha_2, \beta_2) \to (a, b)$ gilt, folgt die Behauptung jetzt aus der Stetigkeit von $\partial_{12} f$ und $\partial_{21} f$ in $\xi = (a, b)$. $\square$

Satz 7.1.2 *Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f : U \to \mathbb{R}$. Die partiellen Ableitungen von f mögen existieren und sie seien in $\xi = (\xi_1, \ldots, \xi_p)^T$ stetig. Dann gilt*

$$\frac{f(x) - f(\xi) - \sum_{j=1}^{p} \partial_j f(\xi)(x_j - \xi_j)}{\|x - \xi\|} \to 0$$

für $x = (x_1, \ldots, x_p)^T \to \xi$. Desweiteren ist f stetig in ξ.

Bemerkungen. 1. Wegen der Äquivalenz der Normen in $\mathbb{R}^p$ (Satz 6.5.6) ist es irrelevant, welche Norm in Satz 7.1.2 betrachtet wird.

2. Aus der Existenz der partiellen Ableitungen allein folgt noch nicht die Stetigkeit, man vgl. etwa das Beispiel nach Satz 6.4.5 oder betrachte eine beliebige Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, die $f(0,t) = f(t,0) = 0$ für $t \in \mathbb{R}$ erfüllt. Dann gilt $\partial_1 f(0,0) = \partial f_2(0,0) = 0$, aber f braucht nicht stetig sein.

3. Mit den Landau-Symbolen lässt sich die Aussage von Satz 7.1.2 auch in der Form

$$f(x) - f(\xi) = \sum_{j=1}^{p} \partial_j f(\xi)(x_j - \xi_j) + o\left(\|x - \xi\|\right)$$

für $x \to \xi$ schreiben. Sie lässt sich wie folgt interpretieren: die durch $(h_1, \ldots, h_n) \mapsto \sum_{j=1}^{p} \partial_j f(\xi) h_j$ definierte lineare Funktion $T : \mathbb{R}^p \to \mathbb{R}$ ist (in der Nähe von 0) eine gute Approximation der durch $h \mapsto f(\xi + h) - f(\xi)$ definierten Funktion, denn

$$f(\xi + h) - f(\xi) = T(h) + o(\|h\|)$$

für $h \to 0$. Wir können dies auch als

$$f(x) - f(\xi) = T(x - \xi) + o(\|x - \xi\|)$$

für $x \to \xi$ schreiben. Leicht sieht man, dass T die einzige lineare Funktion mit dieser Eigenschaft ist. Wir werden diese lineare Abbildung T später Ableitung von f nennen.

Der Graph der linearen Funktion $T : \mathbb{R}^p \to \mathbb{R}$ ist ein p-dimensionaler Teilraum des $\mathbb{R}^{p+1}$, eine sogenannte Hyperebene (durch 0), im Falle $p = 2$ also eine Ebene im $\mathbb{R}^3$ (durch 0). Für $f(x,y) = \sin(x - y) + x^2 + y^4$ und $\xi = (0,0)$ ist der Graph von f und T in Abbildung 27 dargestellt.

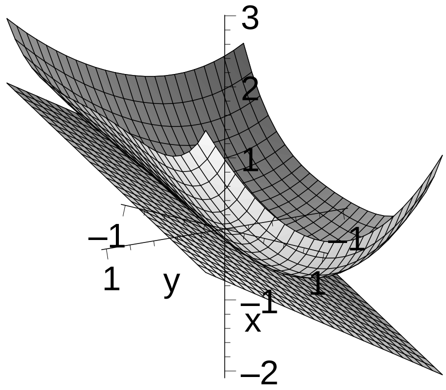

Abbildung 27: Approximation des Graphen durch eine Ebene.

Beweis von Satz 7.1.2. Sei $\varepsilon > 0$ mit $U(\xi, \varepsilon) \subset U$ (wobei $U(\xi, \varepsilon) \subset U$ wieder mit der Norm $\|\cdot\|_\infty$ gebildet sei). Für $0 < \|x - \xi\|_\infty < \varepsilon$ gilt dann

$$\begin{aligned} f(x) - f(\xi) &= f(x_1, \ldots, x_p) - f(\xi_1, \ldots, \xi_p) \\ &= f(x_1, \ldots, x_p) - f(\xi_1, \ldots, \xi_{p-1}, x_p) \\ &\quad + f(\xi_1, \ldots, \xi_{p-1}, x_p) - f(\xi_1, \ldots, \xi_{p-2}, x_{p-1}, x_p) \end{aligned}$$

$$
\begin{aligned}
& +\dots \\
& +f(\xi_1, x_2, \dots, x_p) - f(x_1, \dots, x_p) \\
= & \sum_{j=1}^{p} f(\xi_1, \dots, \xi_j, x_{j+1}, \dots, x_p) - f(\xi_1, \dots, \xi_{j-1}, x_j, \dots, x_p) \\
= & \sum_{j=1}^{p} (x_j - \xi_j) \partial_j f(\xi_1, \dots, \xi_{j-1}, y_j, x_{j+1}, \dots, x_p) \\
= & \sum_{j=1}^{p} (x_j - \xi_j) \partial_j f(\eta^j)
\end{aligned}
$$

mit y_j zwischen ξ_j und x_j, und damit also $\eta^j := (\xi_1, \dots, \xi_{j-1}, y_j, x_{j+1}, \dots, x_p)^T \in U(\xi, \|x-\xi\|_\infty)$ Mit T wie in obiger Bemerkung folgt also

$$
f(x) - f(\xi) - T(x-\xi) = \sum_{j=1}^{p} (x_j - \xi_j) \left(\partial_j f(\eta^j) - \partial_j f(\xi) \right)
$$

und damit

$$
\frac{|f(x) - f(\xi) - T(x-\xi)|}{\|x-\xi\|_\infty} \leq \sum_{j=1}^{p} \left| \partial_j f(\eta^j) - \partial_j f(\xi) \right| .
$$

Die erste Behauptung folgt jetzt aus der Stetigkeit der $\partial_j f$. Die Stetigkeit von f folgt da $T(x-\xi) \to 0$ für $x \to \xi$. $\square$

Die partiellen Ableitungen beschreiben das Änderungsverhalten einer Funktion in Richtung der Koordinatenachsen. Analoges lässt sich auch für andere Richtungen machen.

Definition 7.1.2 Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$, $f : U \to \mathbb{R}$ und $v \in \mathbb{R}^p$ mit $\|v\|_2 = 1$. Dann heißt

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

die *Richtungsableitung* von f an der Stelle ξ in Richtung v, falls der Grenzwert existiert.

Ist $\{e_1, \dots, e_p\}$ die Standardbasis des $\mathbb{R}^p$, so gilt also $\partial_{e_j} = \partial_j$ für alle j.

Satz 7.1.3 *Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$, $f : U \to \mathbb{R}$ und $v \in \mathbb{R}^p$ mit $\|v\|_2 = 1$. Die partiellen Ableitungen von f mögen existieren und sie seien in $\xi = (\xi_1, \dots, \xi_p)^T$ stetig. Dann existiert $\partial_v f(\xi)$ und es gilt*

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

Beweis. Nach Satz 7.1.2 gilt

$$
\lim_{t \to 0} \frac{f(\xi + tv) - f(\xi) - \sum_{j=1}^{p} \partial_j f(\xi) v_j t}{t} = 0,
$$

und daraus folgt die Behauptung. $\square$

Definition 7.1.3 Sei $p \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f : U \to \mathbb{R}$. Dann heißt $\mathrm{grad} f(\xi) := (\partial_1 f(\xi), \ldots, \partial_p f(\xi))$ der *Gradient* von f an der Stelle ξ, falls die partiellen Ableitungen existieren.

Unter den Voraussetzungen von Satz 7.1.3 gilt dann $\partial_v f(\xi) = \mathrm{grad} f(\xi) \cdot v$. (Hierbei steht "$\cdot$" für das Matrixprodukt der $(1 \times p)$-Matrix $\mathrm{grad} f(\xi)$ mit der $(p \times 1)$-Matix $(v_1, \ldots, v_p)^T$. Wir können dies auch als Skalarprodukt auffassen: $\partial_v f(\xi) = \langle \mathrm{grad} f(\xi), v \rangle$.)

In Satz 7.1.2 erhalten wir

$$\frac{f(x) - f(\xi) - \mathrm{grad} f(\xi)(x - \xi)}{\|x - \xi\|} \to 0$$

für $x \to \xi$.

Definition 7.1.4 Seien $p, q \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f = (f_1, \ldots, f_q)^T : U \to \mathbb{R}^q$. Dann heißt die $(q \times p)$-Matrix

$$J_f(\xi) := \begin{pmatrix} \partial_1 f_1(\xi) & \ldots & \partial_p f_1(\xi) \\ \partial_1 f_2(\xi) & \ldots & \partial_p f_2(\xi) \\ \vdots & & \vdots \\ \partial_1 f_q(\xi) & \ldots & \partial_p f_q(\xi) \end{pmatrix}$$

die *Funktionalmatrix* (oder *Jacobimatrix*) von f an der Stelle ξ, falls die partiellen Ableitungen existieren.

Die k-te Zeile von $J_f(\xi)$ ist also durch $\mathrm{grad} f_k(\xi)$ gegeben.

Satz 7.1.4 *Seien $p, q \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f = (f_1, \ldots, f_q)^T : U \to \mathbb{R}^q$. Die partiellen Ableitungen der f_k mögen existieren und sie seien in $\xi = (\xi_1, \ldots, \xi_p)^T$ stetig. Dann gilt*

$$\frac{f(x) - f(\xi) - J_f(\xi)(x - \xi)}{\|x - \xi\|} \to 0$$

für $x \to \xi$.

Beweis. Die Behauptung folgt unmittelbar aus Satz 7.1.2 (bzw. der auf Definition 7.1.3 folgenden Bemerkung), da die k-te Koordinate des Zählers durch $f_k(x) - f_k(\xi) - \mathrm{grad} f_k(\xi)(x - \xi)$ gegeben ist. $\square$

Für $U \subset \mathbb{R}^p$ offen und $n \in \mathbb{N}$ bezeichnet man die Menge aller Funktionen von U nach $\mathbb{R}^q$, für die alle partiellen Ableitungen bis zur Ordnung n existieren und stetig sind, mit $C^n(U, \mathbb{R}^q)$ oder, falls der Zielraum aus dem Zusammenhang klar ist, auch mit $C^n(U)$. Für solche Funktionen sind nach dem Satz von Schwarz die partiellen Ableitungen bis zur Ordnung n unabhängig von der Reihenfolge der Differentiation. Weiter setzt man $C^\infty(U, \mathbb{R}^q) := \bigcap_{n=0}^\infty C^n(U, \mathbb{R}^q)$.

7.2 Differenzierbarkeit

Reelle $(q \times p)$-Matrizen korrespondieren zu linearen Abbildungen von $\mathbb{R}^p$ nach $\mathbb{R}^q$. Insbesondere entspricht die Funktionalmatrix einer linearen Abbildung. Beachtet man noch, dass lineare Abbildungen von $\mathbb{R}^p$ nach $\mathbb{R}^q$ stetig sind (siehe §6.4), so führt dies auf folgende Definition.

Definition 7.2.1 Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $U \subset V$ offen, $\xi \in U$ und $f : U \to W$. Dann heißt f differenzierbar in ξ, falls $T \in L(V,W)$ existiert, so dass

$$\lim_{x\to\xi} \frac{f(x) - f(\xi) - T(x-\xi)}{\|x-\xi\|_V} = \lim_{h\to 0} \frac{f(\xi+h) - f(\xi) - T(h)}{\|h\|_V} = 0.$$

Die Abbildung T ist dann eindeutig bestimmt (siehe die 2. Bemerkung unten) und heißt *(totale) Ableitung* (oder *Fréchet-Ableitung*) von f an der Stelle ξ. Sie wird mit $Df(\xi)$ bezeichnet.

Die Funktion f heißt *(total) differenzierbar*, falls f in jedem Punkt von U differenzierbar ist. Ist dann die durch $x \mapsto Df(x)$ gegebene Funktion $Df : U \to L(V,W)$ stetig, heißt f *stetig differenzierbar*.

Bemerkungen. 1. Mit $L(V,W)$ hatten wir den Vektorraum der stetigen linearen Abbildungen von V nach W bezeichnet; vgl. §6.4.

2. Wir zeigen, dass T in obiger Definition eindeutig bestimmt ist. Dazu nehmen wir an, dass $T_1, T_2 \in L(V,W)$ mit $T_1 \neq T_2$ existieren, die die in der Definition angegebene Eigenschaft haben. Dann gilt

$$\lim_{h\to 0} \frac{T_1(h) - T_2(h)}{\|h\|_V} = \lim_{h\to 0} \frac{(T_1 - T_2)(h)}{\|h\|_V} = 0.$$

Andererseits existiert $y \in V$ mit $T_1(y) \neq T_2(y)$. Es folgt

$$\frac{(T_1 - T_2)(ty)}{\|ty\|_V} = \frac{t(T_1 - T_2)(y)}{t\|y\|_V} = \frac{(T_1 - T_2)(y)}{\|y\|_V}$$

für $t \in \mathbb{R}_+$ und damit

$$\lim_{t\to 0} \frac{(T_1 - T_2)(ty)}{\|ty\|_V} = \frac{(T_1 - T_2)(y)}{\|y\|_V} \neq 0,$$

ein Widerspruch.

3. Die Interpretation der in obiger Definition gegebenen Eigenschaft ist analog zu der von Satz 7.1.2; vgl. Bemerkung 3 dort. Die lineare Abbildung $Df(\xi)$ ist (in der Nähe von 0) eine gute Approximation der durch $h \mapsto f(\xi+h) - f(\xi)$ gegebenen Funktion. Denn ist $R(h) := f(\xi+h) - f(\xi) - Df(\xi)(h)$, so gilt $R(h) = o(\|h\|_V)$ für $h \to 0$. Es ist $Df(\xi)$ die einzige lineare Abbildung mit dieser Eigenschaft; vgl. die vorige Bemerkung.

4. Im allgemeinen hängt Df von den Normen $\|\cdot\|_V, \|\cdot\|_W$ ab. Dies gilt aber nicht, falls $V = \mathbb{R}^p$, $W = \mathbb{R}^q$, da dort alle Normen äquivalent sind (Satz 6.5.6).

