(8.7) **Korollar.** *Sei $M \subset \mathbb{R}^n$ offen und zusammenhängend, sei $F : M \to \mathbb{R}^k$ eine differenzierbare Abbildung mit $DF_X = 0$ für alle $X \in M$. Dann ist F konstant.*

Beweis. Seien $X, Y \in M$. Da M zusammenhängend und offen ist, gibt es Punkte $X_1, \ldots, X_k$ mit $X_1 = X$, $X_k = Y$ und $[X_i, X_{i+1}] \subset M$ für $i = 1, \ldots, k-1$. (Denn für festes $X \in M$ ist die Menge A_1 aller $Y \in M$, die derart mit X durch einen Streckenzug in M verbindbar sind, offen und nicht leer. Die Menge $A_2 := M \setminus A_1$ ist ebenfalls offen. Da $A_1 \cup M_2 = M$, $A_1 \cap A_2 = \emptyset$ und M zusammenhängend ist, muß $A_2 = \emptyset$ sein, also $M = A_1$.) Für $i \in \{1, \ldots, k-1\}$ gilt nach (8.6) die Gleichung $F(X_{i+1}) - F(X_i) = 0$. Hieraus folgt $F(Y) = F(X)$. Da $X, Y \in M$ beliebig waren, folgt die Behauptung. ■

9 Partielle Ableitungen und Koordinatenfunktionen

Bei der Behandlung konkreter differenzierbarer Abbildungen wird man zweckmäßigerweise ihre Koordinatenfunktionen heranziehen. Differentiale und Kettenregel treten dann in Matrizenschreibweise auf. Diese Umformulierungen wollen wir in diesem Abschnitt vornehmen. Dazu müssen wir zunächst partielle Ableitungen reellwertiger Funktionen betrachten.

Im ersten Teil des Folgenden sei stets $M \subset \mathbb{R}^n$ eine offene Menge, $X_0 \in M$ und $f : M \to \mathbb{R}$ eine reelle Funktion.

Für reellwertige Funktionen hängt der Begriff des Differentials eng zusammen mit dem der Richtungsableitung.

(9.1) **Definition.** Sei $E \in \mathbb{R}^n \setminus \{0\}$ ein Vektor. Wenn die Funktion

$$t \mapsto f(X_0 + tE)$$

in 0 differenzierbar ist, so wird ihre Ableitung mit $f'(X_0; E)$ bezeichnet, also

$$f'(X_0; E) := \lim_{h \to 0} \frac{f(X_0 + hE) - f(X_0)}{h}.$$

$f'(X_0; E)$ heißt *Richtungsableitung* von f bezüglich E im Punkt X_0.

(9.2) **Satz.** *Ist f in X_0 differenzierbar, so existieren alle Richtungsableitungen von f in X_0, und es gilt*

$$f'(X_0; E) = Df_{X_0}(E) \qquad \text{für } E \in \mathbb{R}^n.$$

Beweis. Definieren wir $g(t) := f(X_0 + tE)$ für alle $t \in \mathbb{R}$ mit $X_0 + tE \in M$, so ist g nach der Kettenregel (8.5) in 0 differenzierbar, und es gilt $g'(0) = Df_{X_0}(E)$. ■

Behandelt man Aufgaben über differenzierbare Funktionen unter Zuhilfenahme von Koordinaten, so treten insbesondere die Richtungsableitungen in denjenigen Richtungen auf, die durch die Vektoren $E_1, \ldots, E_n$ der