Beispiel. Sei $f \in L(V,W)$. Dann ist f differenzierbar und es gilt $Df(\xi) = f$ für alle $\xi \in V$.

Satz 7.2.1 *Differenzierbare Funktionen sind stetig.*

Beweis. Seien U, V, W, f wie in Definition 7.2.1 und sei f differenzierbar in $\xi \in U$. Für $x \in U$ sei $R(x) := f(x) - f(\xi) - Df(\xi)(x - \xi)$. Nach Definition der Differenzierbarkeit gilt dann $R(x)/\|x - \xi\|_V \to 0$ für $x \to \xi$. Es folgt $R(x) \to 0$ für $x \to \xi$. Da $Df(\xi)$ stetig ist, folgt auch $Df(\xi)(x - \xi) \to 0$ für $x \to \xi$. Insgesamt folgt $f(x) \to f(\xi)$ für $x \to \xi$. Damit ist f stetig in ξ. $\square$

Man beachte, dass im Beweis die Stetigkeit von $Df(\xi)$ benutzt wird. Dies ist ein Grund dafür, in Definition 7.2.1 T als linear *und stetig* zu fordern.

Satz 7.2.2 *Seien $p, q \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen, $\xi \in U$ und $f = (f_1, \dots, f_q)^T : U \to \mathbb{R}^q$. Dann sind äquivalent:*

(i) *f ist differenzierbar in ξ.*

(ii) *Alle partiellen Ableitungen $\partial_k f_j(\xi)$ existieren und es gilt*

$$\frac{f(x) - f(\xi) - J_f(\xi)(x - \xi)}{\|x - \xi\|} \to 0$$

für $x \to \xi$.

Es gilt dann $Df(\xi)(h) = J_f(\xi)(h)$ für alle $h \in \mathbb{R}^p$, d. h., die Matrix $J_f(\xi)$ entspricht der linearen Abbildung $Df(\xi)$ bzgl. der Standardbasis.

Beweis. (ii) $\Rightarrow$ (i). Diese Richtung ist trivial, denn offensichtlich hat die durch $T(h) = J_f(\xi)h$ definierte Funktion $T \in L(V, W)$ die in Definition 7.2.1 verlangte Eigenschaft.

(i) $\Rightarrow$ (ii). Sei $A = (a_{ij})_{\substack{1 \le i \le q \\ 1 \le j \le p}}$ die zu $Df(\xi)$ gehörige Matrix (bzgl. der Standardbasen $\{e_1, e_2, \dots\}$). Dann folgt

$$\frac{f(\xi + h) - f(\xi) - Ah}{\|h\|} \to 0$$

für $h \to 0$ und damit

$$\frac{f_j(\xi + te_k) - f_j(\xi) - a_{jk}t}{t} \to 0$$

für $t \to 0$. Hieraus folgt, dass die partiellen Ableitungen $\partial_k f_j(\xi)$ existieren und dass $a_{jk} = \partial_k f_j(\xi)$ und $A = J_f(\xi)$ gilt. $\square$

Aus der totalen Differenzierbarkeit in ξ folgt also mit Satz 7.2.2 die partielle Differenzierbarkeit in ξ, aber nicht umgekehrt (vgl. die Beispiele nach Satz 7.1.2, wo die partiellen Ableitungen existieren, aber die Funktion sogar unstetig ist). Existieren aber die partiellen Ableitungen in einer Umgebung von ξ und sind sie in ξ stetig, so ist f in ξ total differenzierbar; vgl Satz 7.1.4 und 7.2.2.

Für stetige Differenzierbarkeit sind die Verhältnisse einfacher.

Satz 7.2.3 *Seien $p, q \in \mathbb{N}$, $U \subset \mathbb{R}^p$ offen und $f = (f_1, \dots, f_q)^T : U \to \mathbb{R}^q$. Dann ist f genau dann stetig differenzierbar, wenn $f \in C^1(U, \mathbb{R}^q)$, d. h., wenn alle partiellen Ableitungen der Koordinatenfunktionen von f existieren und stetig sind.*

Beweis. Da aus der Stetigkeit der partiellen Ableitungen die Differenzierbarkeit folgt, können wir f als differenzierbar annehmen. Für $h \in \mathbb{R}^p$ und $x, \xi \in U$ ist dann $(Df(x) - Df(\xi))(h) = (J_f(x) - J_f(\xi))(h)$. Legen wir in $\mathbb{R}^p, \mathbb{R}^q$ die Norm $\|\cdot\|_\infty$ zugrunde, so erhalten wir (vgl. das Beispiel nach Satz 6.4.6)

$$\|Df(x) - Df(\xi)\|_{L(\mathbb{R}^p,\mathbb{R}^q)} = \max_j \sum_{k=1}^p |\partial_k f_j(x) - \partial_k f_j(\xi)|.$$

Hieraus folgt leicht die Behauptung. $\square$

Die üblichen Regeln für die Differenzierbarkeit von $f + g$, und $\lambda \cdot f$, wobei $f, g : U \to W$ und $\lambda \in \mathbb{K}$, übertragen sich unmittelbar. Entsprechendes gilt auch für $f \cdot g$ und f/g, falls $W = \mathbb{R}$ ist; vgl. auch Satz 6.4.4. Wir formulieren die Kettenregel:

Satz 7.2.4 *Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$, $(X, \|\cdot\|_X)$ Banachräume, $U \subset V$ offen, $U' \subset W$ offen, $\xi \in U$, $g : U \to W$ differenzierbar in ξ, $g(U) \subset U'$ und $f : U' \to X$ differenzierbar in $g(\xi)$. Dann ist $f \circ g$ differenzierbar in ξ und*

$$D(f \circ g)(\xi) = Df(g(\xi)) \circ Dg(\xi).$$

Beweis. Es ist

$$R_g(h) := g(\xi + h) - g(\xi) - Dg(\xi)(h) = o(\|h\|_V)$$

für $h \to 0$ und

$$R_f(k) := f(g(\xi) + k) - f(g(\xi)) - Df(g(\xi))(k) = o(\|k\|_W)$$

für $k \to 0$. Mit $\Delta(h) := g(\xi + h) - g(\xi) = Dg(\xi)(h) + R_g(h)$ folgt

$$\begin{aligned}
& f(g(\xi + h)) - f(g(\xi)) \\
= \; & f(g(\xi) + \Delta(h)) - f(g(\xi)) \\
= \; & Df(g(\xi))(\Delta(h)) + R_f(\Delta(h)) \\
= \; & (Df(g(\xi)) \circ Dg(\xi))(h) + \underbrace{Df(g(\xi))(R_g(h)) + R_f(\Delta(h))}_{=:R(h)}.
\end{aligned}$$

Zu zeigen ist, dass $R(h) = o(\|h\|_V)$ für $h \to 0$. Wegen

$$\|Df(g(\xi))(R_g(h))\|_X \le \|Df(g(\xi))\|_{L(W,X)} \|R_g(h)\|_W = o(\|h\|_V)$$

für $h \to 0$ müssen wir nur noch $\|R_f(\Delta(h))\|_X = o(\|h\|_V)$ für $h \to 0$ zeigen. Dies folgt aber, da $\Delta(h) = O(\|h\|_V)$ und damit $\Delta(h) \to 0$ für $h \to 0$, und da

$$\frac{\|R_f(\Delta(h))\|_X}{\|h\|_V} = \underbrace{\frac{\|R_f(\Delta(h))\|_X}{\|\Delta(h)\|_W}}_{=o(1)} \cdot \underbrace{\frac{\|\Delta(h)\|_W}{\|h\|_V}}_{=O(1)}$$

falls $\Delta(h) \neq 0$. $\square$

Wir betrachten jetzt Abbildungen $f : I \to V$ wobei I Intervall und V Banachraum. Hier können wir die Ableitung wieder – wie in Analysis I – über den Differenzenquotienten bilden.

Definition 7.2.2 Sei I Intervall, $(V, \|\cdot\|)$ Banachraum, $f : I \to V$ und $\xi \in I$. Dann heißt

$$f'(\xi) := \lim_{x \to \xi} \frac{f(x) - f(\xi)}{x - \xi} = \lim_{h \to 0} \frac{f(\xi + h) - f(\xi)}{h}$$

Ableitung von f an der Stelle ξ, falls der Grenzwert existiert.

Andererseits können wir ein offenes Intervall (oder das Innere eines beliebigen Intervalls) als offene Teilmenge des Banachraums $\mathbb{R}$ auffassen und die Fréchet-Ableitung betrachten.

Satz 7.2.5 *Seien I, V, f wie in Definition 7.2.2 und $\xi \in \operatorname{int}(I)$. Dann existiert $f'(\xi)$ genau dann, wenn $Df(\xi)$ existiert, und es gilt dann $f'(\xi) = Df(\xi)(1)$.*

Beweis. Existiert $f'(\xi)$, so gilt

$$\lim_{h \to 0} \frac{f(\xi + h) - f(\xi) - hf'(\xi)}{h} = 0,$$

und damit folgt, dass $Df(\xi)$ existiert und durch $h \mapsto hf'(\xi)$ gegeben ist. Es folgt also $f'(\xi) = Df(\xi)(1)$. Analog folgt aus der Existenz von $Df(\xi)$ die von $f'(\xi)$. □

Mit den Bezeichnungen von Satz 7.2.5 gilt $\|f'(\xi)\| = \|Df(\xi)\|_{L(\mathbb{R},V)}$, falls $f'(\xi)$ (und damit $Df(\xi)$) existiert.

Der Mittelwertsatz der Differentialrechnung (Satz 4.2.2) gilt nur für reellwertige Funktionen. Für komplexwertige Funktionen hatten wir nur eine Ungleichung erhalten (Satz 4.2.5). Diese können wir auf die Banachraumsituation übertragen.

Satz 7.2.6 *Sei $(V, \|\cdot\|)$ Banachraum und seien $f : [a, b] \to V$ und $\varphi : [a, b] \to \mathbb{R}$ stetig und in (a, b) differenzierbar. Es gelte $\|f'(t)\| \leq \varphi'(t)$ für $t \in (a, b)$. Dann gilt $\|f(b) - f(a)\| \leq \varphi(b) - \varphi(a)$.*

Gilt

$$M := \sup_{t \in (a,b)} \|f'(t)\| < \infty,$$

so kann man hier $\varphi(t) = Mt$ wählen und erhält

$$\|f(b) - f(a)\| \leq M(b - a).$$

Diese Abschätzung heißt auch *Mittelwertungleichung.*

Beweis von Satz 7.2.6. Für $\varepsilon > 0$ betrachten wir

$$A := \{t \in [a, b] : \|f(t) - f(a)\| \leq \varphi(t) - \varphi(a) + \varepsilon(t - a) + \varepsilon\}.$$

Wir werden zeigen, dass $A = [a, b]$ gilt. Insbesondere gilt dann $b \in A$, und mit $\varepsilon \to 0$ folgt dann die Behauptung.

Um zu zeigen, dass $A = [a, b]$ gilt, beachten wir zunächst, dass A abgeschlossen ist. Dies folgt unmittelbar aus der Stetigkeit von f und φ. Außerdem folgt aus der Stetigkeit, dass eine Umgebung von a in A enthalten ist.

Sei nun $c := \sup\{t \in [a,b] : [a,t] \subset A\}$. Dann gilt $c \in A$ und $c \in (a,b]$. Wir nehmen an, dass $c < b$ gilt. Nach Definition der Differenzierbarkeit existiert $\delta > 0$ mit

$$\|f(t) - f(c) - f'(c)(t-c)\| < \frac{\varepsilon}{2}(t-c)$$

und

$$|\varphi(t) - \varphi(c) - \varphi'(c)(t-c)| < \frac{\varepsilon}{2}(t-c)$$

für $c \leq t \leq c + \delta$ und damit

$$\begin{aligned} \|f(t) - f(c)\| &\leq \|f'(c)\|(t-c) + \frac{\varepsilon}{2}(t-c) \\ &\leq |\varphi'(c)|(t-c) + \frac{\varepsilon}{2}(t-c) \\ &\leq \varphi(t) - \varphi(c) + \frac{\varepsilon}{2}(t-c) + \frac{\varepsilon}{2}(t-c) \\ &= \varphi(t) - \varphi(c) + \varepsilon(t-c). \end{aligned}$$

Außerdem gilt

$$\|f(c) - f(a)\| \leq \varphi(c) - \varphi(a) + \varepsilon(c-a) + \varepsilon$$

wegen $c \in A$. Es folgt

$$\|f(t) - f(a)\| \leq \|f(t) - f(c)\| + \|f(c) - f(a)\| \leq \varphi(t) - \varphi(a) + \varepsilon(t-a) + \varepsilon$$

für $c \leq t \leq c + \delta$ und damit $[a, t+\delta] \subset A$, ein Widerspruch. □

Satz 7.2.7 *Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $U \subset V$ offen, $a, b \in U$ mit $[a,b] \in U$ und $f : U \to W$ differenzierbar. Ist $M := \sup_{x \in [a,b]} \|Df(x)\|_{L(V,W)} < \infty$, so gilt $\|f(b) - f(a)\|_W \leq M\|b-a\|_V$.*

Beweis. Sei $\gamma : [0,1] \to [a,b]$, $t \mapsto a + t(b-a)$, und $g := f \circ \gamma$. Dann ist g differenzierbar in $(0,1)$ mit $Dg(t) = Df(\gamma(t)) \circ D\gamma(t)$ und damit

$$g'(t) = Dg(t)(1) = Df(\gamma(t))(D\gamma(t)(1)) = Df(\gamma(t))(\gamma'(t)) = Df(\gamma(t))(b-a).$$

Es folgt

$$\|g'(t)\|_W \leq \|Df(\gamma(t))\|_{L(V,W)}\|b-a\|_V \leq M\|b-a\|_V.$$

Mit Satz 7.2.6 folgt

$$\|f(b) - f(a)\|_W = \|g(1) - g(0)\|_W \leq M\|b-a\|_V. \quad \square$$

Auch die Ungleichung in Satz 7.2.7 nennt man Mittelwertungleichung. Eine typische Anwendung ist der folgende Satz (vgl. Satz 4.3.1).

Satz 7.2.8 *Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $U \subset V$ offen und zusammenhängend, $f : U \to W$ differenzierbar und $Df(x) = 0$ für alle $x \in U$. Dann ist f konstant, d. h., es existiert $c \in W$ mit $f(x) = c$ für alle $x \in U$.*

Beweis. Sei $x \in U$. Nach Satz 7.2.7 gilt dann $f(y) = f(x)$ für alle $y \in U$ mit $[x, y] \subset U$. Daraus folgt, dass $f(z) = f(x)$ für alle $z \in U$ gilt, die mit x durch einen in U verlaufenden Streckenzug verbunden werden können. Dies sind nach Satz 6.6.4 aber alle $z \in U$. $\square$

Auch der Begriff des Riemann-Integrals lässt sich ohne weiteres auf Funktionen $f : [a, b] \to V$, mit $(V, \|\cdot\|)$ Banachraum, übertragen. Für eine Zerlegung $Z = \{x_0, x_1, \ldots, x_n\}$ von $[a, b]$ mit Stützstellen $\xi = (\xi_1, \ldots, \xi_n)$ heißt $S(f, Z, \xi) = \sum_{k=1}^{n} f(\xi_k)|I_k|$ wieder Riemannsche Summe und f heißt Riemann-integrierbar, falls gilt:

$$\exists S \in V\ \forall \varepsilon \in \mathbb{R}_+ \exists \delta \in \mathbb{R}_+ \forall \text{ Zerlegungen } Z \text{ von } I\ \forall \text{ Stützstellen } \xi \text{ zu } Z:$$
$$|Z| < \delta \Rightarrow \|S(f, Z, \xi) - S\| < \varepsilon.$$

Dieses S ist dann wieder eindeutig und wird mit $\int_a^b f$ oder $\int_a^b f(x)dx$ bezeichnet. Wie in §5.1 gilt wieder das Cauchykriterium und wie dort zeigt man, dass stetige Funktionen integrierbar sind. Man erhält für stetiges f auch wieder

$$\left\| \int_a^b f \right\| \leq \int_a^b \|f\| \leq (b - a) \max_{x \in [a,b]} \|f(x)\|.$$

Für $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $U \subset V$ offen, $a, b \in U$ mit $[a, b] \in U$, $f : U \to W$ stetig differenzierbar, und $\gamma : [0, 1] \to [a, b]$, $t \mapsto a + t(b - a)$, gilt dann

$$f(b) - f(a) = \int_0^1 (f \circ \gamma)'(t)dt = \int_0^1 Df(\gamma(t))(b - a)dt.$$

Hieraus kann man (für *stetig* differenzierbares f) wieder die Aussage von Satz 7.2.7 gewinnen.

7.3 Höhere Ableitungen und Taylorformel

Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $U \subset V$ offen und $f : U \to W$ differenzierbar. Ist dann $Df : U \to L(V, W)$ wieder differenzierbar, so heißt $D(Df)$ *zweite Ableitung* und wird mit $D^2 f$ bezeichnet. Entsprechend definiert man die höheren Ableitungen $D^n f = D(D^{n-1} f)$, $n \geq 3$, falls sie existieren. Wir nennen dann f *n-mal differenzierbar* und, falls $D^n f$ auch stetig ist, *n-mal stetig differenzierbar.*

Es ist $D^2 f : U \to L(V, L(V, W))$, für $\xi \in U$ also $D^2 f(\xi) \in L(V, L(V, W))$. Für $x, y \in V$ gilt also $D^2 f(\xi)(x) \in L(V, W)$ und $(D^2 f(\xi)(x))(y) \in W$.

Der Vektorraum $L(V, L(V, W))$ kann in offensichtlicher Weise mit dem Vektorraum der stetigen, bilinearen Abbildungen von $V \times V$ nach W identifiziert werden. Aus diesem Grunde werden wir im folgenden $D^2 f(\xi)$ als stetige, bilineare Abbildung von $V \times V$ nach W betrachten und schreiben daher $D^2 f(\xi)(x, y)$ statt $(D^2 f(\xi)(x))(y)$.

Analog betrachten wir die höheren Ableitungen als stetige, multilineare Abbildungen und schreiben daher $D^n f(\xi)(x_1, \ldots, x_n)$ statt $(\ldots(D^n f(\xi)(x_1))\ldots)(x_n)$.

Wir wollen jetzt die Resultate aus §4.4 und §5.4 über die Taylorentwicklung übertragen. Für $\xi \in U$ und $x \in V$ setzen wir $h := x - \xi$ und definieren das n-te

Taylorpolynom durch

$$T_n(x) := f(\xi) + Df(\xi)(h) + \frac{1}{2!}D^2 f(\xi)(h,h) + \cdots + \frac{1}{n!}D^n f(\xi)(\underbrace{h,h,\ldots,h}_{n\text{-mal}})$$

und das Restglied $R_n(x)$ wieder durch

$$R_n(x) := f(x) - T_n(x).$$

Wir übertragen Satz 4.4.3. (Alternativ könnten wir auch Satz 5.4.1 entsprechend verallgemeinern.)

Satz 7.3.1 *Sei $(V, \|\cdot\|_V)$ Banachraum, $U \subset V$ offen, $n \in \mathbb{N}$, $\xi, x \in U$ mit $[\xi, x] \subset U$ und $h := x - \xi$. Die Funktion $f : U \to \mathbb{R}$ sei $(n+1)$-mal differenzierbar. Dann existiert $\eta \in [\xi, x]$ mit*

$$R_n(x) = \frac{1}{(n+1)!} D^{n+1} f(\eta)(\underbrace{h,h,\ldots,h}_{(n+1)\text{-mal}}).$$

Hilfssatz 7.3.1 *Seien $(V, \|\cdot\|_V)$, $(W, \|\cdot\|_W)$ Banachräume, $h \in V$ und $A : L(V,W) \to W$, $A(T) = T(h)$. Dann ist A differenzierbar und $DA(T) = A$ für alle $T \in L(V,W)$.*

Beweis. A ist linear und stetig, und nach Beispiel zu Definition 7.2.1 folgt die Behauptung. □

Beweis von Satz 7.3.1 Da $[\xi, x] \subset U$ und U offen existiert $\delta > 0$ mit $\xi + th = (1-t)\xi + tx \in U$ für $t \in I := (-\delta, 1+\delta)$. Wir betrachten $\gamma : I \to U$, $\gamma(t) = \xi + th = (1-t)\xi + tx$, und $g := f \circ \gamma$. Dann ist g nach Kettenregel $(n+1)$-mal differenzierbar und es gilt nach Satz 4.4.3

$$g(1) = g(0) + g'(0) + \frac{1}{2!}g''(0) + \cdots + \frac{1}{n!}g^{(n)}(0) + \frac{1}{(n+1)!}g^{(n+1)}(s)$$

mit $s \in (0,1)$. Nach Kettenregel und Satz 7.2.5 gilt

$$g'(t) = Dg(t)(1) = (Df(\gamma(t)) \circ D\gamma(t))(1) = Df(\gamma(t))\,(\gamma'(t)) = Df(\gamma(t))\,(h)$$

für $t \in I$. Mit A wie in Hilfssatz 7.3.1 gilt also

$$g'(t) = A(Df(\gamma(t))) = (A \circ Df \circ \gamma)(t).$$

Es folgt

$$\begin{aligned}
g''(t) &= Dg'(t)(1) \\
&= D(A \circ Df \circ \gamma)(t)(1) \\
&= \left(DA(\underbrace{(Df \circ \gamma)(t)}_{T}) \circ \big(D^2 f(\gamma(t))\big) \circ D\gamma(t) \right)(1) \\
&= \big(A \circ D^2 f\,(\gamma(t)\big) \circ D\gamma(t))(1) \\
&= \big(A \circ D^2 f\,(\gamma(t)\big)\,(\gamma'(t)) \\
&= \big(D^2 f\,(\gamma(t))\,(\gamma'(t))\big)\,(h) \\
&= \big(D^2 f(\gamma(t))(h)\big)\,(h) \\
&= D^2 f(\gamma(t))(h,h)
\end{aligned}$$

Ebenso zeigt man, dass

$$g^{(k)}(t) = D^k f(\gamma(t))(h, h, \ldots, h)$$

für $k \leq n+1$ gilt. Hieraus folgt sofort die Behauptung. $\square$

Aus Satz 7.3.1 folgt, dass für $(n+1)$-mal stetig differenzierbares f

$$R_n(x) = O(\|x - \xi\|^{n+1})$$

für $x \to \xi$ gilt. Wie in §5.4, insbesondere Satz 5.4.2, erhält man auch eine schwächere Abschätzung, falls f nur n-mal stetig differenzierbar ist.

Satz 7.3.2 *Sei $(V, \|\cdot\|_V)$ Banachraum, $U \subset V$ offen, $n \in \mathbb{N}$, $\xi \in U$, $x \in V$ mit $[\xi, x] \subset U$ und $h := x - \xi$. Die Funktion $f : U \to \mathbb{R}$ sei n-mal stetig differenzierbar. Dann gilt*

$$R_n(x) = o(\|x - \xi\|^n)$$

für $x \to \xi$.

Wir betrachten den Spezialfall $V = \mathbb{R}^p$, also $U \subset \mathbb{R}^p$ offen, $f : U \to \mathbb{R}$ differenzierbar und $\xi \in U$. Mit $h = (h_1, \ldots, h_p)^T \in \mathbb{R}^p$ ist dann

$$Df(\xi)(h) = \mathrm{grad} f(\xi) h = \sum_{j=1}^{p} \partial_j f(\xi) h_j.$$

Die Funktion $Df : U \to L(V, \mathbb{R})$ ist genau dann stetig differenzierbar, wenn dies für die Funktion $\mathrm{grad} f : U \to \mathbb{R}^p$ gilt. Letzteres ist genau dann der Fall, wenn dies für alle Koordinatenfunktionen von $\mathrm{grad} f$, also alle partiellen Ableitungen von f, gilt. Dies ist (nach Satz 7.2.3) genau dann der Fall, wenn alle zweiten partiellen Ableitungen existieren und stetig sind, d. h., wenn $f \in C^2(U, \mathbb{R})$.

Definition 7.3.1 Sei $U \subset \mathbb{R}^p$ offen, $f \in C^2(U, \mathbb{R})$ und $\xi \in U$. Dann heißt

$$H_f(\xi) := (\partial_{ij} f(\xi))_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}}$$

Hesse-Matrix von f an der Stelle ξ.

Für $f \in C^2(U, \mathbb{R})$ gilt dann

$$D^2 f(\xi)(h, h) = h^T H_f(\xi) h = \sum_{i,j=1}^{n} \partial_{ij} f(\xi) h_i h_j.$$

Man beachte, dass die Hesse-Matrix nach dem Satz von Schwarz symmetrisch ist.

7.4 Lokale Extrema

In Definition 4.2.1 haben wir für Funktionen $f : M \to \mathbb{R}$ mit $M \subset \mathbb{C}$ definiert, wann f ein lokales bzw. globales Extremum hat. Diese Begriffe können wörtlich auf den Fall, dass M metrischer Raum ist, übertragen werden.

Für $f \in C^2(I, \mathbb{R})$ mit einem Intervall I und $\xi \in \operatorname{int}(I)$ hatten wir gezeigt:

$$\begin{aligned} & f'(\xi) = 0 \text{ und } f''(\xi) > 0 \text{ (bzw. } < 0) \\ \overset{\text{Satz 4.3.3}}{\Rightarrow} \quad & f \text{ hat lokales Minimum (bzw. Maximum) in } \xi \\ \overset{\text{Satz 4.2.1}}{\Rightarrow} \quad & f'(\xi) = 0 \end{aligned}$$

Wir wollen diese Aussagen jetzt auf Funktionen $f \in C^2(U, \mathbb{R})$ mit $U \subset \mathbb{R}^p$ übertragen. Dabei ersetzen wir f' durch den Gradienten und f'' durch die Hesse-Matrix.

Satz 7.4.1 *Sei $U \subset \mathbb{R}^p$ offen und $f \in C^1(U, \mathbb{R})$ habe in $\xi \in U$ ein lokales Extremum. Dann gilt* $\operatorname{grad} f(\xi) = 0$.

Beweis. Für $j \in \{1, \ldots, p\}$ hat die durch $t \mapsto f(\xi + te_j)$ definierte Funktion g_j ein lokales Extremum in 0. Nach Satz 4.3.3 folgt $\partial_j f(\xi) = g_j'(0) = 0$. $\square$

Es sei angemerkt, dass ein entsprechender Satz auch für Funktionen gilt, die auf offenen Teilmengen von Banachräumen definiert sind. (Beweis selbst).

Beispiel. Sei $K := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4, y \geq 0\}$ und $f : K \to \mathbb{R}$, $f(x, y) = x^2 + y^2 - \sqrt{3}x - y$. Man bestimme $\max_{(x,y)\in K} f(x, y)$ und $\min_{(x,y)\in K} f(x, y)$. (Die Existenz von Maximum und Minimum folgt, da K kompakt und f stetig.)

Für ein im Inneren von K gelegenes Extremum (x, y) folgt nach Satz 7.4.1, dass $(0, 0) = \operatorname{grad} f(x, y) = (2x - \sqrt{3}, 2y - 1)$, also $(x, y) = \left(\frac{1}{2}\sqrt{3}, \frac{1}{2}\right)$.

Wir untersuchen jetzt den Rand von K auf Extrema. Mit $\Gamma_1 := \{(x, 0) : -2 \leq x \leq 2\}$ und $\Gamma_2 := \{(2\cos t, 2\sin t) : 0 \leq t \leq \pi\}$ gilt $\partial K = \Gamma_1 \cup \Gamma_2$. Desweiteren ist $\max_{(x,y)\in\Gamma_1} f(x, y) = \max_{-2\leq x\leq 2} f(x, 0) = \max_{-2\leq x\leq 2} g(x)$ mit $g(x) := x^2 - \sqrt{3}x$ und $\max_{(x,y)\in\Gamma_2} f(x, y) = \max_{0\leq t\leq\pi} f(2\cos t, 2\sin t) = \max_{0\leq t\leq\pi} h(t)$ mit $h(t) := 4 - 2\sqrt{3}\cos t - 2\sin t$. Analoges gilt für die Minima.

Wegen $g'(x) = 2x - \sqrt{3}$ hat g' nur die Nullstelle $\frac{1}{2}\sqrt{3}$ und damit sind Extremwerte von g nur in $-2, \frac{1}{2}\sqrt{3}$ und 2 möglich. Wegen $h'(t) = 2\sqrt{3}\sin t - 2\cos t$ hat h' in $(0, \pi)$ nur die Nullstelle $\frac{\pi}{6}$ und damit sind Extremwerte von h nur in $0, \frac{\pi}{6}$ und π möglich.

Insgesamt erhält man 5 Kandidaten für mögliche Extremstellen: $\xi_1 = \left(\frac{1}{2}\sqrt{3}, \frac{1}{2}\right)$, $\xi_2 = \left(\frac{1}{2}\sqrt{3}, 0\right)$, $\xi_3 = \left(2\cos\frac{\pi}{6}, 2\sin\frac{\pi}{6}\right) = (\sqrt{3}, 1)$, $\xi_4 = (-2, 0)$ und $\xi_5 = (2, 0)$.

Wegen $f(\xi_1) = -1$, $f(\xi_2) = g\left(\frac{1}{2}\sqrt{3}\right) = -\frac{3}{4}$, $f(\xi_3) = h\left(\frac{\pi}{6}\right) = 0$, $f(\xi_4) = g(-2) = h(\pi) = 4 + 2\sqrt{3}$ und $f(\xi_5) = g(2) = h(0) = 4 - 2\sqrt{3}$ folgt, dass f sein globales Maximum $4 + 2\sqrt{3}$ in $\xi_4 = (-2, 0)$ und sein globales Minimum -1 in $\xi_1 = \left(\frac{1}{2}\sqrt{3}, \frac{1}{2}\right)$ annimmt.

Wir haben hier die Extrema zur Übung mit Hilfe der Differentialrechnung bestimmt. Im konkreten Beispiel hätte man das leichter tun können, in dem man $f(x, y)$ in der Form $f(x, y) = \left(x - \frac{1}{2}\sqrt{3}\right)^2 + \left(y - \frac{1}{2}\right)^2 - 1$ schreibt.

Definition 7.4.1 Sei A reelle, symmetrische $(p \times p)$-Matrix. Dann heißt A

(i) *positiv definit*, falls $x^T Ax > 0$ für alle $x \in \mathbb{R}^p \setminus \{0\}$;

(ii) *negativ definit*, falls $x^T Ax < 0$ für alle $x \in \mathbb{R}^p \setminus \{0\}$;

(iii) *positiv semidefinit*, falls $x^T Ax \geq 0$ für alle $x \in \mathbb{R}^p$;

(iv) *negativ semidefinit*, falls $x^T Ax \leq 0$ für alle $x \in \mathbb{R}^p$;

(v) *indefinit*, falls A weder positiv noch negativ semidefinit ist, das heißt, falls $x, y \in \mathbb{R}^p$ mit $x^T Ax > 0$ und $y^T Ay < 0$ existieren.

Ist A positiv definit, so ist wegen Satz 6.5.5

$$c := \min_{\substack{x \in \mathbb{R}^p \\ \|x\|=1}} x^T Ax > 0$$

und es gilt

$$x^T Ax = \left(\frac{x}{\|x\|} A \frac{x}{\|x\|} \right) \|x\|^2 \geq c\|x\|^2$$

für alle $x \in \mathbb{R}^p$. Entsprechendes gilt im negativ definiten Fall.

Satz 7.4.2 *Sei A reelle, symmetrische $(p \times p)$-Matrix. Dann sind alle Eigenwerte von A reell und es gilt:*

(i) *A ist positiv definit $\Leftrightarrow$ Alle Eigenwerte von A sind positiv.*

(ii) *A ist negativ definit $\Leftrightarrow$ Alle Eigenwerte von A sind negativ.*

(iii) *A ist positiv semidefinit $\Leftrightarrow$ A hat keine negativen Eigenwerte.*

(iv) *A ist negativ semidefinit $\Leftrightarrow$ A hat keine positiven Eigenwerte.*

(v) *A ist indefinit $\Leftrightarrow$ A hat sowohl positive wie negative Eigenwerte.*

Satz 7.4.3 *Sei*

$$A = (a_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}}$$

reelle, symmetrische $(p \times p)$-Matrix. Für $1 \leq k \leq p$ sei A_k die $(k \times k)$-Matrix, die man aus A durch Weglassen der letzten $p - k$ Zeilen und Spalten erhält, also

$$A_k = (a_{ij})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k}} .$$

Dann gilt:

(i) *A ist positiv definit $\Leftrightarrow$ Für alle $k \in \{1, \ldots, p\}$ gilt $\det(A_k) > 0$.*

(ii) *A ist negativ definit $\Leftrightarrow$ Für alle $k \in \{1, \ldots, p\}$ gilt $(-1)^k \det(A_k) > 0$.*

Für den Beweis der Sätze 7.4.2 und 7.4.3 sei auf die Lineare Algebra verwiesen.

Satz 7.4.4 *Sei $U \subset \mathbb{R}^p$ offen, $f \in C^2(U, \mathbb{R})$, $\xi \in U$ und $\operatorname{grad} f(\xi) = 0$. Dann gilt:*

(i) *Ist $H_f(\xi)$ positiv definit, so hat f ein lokales Minimum in ξ.*

(ii) *Ist $H_f(\xi)$ negativ definit, so hat f ein lokales Maximum in ξ.*

(iii) *Ist $H_f(\xi)$ indefinit, so hat f kein lokales Extremum in ξ.*

Im semidefiniten Fall ist keine allgemeine Aussage möglich. (Dies entspricht dem Fall $f''(\xi) = 0$ im Falle einer Veränderlichen.)

Beweis von Satz 7.4.4. Sei $h \in \mathbb{R}^p$ mit $\xi + h \in U$. Aufgrund der Taylorformel gilt

$$f(\xi + h) - f(\xi) = \underbrace{Df(\xi)(h)}_{=0} + \frac{1}{2}D^2 f(\xi)(h,h) + R(h) = \frac{1}{2}h^T H_f(\xi)h + R(h)$$

mit $R(h) = o(\|h\|^2)$ für $h \to 0$.

Zu (i). Ist $H_f(\xi)$ positiv definiert, so existiert $c > 0$ mit $h^T H_f(\xi)h \geq c\|h\|^2$ für alle $h \in \mathbb{R}^p$. Weiter existiert $\delta > 0$ mit $|R(h)| < \frac{c}{2}\|h\|^2$ für $0 < \|h\| \leq \delta$. Es folgt

$$f(\xi + h) - f(\xi) \geq \frac{1}{2}h^T H_f(\xi)h - |R(h)| > 0$$

für $0 < \|h\| \leq \delta$. Also ist ξ lokales Minimum.

Zu (ii). Dies folgt analog zu (i).

Zu (iii). Ist $H_f(\xi)$ indefinit, so existieren $h_1, h_2 \in \mathbb{R}^p$ mit $c_1 := h_1^T H_f(\xi)h_1 < 0$ und $c_2 := h_2^T H_f(\xi)h_2 > 0$. Für $t \in \mathbb{R}$ und $j \in \{1,2\}$ folgt $(th_j)^T H_f(\xi)(th_j) = t^2 c_j$ und $R(th_j) = o(\|th_j\|^2\|) = o(t^2)$ für $t \to 0$. Für genügend kleines t ist dann $|R(th_j)| < \frac{1}{2}(th_j)^T H_f(\xi)(th_j)$ und damit

$$f(\xi + th_1) - f(\xi) \leq \frac{1}{2}h^T H_f(\xi)h + |R(h)| < 0$$

und

$$f(\xi + th_2) - f(\xi) \geq \frac{1}{2}h^T H_f(\xi)h - |R(h)| > 0.$$

Es folgt, dass ξ kein lokales Extremum ist. □

Für $p = 2$ sieht im Falle (iii) der Graph von f (also die Menge $\{(x, y, f(x, y)) : (x, y) \in U\} \subset \mathbb{R}^3$) lokal wie ein "Sattel" aus, im Beispiel $f(x, y) = x^2 - y^2$, $\xi = (0, 0)$ ist dies in Abbildung 28 dargestellt. Man spricht daher auch von einem *Sattelpunkt*.

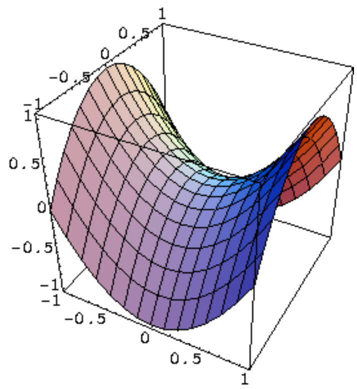

Abbildung 28: Ein Sattelpunkt.

Wir betrachten den Spezialfall $p = 2$ genauer. Für eine symmetrische (2×2)-Matrix

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

ist die Determinante $\det(A)$ durch $\det(A) = ac - b^2$ gegeben und es gilt:

(i) A ist positiv definit $\Leftrightarrow a > 0$ und $\det(A) > 0$.

(ii) A ist negativ definit $\Leftrightarrow a < 0$ und $\det(A) > 0$.

(iii) A ist indefinit $\Leftrightarrow \det(A) < 0$.

Damit erhalten wir das folgende Resultat direkt aus Satz 7.4.4.

Satz 7.4.5 *Sei $U \subset \mathbb{R}^2$ offen, $f \in C^2(U, \mathbb{R})$, $\xi \in U$ und* $\mathrm{grad} f(\xi) = 0$. *Sei $\Delta(\xi) := \partial_{11} f(\xi) \partial_{22} f(\xi) - (\partial_{12} f(\xi))^2$. Dann gilt:*

(i) *Ist $\Delta(\xi) > 0$ und $\partial_{11} f(\xi) > 0$, so hat f ein lokales Minimum in ξ.*

(ii) *Ist $\Delta(\xi) > 0$ und $\partial_{11} f(\xi) < 0$, so hat f ein lokales Maximum in ξ.*

(iii) *Ist $\Delta(\xi) < 0$, so hat f kein lokales Extremum (sondern einen Sattelpunkt) in ξ.*

Beispiel 1. Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = xy - x^3 - y^2$. Dann ist $f \in C^2(\mathbb{R}^2, \mathbb{R})$ und $\partial_1 f(x, y) = y - 3x^2$ und $\partial_2 f(x, y) = x - 2y$. Aus $\mathrm{grad} f(x, y) = 0$ folgt $y = 3x^2$ und $x = 2y$, also $y = 3(2y)^2 = 12y^2$ und damit $y = 0$ oder $y = \frac{1}{12}$. Zusammen mit $x = 2y$ folgt $(x, y) = \xi_1 := (0, 0)$ oder $(x, y) = \xi_2 := \left(\frac{1}{6}, \frac{1}{12}\right)$. Tatsächlich gilt auch $\mathrm{grad} f(\xi_j) = 0$ für $j \in \{1, 2\}$.

Wegen $\partial_{11} f(x, y) = -6x$, $\partial_{12} f(x, y) = \partial_{21} f(x, y) = 1$ und $\partial_{22} f(x, y) = -2$ folgt

$$H_f(x, y) = \begin{pmatrix} -6x & 1 \\ 1 & -2 \end{pmatrix}$$

und damit $\Delta(x, y) = \det H_f(x, y) = 12x - 1$.

Somit ist $\Delta(\xi_1) = \Delta(0, 0) = -1 < 0$ und folglich ist ξ_1 Sattelpunkt. Weiter ist $\Delta(\xi_1) = \Delta\left(\frac{1}{6}, \frac{1}{12}\right) = 1 > 0$ und wegen $\partial_{11} f(\xi_2) = -1 < 0$ hat f in ξ_2 ein lokales Maximum.

Der Graph von f – mit den Punkten ξ_1, ξ_2 markiert – ist in Abbildung 29 dargestellt.

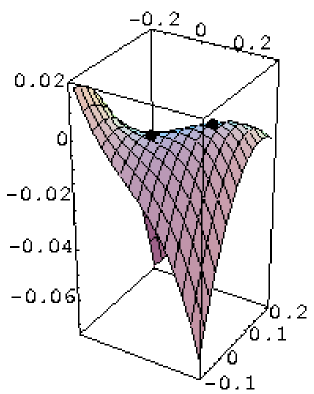

Abbildung 29: Der Graph der Funktion aus Beispiel 1.

Beispiel 2. Sei $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = 2x^4 + y^4 + z^4 - 4xyz - 2x^2$. Dann ist $f \in C^2(\mathbb{R}^3, \mathbb{R})$ und $\partial_1 f(x, y, z) = 8x^3 - 4yz - 4x$, $\partial_2 f(x, y, z) = 4y^3 - 4xz$, $\partial_3 f(x, y, z) =$

$4z^3 - 4xy$. Eine längere Rechnung zeigt, dass $\mathrm{grad} f(x,y,z) = 0$ für genau 7 Punkte (x,y,z) gilt, nämlich für $(x,y,z) = \xi_j$, $j \in \{1,\ldots,7\}$, mit $\xi_1 := (0,0,0)$, $\xi_2 := \left(\frac{1}{\sqrt{2}}, 0, 0\right)$, $\xi_3 := \left(-\frac{1}{\sqrt{2}}, 0, 0\right)$, $\xi_4 := (1,1,1)$, $\xi_5 := (1,-1,-1)$, $\xi_6 := (-1,1,-1)$ und $\xi_7 := (-1,-1,1)$. Es ist

$$H_f(x,y,z) = \begin{pmatrix} 24x^2 - 4 & -4z & -4y \\ -4z & 12y^2 & -4x \\ -4y & -4x & 12z^2 \end{pmatrix}$$

und damit

$$H_f(\xi_1) = \begin{pmatrix} -4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Diese Matrix ist negativ semidefinit, Satz 7.4.4 also nicht anwendbar. Es ist aber $f(0,0,z) > 0$ für alle $z \in \mathbb{R}$ und $f(x,0,0) < 0$ für $0 < x < 1$, und daher hat f in ξ_1 kein lokales Extremum. Weiter gilt

$$H_f(\xi_{2,3}) = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 0 & \pm 2\sqrt{2} \\ 0 & \pm 2\sqrt{2} & 0 \end{pmatrix}.$$

Das charakteristische Polynom dieser Matrix ist durch $(8-\lambda)(\lambda^2 - 8)$ gegeben. Damit hat $H_f(\xi_{2,3})$ sowohl positive wie negative Eigenwerte, ist also indefinit, und folglich ist in $\xi_{2,3}$ kein lokales Extremum. Weiter ist

$$H_f(\xi_4) = \begin{pmatrix} 20 & -4 & -4 \\ -4 & 12 & -4 \\ -4 & -4 & 12 \end{pmatrix}.$$

Diese Matrix ist positiv definit, wie man etwa durch Betrachtung der Abschnittsdeterminanten feststellt. Damit liegt in ξ_4 ein lokales Minimum vor.

Analog – oder durch Symmetrieüberlegungen – zeigt man, dass f auch in ξ_5, ξ_6, ξ_7 lokale Minima hat.

7.5 Der Satz über implizite Funktionen

Wir betrachten die Frage, wann sich eine Gleichung der Form $F(x,y) = 0$ (mit einer differenzierbaren, auf einer Teilmenge von $\mathbb{R}^2$ definierten Funktion F) nach y auflösen lässt, d. h., wann eine Funktion f existiert, so dass die Gleichung $F(x,y) = 0$ äquivalent zur Gleichung $y = f(x)$ ist. Es gilt dann $F(x, f(x)) = 0$. Wir nennen dann $F(x,y) = 0$ eine *implizite* Form der *explizit* gegebenen Gleichung $y = f(x)$.

Etwas genauer werden wir folgende Fragestellung untersuchen: Gegeben sei $(\xi, \eta) \in \mathbb{R}^2$ mit $F(\xi, \eta) = 0$. Existieren dann Umgebungen U und V von ξ und η und eine (differenzierbare) Funktion $f : U \to V$ mit $f(\xi) = \eta$ und $F(x, f(x)) = 0$ für alle $x \in U$? Inwieweit ist f eindeutig bestimmt?

Für eine konkrete Funktion F könnte $\{(x,y) \in \mathbb{R}^2 : F(x,y) = 0\}$ etwa wie in Abbildung 30 aussehen. Hier scheint es so zu sein, dass für $(\xi, \eta) = P$ Umgebungen U, V und eine Funktion f mit den verlangten Eigenschaften existieren, während das für $(\xi, \eta) = Q$ nicht der Fall ist.

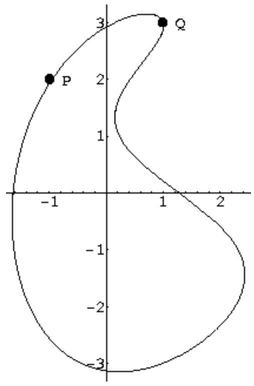

Abbildung 30: Bild von $\{(x, y) \in \mathbb{R}^2 : F(x, y) = 0\}$ für eine Funktion F.

Man erkennt nun, dass für $(\xi, \eta) = Q$ die durch $y \mapsto F(\xi, y)$ definierte Funktion in einer Umgebung von η immer ≤ 0 oder immer ≥ 0 ist, also in η ein lokales Extremum hat. Es folgt, dass $\partial_2 F(Q) = 0$ gilt.

Wir werden zeigen, dass (für stetig differenzierbares F) Umgebungen U, V und eine Funktion f mit den verlangten Eigenschaften existieren, wenn $\partial_2 F(\xi, \eta) \neq 0$ gilt.

Zum Beweis benutzt man einen dem Newtonverfahren ähnlichen Iterationsprozess: Wir suchen – für festes x – eine Nullstelle der durch $y \mapsto \alpha(y) := F(x, y)$ gegebenen Funktion. Das Newtonverfahren besteht aus der Iteration der Funktion

$$y \mapsto y - \frac{\alpha(y)}{\alpha'(y)} = y - \frac{F(x, y)}{\partial_2 F(x, y)}.$$

Für unsere Zwecke ist es einfacher, stattdessen die Funktion

$$y \mapsto y - \frac{F(x, y)}{\partial_2 F(\xi, \eta)}$$

zu iterieren. Beginnend mit einer Funktion f_0 betrachtet man also die rekursiv definierte Folge (f_n), die durch durch $f_{n+1} = A(f_n)$,

$$A(f)(x) := f(x) - \frac{F(x, f(x))}{\partial_2 F(\xi, \eta)}$$

gegeben ist.

Zur Umsetzung dieser Beweisidee benötigen wir den folgenden Satz.

Satz 7.5.1 (Banachscher Fixpunktsatz) *Sei (M, d) metrischer Raum und $f : M \to M$. Es existiere $\alpha \in [0, 1)$ mit $d(f(x), f(y)) \leq \alpha d(x, y)$ für alle $x, y \in M$. Dann hat f genau einen Fixpunkt, d. h., es existiert genau ein $\xi \in M$ mit $f(\xi) = \xi$. Darüberhinaus gilt $\lim_{n\to\infty} f^n(x) = \xi$ für alle $x \in M$. (Dabei ist f^n die n-te Iterierte von f. Diese sind rekursiv durch $f^1 := f$ und $f^{n+1} := f \circ f^n$ definiert.)*

Beweis. Übung.

Wir werden die oben skizzierten Aussagen über implizite Funktionen in allgemeinerer Form beweisen. Seien dazu $G \subset \mathbb{R}^p$ offen, $H \subset \mathbb{R}^q$ offen, $\xi \in G$, $\eta \in H$

und $F \in C^1(G \times H, \mathbb{R}^q)$ mit $F(\xi, \eta) = 0$. Wir fragen nach der Existenz von Umgebungen U und V von ξ und η und einer differenzierbaren Funktion $f : U \to V$ mit $f(\xi) = \eta$ und $F(x, f(x)) = 0$ für alle $x \in U$.

Um die oben (für $p = q = 1$) diskutierte Bedingung $\partial_2 F(\xi, \eta) \neq 0$ zu verallgemeinern, betrachten wir für $x \in G$ die Funktion $F_2^x : H \to \mathbb{R}^q$, $y \mapsto F(x, y)$ und setzen

$$D_2 F(x, y) := DF_2^x(y)$$

und

$$\partial_2^* F(x, y) := J_{F_2^x}(y) = \begin{pmatrix} \partial_{p+1} F_1(x, y) & \dots & \partial_{p+q} F_1(x, y) \\ \partial_{p+1} F_2(x, y) & \dots & \partial_{p+q} F_2(x, y) \\ \vdots & & \vdots \\ \partial_{p+1} F_q(x, y) & \dots & \partial_{p+q} F_q(x, y) \end{pmatrix}.$$

Man schreibt auch $\frac{\partial F}{\partial y}$ statt $\partial_2^* F$. Analog definiert man $D_1 F$ und $\partial_1^* F$, in dem man $F_1^y : G \to \mathbb{R}^p$, $x \mapsto F(x, y)$, betrachtet.

An Stelle der für $p = q = 1$ genannte Bedingung $\partial_2 F(\xi, \eta) \neq 0$ benötigt man jetzt, dass $D_2 F(\xi, \eta)$ invertierbar ist. Dies ist nach Linearer Algebra äquivalent zu $\det(\partial_2^* F(\xi, \eta)) \neq 0$.

Satz 7.5.2 (Satz über implizite Funktionen) *Sei $G \subset \mathbb{R}^p$ offen, $H \subset \mathbb{R}^q$ offen, $\xi \in G$, $\eta \in H$ und $F \in C^1(G \times H, \mathbb{R}^q)$ mit $F(\xi, \eta) = 0$. Es sei $D_2 F(\xi, \eta)$ invertierbar. Dann existieren offene Umgebungen $U \subset G$ und $V \subset H$ von ξ und η und eine Funktion $f \in C^1(U, V)$ mit $f(\xi) = \eta$ und $F(x, f(x)) = 0$ für $x \in U$, und umgekehrt folgt für $(x, y) \in U \times V$ mit $F(x, y) = 0$, dass $y = f(x)$ gilt.*

Desweiteren gilt $Df(x) = -D_2 F(x, f(x))^{-1} \circ D_1 F(x, f(x))$ für $x \in U$, insbesondere also $Df(\xi) = -D_2 F(\xi, \eta)^{-1} \circ D_1 F(\xi, \eta)$.

Beweis. Wir werden nur die Existenz von $f \in C(U, V)$ mit den verlangten Eigenschaften beweisen, und den Beweis der stetigen Differenzierbarkeit von f nur skizzieren. Ist aber die Differenzierbarkeit von f bekannt, so folgt die Formel für Df aus der Kettenregel.

Zur Abkürzung sei $D := D_2 F(\xi, \eta) \in L(\mathbb{R}^q, \mathbb{R}^q)$. Nach Voraussetzung existiert dann D^{-1}. Seien $\varepsilon, \delta > 0$, $U := U(\xi, \delta)$ und $V := U(\eta, \varepsilon)$. Dabei können wir ε, δ so klein wählen, dass die folgenden drei Bedingungen gelten:

(i) $\overline{U} \subset G$ und $\overline{V} \subset H$.

(ii) Für $x \in U$ und $y \in V$ gilt

$$\left\| id_{\mathbb{R}^q} - D^{-1}(D_2 F(x, y)) \right\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \leq \frac{1}{2}.$$

(iii) Für $x \in U$ gilt

$$\left\| D^{-1}(F(x, \eta)) \right\| \leq \frac{\varepsilon}{4}.$$

Dabei können (ii) und (iii) erreicht werden, weil die linken Seiten stetig in $G \times H$ bzw. G sind und für $(x, y) = (\xi, \eta)$ bzw. $x = \xi$ verschwinden. Wir setzen nun

$X := C(\overline{U}, \mathbb{R}^q) = \{g : \overline{U} \to \mathbb{R}^q : g \text{ stetig}\}$ und $\|g\|_X := \max_{x \in \overline{U}} \|g(x)\|$ für $g \in X$. Dann ist $(X, \|\cdot\|_X)$ Banachraum; vgl. Beispiel 1 zu Satz 6.3.4. Weiter setzen wir

$$M := \left\{ g \in X : g(\xi) = \eta, \forall x \in \overline{U} : \|g(x) - \eta\|_X \leq \frac{\varepsilon}{2} \right\}.$$

Dann ist $M \neq \emptyset$, da die konstante Funktion $g = \eta$ in M ist. Außerdem ist M abgeschlossen. Nach Satz 6.3.4 ist M also vollständig.

Wir betrachten jetzt die Funktion $A : M \to X$,

$$A(g)(x) := g(x) - D^{-1}(F(x, g(x))).$$

Wir werden zeigen:

(iv) Für $g, h \in M$ gilt $\|A(g) - A(h)\|_X \leq \frac{1}{2}\|g - h\|_X$.

(v) $A(M) \subset M$.

Der Banachsche Fixpunktsatz liefert nun $f \in M \subset C\left(\overline{U}, \overline{U(\eta, \frac{\varepsilon}{2})}\right)$ und damit $f \in C(U, V)$ mit $A(f) = f$. Daraus folgt $D^{-1}(F(x, f(x))) = 0$, also $F(x, f(x)) = 0$ für $x \in U$.

zu (iv): Für $x \in \overline{U}$ betrachten wir $\phi^x : V \to \mathbb{R}^q$, $y \mapsto y - D^{-1}(F(x, y))$, also $\phi^x = id_{\mathbb{R}^q}|V - D^{-1} \circ F_2^x$. Es ist dann $\phi^x(g(x)) = A(g)(x)$ für $g \in M$.

Nach Kettenregel ist ϕ^x differenzierbar mit

$$D\phi^x(y) = id_{\mathbb{R}^q} - D^{-1}\left(D_2 F(x, y)\right).$$

Nach (ii) folgt $\|D\phi^x(y)\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \leq \frac{1}{2}$ für $y \in V$. Nach Mittelwertungleichung (Satz 7.2.7) folgt

$$\|\phi^x(y_1) - \phi^x(y_2)\| \leq \sup_{y \in [y_1, y_2]} \|D\phi^x(y)\|_{L(\mathbb{R}^q, \mathbb{R}^q)} \|y_1 - y_2\|$$

für $y_1, y_2 \in V$, und damit

$$\|\phi^x(y_1) - \phi^x(y_2)\| \leq \frac{1}{2}\|y_1 - y_2\|,$$

und das gilt aus Stetigkeitsgründen sogar für für $y_1, y_2 \in \overline{V}$. Es folgt

$$\|A(g)(x) - A(h)(x)\| = \|\phi^x(g(x)) - \phi^x(h(x))\| \leq \frac{1}{2}\|g(x) - h(x)\|$$

für $g, h \in M$ und $x \in \overline{U}$ und damit (iv).

zu (v): Für $g \in M$ und $x \in \overline{U}$ gilt

$$\begin{aligned} \|A(g)(x) - \eta\| &\leq \|A(g)(x) - A(\eta)(x)\| + \|A(\eta)(x) - \eta\| \\ &\overset{\text{(iv)}}{\leq} \frac{1}{2}\|g(x) - \eta\| + \|D^{-1}(F(x, \eta))\| \\ &\overset{\text{(iii)}}{\leq} \frac{1}{2} \cdot \frac{\varepsilon}{2} + \frac{\varepsilon}{4}, \end{aligned}$$

und daraus folgt (v).

Wir haben also die Existenz einer Funktion $f \in C(U,V)$ mit $F(x,f(x)) = 0$ für $x \in U$ nachgewiesen. Seien nun $x \in U$ und $y \in V$ mit $F(x,y) = 0$. Dann ist $\phi^x(y) = y$ und damit $\|f(x) - y\| = \|\phi^x(f(x)) - \phi^x(y)\| \leq \frac{1}{2}\|f(x) - y\|$ und damit $\|f(x) - y\| = 0$, also $f(x) = y$.

Um die Differenzierbarkeit von f in ξ zu zeigen, betrachtet man die Entwicklung

$$\begin{aligned} 0 &= \underbrace{F(x,f(x))}_{=0} - \underbrace{F(\xi,\eta)}_{=0} \\ &= D_1F(\xi,\eta)(x-\xi) + D_2F(\xi,\eta)(f(x)-\eta) + \ldots, \end{aligned}$$

die wegen $f(\xi) = \eta$

$$f(x) - f(\xi) = -(D_2F(\xi,\eta))^{-1} \circ D_1F(\xi,\eta))(x-\xi) + \ldots$$

liefert. Die Differenzierbarkeit von f in ξ erhält man nun durch eine Analyse der hier nur durch ... angedeuteten Restglieder. Die Differenzierbarkeit von f in einer Umgebung von ξ erhält man analog. (Dabei muss das im ersten Teil des Beweises gewählte δ eventuell noch verkleinert werden.) $\square$

Beispiel 1. Sei $F : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, $F(x,y) = x^y - y^x$. Es gilt $F(2,4) = 0$ und $\partial_2F(x,y) = x^y \ln x - xy^{x-1}$, also $\partial_2F(2,4) = 2^4 \ln 2 - 2 \cdot 4 = 8(2\ln 2 - 1) > 0$. Nach dem Satz über implizite Funktionen existieren $\varepsilon, \delta > 0$ und stetig differenzierbares $f : (2-\delta, 2+\delta) \to (4-\varepsilon, 4+\varepsilon)$ mit $F(x,f(x)) = 0$, d. h., $x^{f(x)} = f(x)^x$.

Es ist $\partial_1F(x,y) = yx^{y-1} - y^x \ln y$, also $\partial_1F(2,4) = 4 \cdot 2^3 - 4^2 \ln 4 = 32(1 - \ln 2)$. Es folgt

$$f'(2) = -\frac{\partial_1F(2,4)}{\partial_2F(2,4)} = -4\frac{1-\ln 2}{2\ln 2 - 1} = -3,177\ldots.$$

Eine Lösung der Gleichung $F(x,y) = 0$ ist natürlich auch durch $y = x$ gegeben. Wegen $\partial_2F(x,x) = x^x(\ln x - 1) \neq 0$ für $x \neq e$ folgt aus dem Satz über implizite Funktionen, dass jeder Punkt $x \neq e$ eine Umgebung U besitzt, so dass $F(x,y) \neq 0$ für $x, y \in U$ mit $x \neq y$. Dies ist für $x = e$ nicht der Fall.

Die durch $F(x,y) = 0$ gegebene Teilmenge des $\mathbb{R}^2$ ist in Abbildung 31 dargestellt.

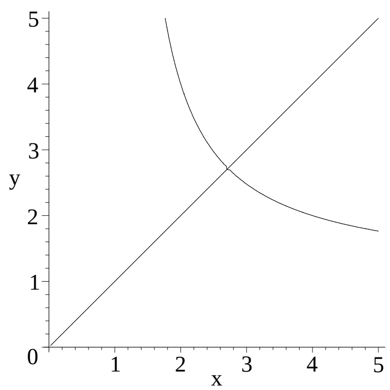

Abbildung 31: Die Menge der $(x,y) \in \mathbb{R}^2$ mit $x^y = y^x$.

Beispiel 2. Sei $G = \mathbb{R}$, $H = \mathbb{R}_+ \times \mathbb{R}_+$ und $F = (F_1, F_2)^T : G \times H \to \mathbb{R}^2$,

$$F(x, y_1, y_2) = \begin{pmatrix} F_1(x, y_1, y_2) \\ F_2(x, y_1, y_2) \end{pmatrix} = \begin{pmatrix} x \ln y_1 + y_1 e^{xy_2} - 1 \\ x e^{y_1} + \sqrt{y_1 y_2} - y_2 e^x \end{pmatrix}.$$

Es ist dann $F \in C^1(G \times H, \mathbb{R}^2)$ und $F(0, 1, 1) = 0$. Es gilt

$$\begin{aligned} \partial_2^* F(x, y_1, y_2) &= \begin{pmatrix} \partial_2 F_1(x, y_1, y_2) & \partial_3 F_1(x, y_1, y_2) \\ \partial_2 F_2(x, y_1, y_2) & \partial_3 F_2(x, y_1, y_2) \end{pmatrix} \\ &= \begin{pmatrix} \frac{x}{y_1} + e^{xy_2} & y_1 x e^{xy_2} \\ x e^{y_1} + \frac{1}{2}\sqrt{\frac{y_2}{y_1}} & \frac{1}{2}\sqrt{\frac{y_1}{y_2}} - e^x \end{pmatrix}, \end{aligned}$$

also

$$A := \partial_2^* F(0, 1, 1) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Wegen $\det A = -\frac{1}{2} \neq 0$ ist A invertierbar. Es gilt

$$A^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix}.$$

Nach dem Satz über implizite Funktionen existiert also eine Umgebung $U \subset \mathbb{R}$ von 0 und eine Umgebung $V \subset \mathbb{R}_+ \times \mathbb{R}_+$ von $(1, 1)$ sowie eine Funktion $f = (f_1, f_2)^T : U \to V$ mit $F(x, f_1(x), f_2(x)) = 0$ für alle $x \in U$. Wegen

$$\partial_1^* F(x, y_1, y_2) = \begin{pmatrix} \partial_1 F_1(x, y_1, y_2) \\ \partial_1 F_2(x, y_1, y_2) \end{pmatrix} = \begin{pmatrix} \ln y_1 + y_1 y_2 e^{xy_2} \\ e^{y_1} - y_2 e^x \end{pmatrix}$$

ist

$$\partial_1^* F(0, 1, 1) = \begin{pmatrix} 1 \\ e - 1 \end{pmatrix}$$

und damit

$$f'(0) = -\partial_2^* F(0, 1, 1)^{-1} \partial_1^* F(0, 1, 1) = -\begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ e - 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2e - 3 \end{pmatrix}.$$

Satz 7.5.3 (Umkehrsatz) *Sei $G \subset \mathbb{R}^p$ offen, $f \in C^1(G, \mathbb{R}^p)$, $\alpha \in G$ und $Df(\alpha)$ invertierbar. Dann existieren offene Umgebungen $U \subset G$ von α und $V \subset \mathbb{R}^p$ von $\beta := f(\alpha)$, so dass $f : U \to V$ bijektiv ist (d. h., $f|U$ ist injektiv und $V := f(U)$ ist eine offene Umgebung von β).*

Die Umkehrfunktion $g : V \to U$ von $f : U \to V$ ist stetig differenzierbar und es gilt $Dg(x) = Df(g(x))^{-1}$ für $x \in V$.

Beweis. Wir wenden den Satz über implizite Funktionen auf die Funktion $F : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$, $F(x, y) = f(y) - x$, an. Es ist dann $F(\beta, \alpha) = 0$ und $D_2 F(x, y) = Df(y)$, also $D_2 F(\beta, \alpha) = Df(\alpha)$.

Nach dem Satz über implizite Funktionen existieren offene Umgebungen W von α und V von β sowie eine Funktion $g \in C^1(V, W)$ mit $0 = F(x, g(x)) = f(g(x)) - x$ für $x \in V$, und $F(x, y) \neq 0$ falls $x \in V$, $y \in W$, $y \neq g(x)$.

Wegen $f(g(x)) - x = 0$ für $x \in V$ folgt, dass g injektiv ist. Mit $U := g(V)$ folgt also, dass $f : U \to V$ bijektiv mit Umkehrfunktion $g : V \to U$ ist.

Noch zu zeigen ist, dass U offen ist. Dazu zeigen wir zunächst, dass $U = f^{-1}(V) \cap W$ gilt. Sei dazu $y \in U$. Dann existiert $x \in V$ mit $y = g(x)$. Es folgt $f(y) = f(g(x)) = x \in V$, also $y \in f^{-1}(V)$, und wegen $U \subset W$ also $U \subset f^{-1}(V) \cap W$. Ist umgekehrt $y \in f^{-1}(V) \cap W$, so folgt $x := f(y) \in V$. Wegen $F(x, y) = f(y) - x = 0$ folgt $y = g(x) \in U$. Insgesamt erhalten wir $U = f^{-1}(V) \cap W$. Da f stetig und V offen ist, ist $f^{-1}(V)$ offen und damit auch $f^{-1}(V) \cap W$ offen.

Die Formel $Dg(x) = Df(g(x))^{-1}$ folgt aus dem Satz über implizite Funktionen – oder direkt aus der Kettenregel. $\square$

Der Satz ist nur lokal, d. h., auch wenn $Df(x)$ für alle $x \in G$ invertierbar ist, muss $f : G \to f(G)$ nicht bijektiv sein. Man betrachte etwa die Funktion $f : \mathbb{C} \to \mathbb{C}\backslash\{0\}$, $f(z) = e^z$. In reeller Schreibweise haben wir

$$f(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}$$

und damit

$$\det J_f(x, y) = \det \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} = e^x \neq 0,$$

womit $Df(x)$ für alle $x \in \mathbb{R}^2$ invertierbar ist. Andererseits gilt aber $f(0, 2\pi) = f(0, 0)$, womit f nicht injektiv ist.

Definition 7.5.1 Seien (M, d_M), (N, d_N) metrische Räume. Eine Abbildung $f : M \to N$ heißt *offen*, wenn $f(U)$ für jedes offene $U \subset M$ offen ist.

Man vergleiche diese Definition mit der durch Satz 6.4.3 gegebenen Charakterisierung der Stetigkeit. ("Urbilder offener Mengen sind offen.")

Eine Folgerung aus dem Umkehrsatz ist das folgende Resultat.

Satz 7.5.4 *Sei $G \subset \mathbb{R}^p$ offen, $f \in C^1(G, \mathbb{R}^p)$ und $Df(x)$ invertierbar für alle $x \in G$. Dann ist f offen.*

7.6 Extrema unter Nebenbedingungen

In §4.3.3 haben wir für $U \subset \mathbb{R}^p$ Funktionen $f : U \to \mathbb{R}$ auf lokale Extrema untersucht. Jetzt betrachten wir den Fall, dass zusätzlich noch eine Funktion $g : U \to \mathbb{R}^q$ gegeben ist, und wollen lokale Extrema von f unter der Nebenbedingung $g = 0$ bestimmen, d. h., wir untersuchen $f|\{x \in U : g(x) = 0\} = f|g^{-1}(0)$ auf lokale Extrema.

Beispiel. Man maximiere (oder minimiere) $f(x, y) = x + y$ unter der Nebenbedingung $g(x, y) = x^2 + y^2 - 1 = 0$.

Eine Methode besteht darin, die Gleichung $g(x, y) = 0$ nach y aufzulösen und das Resultat $y = h(x)$ in f einzusetzen, also das Extremum von $\tilde{f}(x) = f(x, h(x))$ zu bestimmen.

Im obigen *Beispiel* ist $h(x) = \pm\sqrt{1 - x^2}$ und $\tilde{f}(x) = x \pm \sqrt{1 - x^2}$, $x \in [-1, 1]$.

Der Nachteil dieser Methode ist, dass eine Auflösung der Gleichung $g(x, y)$ nach y (oder x) nicht immer möglich ist, oder die Auflösung oft kompliziert ist. Stattdessen werden wir jetzt den Satz über implizite Funktionen benutzen.

Satz 7.6.1 (Lagrangesche Multiplikatorenregel) *Sei $q < p$, $G \subset \mathbb{R}^p$ offen, $f \in C^1(G, \mathbb{R})$, $g \in C^1(G, \mathbb{R}^q)$, $\xi \in G$ und $g(\xi) = 0$. Die Funktion f habe in ξ ein lokales Extremum unter der Nebenbedingung $g = 0$. Hat dann $Dg(\xi)$ Rang q, so existiert $\Lambda \in L(\mathbb{R}^q, \mathbb{R})$ mit $\Lambda \circ Dg(\xi) = Df(\xi)$.*

Ist $\lambda = (\lambda_1, \ldots, \lambda_q)$ die Λ darstellende $(1 \times q)$-Matrix bezüglich der Standardbasis, so erhält die Gleichung $\Lambda \circ Dg(\xi) = Df(\xi)$ die Form

$$\mathrm{grad} f(\xi) = \lambda \cdot J_g(\xi),$$

mit $g = (g_1, \ldots, g_q)^T$ also

$$\sum_{j=1}^{q} \lambda_j \partial_k g_j(\xi) = \partial_k f(\xi).$$

Die λ_j nennt man *Lagrangesche Multiplikatoren.*

Beweis von Satz 7.6.1. Nach der Voraussetzung über den Rang von $Dg(\xi)$ existieren q linear unabhängige Spalten in $J_g(\xi)$. Ohne Beschränkung der Allgemeinheit seien die letzten q Spalten linear unabhängig. Wir setzen nun $r := p - q$. Dann ist $\mathbb{R}^p = \mathbb{R}^r \times \mathbb{R}^q$ und f und g können als Abbildungen von $\mathbb{R}^r \times \mathbb{R}^q$ nach $\mathbb{R}$ aufgefasst werden. Dementsprechend schreiben wir $x \in \mathbb{R}^p$ in der Form $x = (z, y)$ mit $z \in \mathbb{R}^r$ und $y \in \mathbb{R}^q$ und damit $f(z, y)$ bzw. $g(z, y)$ statt $f(x)$ bzw. $g(x)$. Wir definieren $D_1 f, D_2 f, D_1 g$ und $D_2 g$ wie zuvor.

Nach obiger Annahme über die Spalten von $J_g(\xi)$ ist dann $D_2 g(\xi)$ invertierbar. Setzen wir noch $\xi = (\zeta, \eta)$, so folgt mit dem Satz über implizite Funktionen, dass Umgebungen $U \subset \mathbb{R}^r$ von ζ und $V \subset \mathbb{R}^q$ von η und eine Funktion $h \in C^1(U, V)$ existieren, so dass $h(\zeta) = \eta$ und $g(z, h(z)) = 0$ für $z \in U$. Damit ist in ζ ein lokales Extremum der durch $z \mapsto f(z, h(z))$ definierten Funktion $\varphi : U \to \mathbb{R}$. Es folgt $D\varphi(\zeta) = 0$.

Nun ist (nach Kettenregel)

$$D\varphi(\zeta) = D_1 f(\zeta, \eta) + D_2 f(\zeta, \eta) \circ Dh(\zeta)$$

und (nach Satz über implizite Funktionen)

$$Dh(\zeta) = -D_2 g(\zeta, \eta)^{-1} \circ D_1 g(\zeta, \eta).$$

Es folgt

$$0 = D\varphi(\zeta) = D_1 f(\xi) - D_2 f(\xi) \circ D_2 g(\xi)^{-1} \circ D_1 g(\xi).$$

Mit $\Lambda := D_2 f(\xi) \circ D_2 g(\xi)^{-1}$ folgt also $D_1 f(\xi) = \Lambda \circ D_1 g(\xi)$. Trivialerweise ist auch $D_2 f(\xi) = \Lambda \circ D_2 g(\xi)$. Es folgt $Df(\xi) = \Lambda \circ Dg(\xi)$. $\square$

Beispiel 1. Es seien $f, g : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x + y$, $g(x, y) = x^2 + y^2 - 1 = 0$. Man bestimme das Maximum und Minimum von f unter der Nebenbedingung

$g = 0$. (Die Existenz von Maximum und Minimum folgt, da f stetig und $g^{-1}(0) = \{(x,y) \in \mathbb{R}^2 : g(x,y) = 0\}$ kompakt.

Es sind $f, g \in C^1(\mathbb{R}^2, \mathbb{R})$ mit $\operatorname{grad} f(x,y) = (1,1)$ und $\operatorname{grad} g(x,y) = (2x, 2y)$. Aus $\operatorname{grad} g(x,y) = (0,0)$ folgt $(x,y) = (0,0)$. Da aber $g(0,0) = -1 \neq 0$, ist die Voraussetzung über den Rang von Dg in allen Punkte von $g^{-1}(0)$ erfüllt.

Wir betrachten die Gleichung

$$\operatorname{grad} f(x,y) = \lambda \operatorname{grad} g(x,y),$$

also $(1,1) = \lambda(2x, 2y)$. Es folgt $\lambda \neq 0$ und damit $x = y$. Zusammen mit $g(x,y) = x^2 + y^2 - 1 = 0$ erhalten wir $(x,y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) =: \xi_1$ oder $(x,y) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) =: \xi_2$. Es folgt, dass f sein Maximum $\sqrt{2}$ in ξ_1 und sein Minimum $-\sqrt{2}$ in ξ_2 annimmt.

Beispiel 2. Man maximiere das Volumen einer Konservendose bei konstanter Oberfläche, d. h., man maximiere die durch $V(r,h) = \pi r^2 h$ gegebene Funktion $V : \mathbb{R}_+^2 \to \mathbb{R}$ unter der Nebenbedingung $g(r,h) := 2\pi r^2 + 2\pi r h - F = 0$, wobei $F \in \mathbb{R}_+$ eine gegebene Konstante ist, und $g : \mathbb{R}_+^2 \to \mathbb{R}$.

Zunächst ist wegen $2\pi r^2 < F$ für $g(r,h) = 0$ die Menge $\{r \in \mathbb{R}_+ : \exists h \in \mathbb{R}_+ \text{ mit } g(r,h) = 0\}$ beschränkt und damit gilt $V(r,h) \to 0$ für $h \to 0$ (und $g(r,h) = 0$). Außerdem gilt für $g(r,h) = 0$ auch $2\pi r h < F$ und damit $r < F/2\pi h$, also $V(r,h) < \pi h F^2/(4\pi^2 h^2) = F^2/4\pi h$. Es folgt $V(r,h) \to 0$ für $h \to \infty$ (und $g(r,h) = 0$). Zusammen mit der Stetigkeit von V folgt aus diesen Überlegungen, dass das gesuchte Maximum existiert.

Es gilt nun $V, g \in C^1(\mathbb{R}^2, \mathbb{R})$ mit $\operatorname{grad} V(r,h) = (2\pi r h, \pi r^2)$ und $\operatorname{grad} g(r,h) = (4\pi r + 2\pi h, 2\pi r)$. Aus $\operatorname{grad} g(r,h) = (0,0)$ folgt $(r,h) = (0,0)$. Da $(0,0)$ nicht im Definitionsbereich von g liegt, ist die Voraussetzung über den Rang von Dg wieder in allen Punkte von $g^{-1}(0)$ erfüllt. (Auch wenn man g und F im Definitionsbereich $\mathbb{R} \times \mathbb{R}$ betrachten würde, wäre die Voraussetzung über den Rang von Dg wegen $g(0,0) = -F \neq 0$ in allen Punkte von $g^{-1}(0)$ erfüllt.) Die Gleichung

$$\operatorname{grad} V(r,h) = \lambda \operatorname{grad} g(r,h)$$

hat jetzt die Form

$$\begin{aligned} 2\pi r h &= \lambda\,(4\pi r + 2\pi h), \\ \pi r^2 &= \lambda\, 2\pi r. \end{aligned}$$

Aus der zweiten Gleichung folgt $\lambda = \frac{r}{2}$ und Einsetzen in die erste liefert $2\pi r h = \frac{r}{2}(4\pi r + 2\pi h) = 2\pi r^2 + \pi r h$ und damit $rh = 2r^2$, also $h = 2r$.

Bemerkung. In Beispiel 2 kann man natürlich auch die Nebenbedingung $g(r,h) = 0$ nach h auflösen und erhält $h = \alpha(r) := \frac{F - 2\pi r^2}{2\pi r}$. Man hat dann

$$\beta(r) := V(r, \alpha(r)) = \pi r^2 \frac{F - 2\pi r^2}{2\pi r} = \frac{F}{2} r - \pi r^3$$

zu maximieren. Aus $\beta'(r) = \frac{F}{2} - 3\pi r^2$ folgt, dass β sein Maximum bei $r = r_0 := \sqrt{\frac{F}{6\pi}}$ annimmt. Der zugehörige Wert von h ist durch

$$h_0 := \alpha(r_0) = \frac{F - 2\pi r_0^2}{2\pi r_0} = \cdots = 2\sqrt{\frac{F}{6\pi}}$$

gegeben.

Die interessante Aussage in diesem Beispiel ist, dass das Maximum für $h = 2r$ angenommen wird. Diese erhält man mit Hilfe der Lagrangeschen Multiplikatoren direkt. Durch das Auflösen der Nebenbedingung nach einer Variablen erhält man zunächst die eher uninteressante Aussage, wie sich an der Stelle, wo das Maximum angenommen wird, r und h aus F berechnen lassen, und man kann die Bedingung $r = 2h$ dann erst durch Elimination von F daraus gewinnen.

7.7 Vertauschung von Grenzprozessen

Gegeben sei eine Funktion f von zwei Veränderlichen. Der Satz von Schwarz besagt, dass unter geeigneten Voraussetzungen die Differentiation von $f(x, y)$ nach x mit der nach y vertauscht werden kann.

Wir untersuchen die Frage, wann die Differentiation nach x mit der Integration über y vertauscht werden kann und wann Integration über x mit Integration über y vertauscht werden kann.

Satz 7.7.1 *Seien I, J kompakte Intervalle und sei $f : I \times J \to \mathbb{R}$ stetig. Dann ist auch $F : I \to \mathbb{R}$,*

$$F(x) = \int_J f(x, y)dy$$

stetig.

Existiert $\partial_1 f$ und ist $\partial_1 f$ stetig (in $I \times J$), so ist F stetig differenzierbar und es gilt

$$F'(x) = \int_J \partial_1 f(x, y)dy.$$

In den Randpunkten von $I \times J$ ist dabei $\partial_1 f$ als "einseitige" Ableitung zu verstehen. (Wir haben partielle Ableitungen nur für auf offenen Mengen definierte Funktionen eingeführt, aber die Erweiterung ist hier offensichtlich.)

Man kann die Formel in obigem Satz auch einprägsam als

$$\frac{d}{dx} \int_J f(x, y)dy = \int_J \frac{\partial}{\partial x} f(x, y)dy$$

schreiben.

Die Erweiterung des Satzes auf Funktionen von $p \geq 3$ Veränderlichen ist offensichtlich. Es folgt

$$\frac{d}{dx_j} \int_{I_k} f(x_1, \ldots, x_p)dx_k = \int_{I_k} \frac{\partial}{\partial x_j} f(x_1, \ldots, x_p)dx_k$$

für stetiges $f, \frac{\partial f}{\partial f_j} : I_1 \times \cdots \times I_p \to \mathbb{R}$ (und $k \neq j$).

Beweis von Satz 7.7.1. Sei $\xi \in I$. Da $I \times J$ kompakt ist, ist f gleichmäßig stetig und damit existiert zu $\varepsilon > 0$ ein $\delta > 0$ mit

$$|f(x, y) - f(\xi, y)| < \frac{\varepsilon}{|J|}$$

für $|x-\xi|<\delta$. Es folgt

$$|F(x)-F(\xi)| = \left|\int_I f(x,y)-f(\xi,y)dy\right| \leq \int_I |f(x,y)-f(\xi,y)|\,dy \leq \varepsilon.$$

Damit ist F stetig in ξ.

Sei nun $\partial_1 f$ stetig und damit gleichmäßig stetig. Nach Mittelwertsatz existiert dann zu $x \in I$ und $y \in J$ ein z zwischen ξ und x mit

$$\frac{f(x,y)-f(\xi,y)}{x-\xi} = \partial_1 f(z,y).$$

Es folgt, dass zu $\varepsilon > 0$ ein $\delta > 0$ existiert mit

$$\left|\frac{f(x,y)-f(\xi,y)}{x-\xi} - \partial_1 f(\xi,y)\right| < \frac{\varepsilon}{|J|}$$

für $|x-\xi|<\delta$. Durch Integration über y erhält man

$$\left|\frac{F(x)-F(\xi)}{x-\xi} - \int_J \partial_1 f(\xi,y)dy\right| \leq \varepsilon$$

für $|x-\xi|<\delta$. Dies liefert die Differenzierbarkeit von F und die angegebene Darstellung von F'. Diese Darstellung, zusammen mit dem zuerst bewiesenen, liefert wiederum die Stetigkeit von F'. $\square$

Beispiel. Sei $F : \mathbb{R} \to \mathbb{R}$,

$$F(x) = \int_1^2 \frac{e^{xy}}{y} dy.$$

Dann ist F differenzierbar und es gilt

$$F'(x) = \int_1^2 e^{xy} dy = \frac{e^{xy}}{x}\bigg|_{y=1}^{y=2} = \frac{e^{2x}-e^x}{x}.$$

Der folgende Satz – und Verallgemeinerungen davon – werden nach Fubini benannt.

Satz 7.7.2 *Seien I, J kompakte Intervalle und sei $f : I \times J \to \mathbb{R}$ stetig. Dann gilt*

$$\int_I \left(\int_J f(x,y)dy\right) dx = \int_J \left(\int_I f(x,y)dx\right) dy.$$

Beweis. Sei $I=[a,b]$. Wir betrachten $\phi : I \to \mathbb{R}$,

$$\phi(t) = \int_a^t \left(\int_J f(x,y)dy\right) dx,$$

und $\psi : I \to \mathbb{R}$,

$$\psi(t) = \int_J \left(\int_a^t f(x,y)dx\right) dy.$$

Nach dem Hauptsatz ist ϕ differenzierbar mit

$$\phi'(x) = \int_J f(x,y)dy.$$

Nach Satz 7.7.1 und dem Hauptsatz ist aber auch ψ differenzierbar mit

$$\psi'(t) = \int_J f(x,y)dy.$$

Es folgt $\phi' = \psi'$ und wegen $\phi(a) = 0 = \psi(a)$ damit $\phi(b) = \psi(b)$. $\square$

7.8 Taylorreihen

Es sei $U \subset \mathbb{R}^p$ offen und $f : U \to \mathbb{R}$. Wie in §7.3 sehen wir, dass f genau dann n-mal stetig differenzierbar ist, wenn $f \in C^n(U, \mathbb{R})$, d. h., wenn die partiellen Ableitungen der Ordnung n existieren und stetig sind. Es gilt dann

$$D^n f(\xi)(h, \dots, h) = \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \cdots \sum_{j_n=1}^{p} \partial_{j_1 j_2 \dots j_n} f(\xi) h_{j_1} h_{j_2} \dots h_{j_n}.$$

Nach dem Satz von Schwarz kommt es auf die Reihenfolge der partiellen Ableitungen nicht an. Wir wollen daher gleiche Ableitungen zusammenfassen.

Für $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathbb{N}_0^p$ und $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ setzen wir

$$\begin{aligned} |\alpha| &:= \alpha_1 + \alpha_2 + \cdots + \alpha_p, \\ \alpha! &:= \alpha_1! \cdot \alpha_2! \cdot \cdots \cdot \alpha_p!, \\ \partial^\alpha f &:= \partial_1^{\alpha_1} \partial_2^{\alpha_2} \dots \partial_p^{\alpha_p} f, \\ x^\alpha &:= x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \cdots \cdot x_p^{\alpha_p}. \end{aligned}$$

Wir nennen $\alpha = (\alpha_1, \dots, \alpha_p)$ auch *Multiindex.*

Eine kombinatorische Überlegung (die wir hier nicht weiter ausführen) liefert folgendes Ergebnis.

Hilfssatz 7.8.1 *Die Anzahl der n-Tupel $(j_1, \dots, j_n) \in \mathbb{N}_0^n$, für welche für alle $\ell \in \{1, \dots, n\}$ die Menge $\{k : j_k = \ell\}$ genau α_ℓ Elemente hat, ist durch $\frac{n!}{\alpha!}$ gegeben.*

Der Hilfssatz besagt, dass die partielle Ableitung $\partial^\alpha f$ auf $\frac{n!}{\alpha!}$ verschiedene Arten in der Form $\partial_{j_1 j_2 \dots j_n} f$ geschrieben werden kann. Wir erhalten damit

$$D^n f(\xi)(h, \dots, h) = \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \cdots \sum_{j_n=1}^{p} \partial_{j_1 j_2 \dots j_n} f(\xi) h_{j_1} h_{j_2} \dots h_{j_n} = \sum_{\substack{\alpha \in \mathbb{N}_0^p \\ |\alpha| = n}} \frac{n!}{\alpha!} \partial^\alpha f(\xi) h^\alpha.$$

Das Taylorpolynom erhält damit die Gestalt

$$T_n(x) = f(\xi) + \sum_{k=1}^{n} \sum_{\substack{\alpha \in \mathbb{N}_0^p \\ |\alpha| = k}} \frac{1}{\alpha!} \partial^\alpha f(\xi) h^\alpha.$$

Beispiel. Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = e^{x^2+y}$. Es ist dann $\partial_2 f = f$ und damit $\partial_2^k f = f$ für alle $k \in \mathbb{N}$. Es folgt, dass $\partial^{(j,k)} f = \partial_1^j \partial_2^k f = \partial_1^j f$ für alle $j, k \in \mathbb{N}_0$. Nun ist

$$\begin{aligned} \partial_1 f(x, y) &= 2x e^{x^2+y} \\ \partial_1^2 f(x, y) &= (2 + 4x^2) e^{x^2+y} \\ \partial_1^3 f(x, y) &= (12x + 8x^3) e^{x^2+y} \end{aligned}$$

Das zweite Taylorpolynom zum Entwicklungspunkt $\xi = (0, 0)$ ist damit gegeben durch

$$\begin{aligned} T_2(x, y) &= f(0,0) + \partial_1 f(0,0)x + \partial_2 f(0,0)y \\ &\quad + \frac{1}{2}\partial_1^2 f(0,0)x^2 + \partial_{12} f(0,0)xy + \frac{1}{2}\partial_2^2 f(0,0)y^2 \\ &= 1 + y + x^2 + \frac{1}{2}y^2 \end{aligned}$$

Die Graphen von f sowie der Taylorpolynome vom Grad 1 und 2 sind in Abbildung 32 dargestellt.

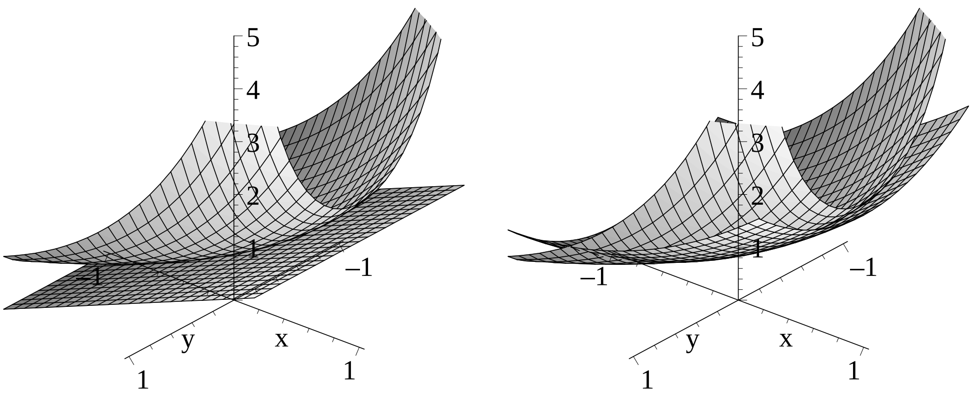

Abbildung 32: Taylorpolynome vom Grad 1 (links) und 2 (rechts).

Nach Satz 7.3.1 existiert $(\sigma, \tau) \in [(0,0),(x,y)]$ mit

$$\begin{aligned} & R_2(x,y) \\ = \; & f(x,y) - T_2(x,y) \\ = \; & \frac{1}{3!}\partial_1^3 f(\sigma,\tau)x^3 + \frac{1}{2!}\partial_1^2\partial_2 f(\sigma,\tau)x^2y + \frac{1}{2!}\partial_1\partial_2^2 f(\sigma,\tau)xy^2 + \frac{1}{3!}\partial_2^3 f(\sigma,\tau)y^3. \end{aligned}$$

Ist etwa $t := \|(x,y)\|_\infty < 1$, so folgt $\|(\sigma,\tau)\|_\infty < 1$ und damit

$$|R_2(x,y)| \le \left(\frac{1}{3}20 + \frac{1}{2}6 + \frac{1}{2}2 + \frac{1}{3}\right) e^2 t^3 = \frac{15e^2}{2} t^3.$$

Man kann nun auch Potenzreihen und Taylorreihen in mehreren Veränderlichen betrachten. Eine *Potenzreihe* in p Variablen ist dabei eine Reihe der Form

$$\sum_{\alpha \in \mathbb{N}_0^p} c_\alpha (x-\xi)^\alpha,$$

wobei $p \in \mathbb{N}$, $c_\alpha \in \mathbb{R}$ und $x, \xi \in \mathbb{R}^p$. (Auch $c_\alpha \in \mathbb{C}$ und $x, \xi \in \mathbb{C}^p$ wäre hier möglich.) Für $f \in C^\infty(U, \mathbb{R})$ und $\xi \in U$ heißt

$$\sum_{\alpha \in \mathbb{N}_0^p} \frac{1}{\alpha!} \partial^\alpha f(\xi)(x-\xi)^\alpha$$

Taylorreihe von f um ξ. Eine Taylorreihe ist also eine Potenzreihe, deren Koeffizienten c_α die Form $c_\alpha = \partial^\alpha f(\xi)(x-\xi)^\alpha$ haben. Da für $p \ge 2$ keine Summationsreihenfolge besonders ausgezeichnet ist, betrachten wir hier nur absolut konvergente Potenz- und Taylorreihen, denn diese sind ja auch unbedingt konvergent (Satz 2.7.2).

Satz 7.8.1 *Es seien $\xi, \eta \in \mathbb{R}^p$ und die Potenzreihe*

$$\sum_{\alpha \in \mathbb{N}_0^p} c_\alpha (x-\xi)^\alpha$$

konvergiere absolut für $x = \eta$. Dann konvergiert die Potenzreihe absolut und gleichmäßig in

$$\{x \in \mathbb{R}^p : \forall j \in \{1, \dots, p\} : |x_j - \xi_j| \leq |\eta_j - \xi_j|\}.$$

Beweis. Gilt $|x_j - \xi_j| \leq |\eta_j - \xi_j|$ für alle j, so ist

$$|c_\alpha (x - \xi)^\alpha| = |c_\alpha| \cdot |x_1 - \xi_1|^{\alpha_1} \cdot \dots \cdot |x_p - \xi_p|^{\alpha_p} \leq |c_\alpha| \cdot |\eta_1 - \xi_1|^{\alpha_1} \cdot \dots \cdot |\eta_p - \xi_p|^{\alpha_p},$$

und die Behauptung folgt aus dem Majorantenkriterium. □

Den folgenden Satz erhält man nun wie in einer Veränderlichen (vgl. Satz 4.4.2). Für $\alpha = (\alpha_1, \dots, \alpha_p), \beta = (\beta_1, \dots, \beta_p) \in \mathbb{N}_0^p$ schreiben wir dabei $\alpha \geq \beta$, falls $\alpha_j \geq \beta_j$ für alle j gilt.

Satz 7.8.2 *Es seien die Voraussetzungen von Satz* 7.8.1 *gegeben und es sei*

$$U := \{x \in \mathbb{R}^p : \forall j \in \{1, \dots, p\} : |x_j - \xi_j| < |\eta_j - \xi_j|\}$$

und $f : U \to \mathbb{R}$,

$$f(x) := \sum_{\alpha \in \mathbb{N}_0^p} c_\alpha (x - \xi)^\alpha.$$

Dann gilt $f \in C^\infty(U, \mathbb{R})$ und die partiellen Ableitungen dürfen gliedweise gebildet werden, d. h., für $\beta \in \mathbb{N}_0^p$ gilt

$$\partial^\beta f(x) = \sum_{\alpha \geq \beta} c_\alpha \frac{\alpha!}{(\alpha - \beta)!} (x - \xi)^{\alpha - \beta}.$$

Insbesondere gilt $\partial^\beta f(\xi) = \beta! c_\beta$.

Wie bei Funktionen einer Veränderlichen muss die Taylorreihe einer C^∞-Funktion nicht konvergieren, und wenn sie konvergiert, muss ihre Summe nicht mit dem Funktionswert übereinstimmen. Es gilt aber z. B. der folgende Satz.

Satz 7.8.3 *Sei $\xi \in \mathbb{R}^p$, $r \in \mathbb{R}_+^p$, $U := \{x \in \mathbb{R}^p : \forall j \in \{1, \dots, p\} : |x_j - \xi_j| < r_j\}$ und $f \in C^\infty(U, \mathbb{R})$. Es existiere $C \in \mathbb{R}_+$ mit*

$$|\partial^\alpha f(x)| \leq C \frac{\alpha!}{r^\alpha}$$

für alle $x \in U$ und alle $\alpha \in \mathbb{N}_0^p$. Dann konvergiert die Taylorreihe von f um ξ auf jeder kompakten Teilmenge von U absolut und gleichmäßig, und es gilt

$$f(x) = \sum_{\alpha \in \mathbb{N}_0^p} \frac{1}{\alpha!} \partial^\alpha f(\xi) (x - \xi)^\alpha.$$

für alle $x \in U$.

Beweis. Für $\rho \in \mathbb{R}_+^p$ mit $0 < \rho_j < r_j$ für alle j sei

$$Q_\rho := \{x \in \mathbb{R}^p : \forall j \in \{1, \dots, p\} : |x_j - \xi_j| < \rho_j\}.$$

Dann ist $Q_\rho \subset U$ und für $x \in Q_\rho$ und $\alpha \in \mathbb{N}_0^p$ gilt

$$\frac{1}{\alpha!}|\partial^\alpha f(\xi)(x-\xi)^\alpha| \leq C\frac{1}{r^\alpha}\rho^\alpha = C\left(\frac{\rho_1}{r_1}\right)^{\alpha_1} \cdot \dots \cdot \left(\frac{\rho_p}{r_p}\right)^{\alpha_p}.$$

Wegen

$$\sum_{\alpha \in \mathbb{N}_0^p} \left(\frac{\rho_1}{r_1}\right)^{\alpha_1} \cdot \dots \cdot \left(\frac{\rho_p}{r_p}\right)^{\alpha_p} = \frac{1}{1-\frac{\rho_1}{r_1}} \cdot \dots \cdot \frac{1}{1-\frac{\rho_p}{r_p}} < \infty$$

folgt die absolute und in Q_ρ gleichmäßige Konvergenz der Taylorreihe nach Vergleichskriterium. Da jede kompakte Teilmenge von U in einer der Mengen Q_ρ enthalten ist, folgt hieraus die erste Behauptung.

Sei nun $T(x)$ die Summe der Taylorreihe. Um zu zeigen, dass $f(x) = T(x)$ gilt, reicht es zu zeigen, dass $R_n(x) := f(x) - T_n(x) \to 0$ für $n \to \infty$ gilt. Nach Satz 7.3.1 existiert aber zu $x \in Q_\rho$ ein $\eta \in [\xi, x] \subset Q_\rho$ mit

$$\begin{aligned}
|R_n(x)| &= \left| \sum_{\substack{\alpha \in \mathbb{N}_0^p \\ |\alpha| = n+1}} \frac{1}{\alpha!}\partial^\alpha f(\eta)(x-\xi)^\alpha \right| \\
&\leq C \sum_{\substack{\alpha \in \mathbb{N}_0^p \\ |\alpha| = n+1}} \frac{1}{r^\alpha}\,|(x-\xi)^\alpha| \\
&\leq C \sum_{\substack{\alpha \in \mathbb{N}_0^p \\ |\alpha| = n+1}} \left(\frac{\rho_1}{r_1}\right)^{\alpha_1} \cdot \dots \cdot \left(\frac{\rho_p}{r_p}\right)^{\alpha_p},
\end{aligned}$$

und $R_n(x) \to 0$ folgt ähnlich wie oben. □

Man beachte in obigem Beweis, dass aus $R_n(x) \to 0$ noch nicht die Konvergenz der Reihe folgen muss.

Beispiel. Sei $U := \{(x, y) \in \mathbb{R}^2 : x < 1 \text{ und } y < 1\}$ und $f : U \to \mathbb{R}$, $f(x, y) = \frac{1}{1-x} - \frac{1}{1-y}$. Dann gilt $f \in C^\infty(U, \mathbb{R})$ mit

$$\partial^{(m,n)} f(x, y) = \begin{cases} \dfrac{m!}{(1-x)^{m+1}} & \text{falls } n = 0,\ m \geq 1, \\ -\dfrac{n!}{(1-y)^{n+1}} & \text{falls } m = 0,\ n \geq 1, \\ 0 & \text{sonst.} \end{cases}$$

Die Taylorreihe $T(x, y)$ um $\xi = (0, 0)$ ist also durch

$$T(x, y) = \sum_{m,n=0}^{\infty} c_{mn} x^m y^n = \sum_{m=0}^{\infty} x^m - \sum_{n=0}^{\infty} y^n,$$

gegeben, mit

$$c_{mn} = \begin{cases} 1 & \text{falls } n = 0,\ m \geq 1, \\ -1 & \text{falls } m = 0,\ n \geq 1, \\ 0 & \text{sonst.} \end{cases}$$

Sie konvergiert genau dann, wenn $\|(x, y)\|_\infty < 1$, also wenn $|x| < 1$ und $|y| < 1$. Weiter gilt $f(x, x) = T_n(x, x) = 0$ für alle $x \in (-\infty, 1)$. Für $x \in (-\infty, -1)$ gilt also $R_n(x, x) = 0$, obwohl die Taylorreihe divergiert.

In konkreten Fällen berechnet man – wie auch in obigem Beispiel – die Taylorreihe am einfachsten oft nicht durch sukzessives Berechnen der partiellen Ableitungen, sondern durch Rückgriff auf bekannte Potenzreihenentwicklungen in einer Veränderlichen.

Beispiel. Sei $f : \{(x, y) \in \mathbb{R}^2 : x < 1\} \to \mathbb{R}$,

$$f(x, y) = \frac{e^{xy}}{1 - x}.$$

Für $|x| < 1$ und $y \in \mathbb{R}$ ist dann

$$\begin{aligned} f(x, y) &= e^{xy} \frac{1}{1 - x} \\ &= \left(\sum_{k=0}^{\infty} \frac{1}{k!} (xy)^k \right) \left(\sum_{\ell=0}^{\infty} x^\ell \right) \\ &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{1}{k!} x^{k+\ell} y^k \\ &= \sum_{k=0}^{\infty} \sum_{j=k}^{\infty} \frac{1}{k!} x^j y^k \\ &= \sum_{(j,k) \in \mathbb{N}_0^2} c_{jk} x^j y^k \end{aligned}$$

mit

$$c_{jk} = \begin{cases} \dfrac{1}{k!} & \text{falls } j \geq k, \\ 0 & \text{sonst.} \end{cases}$$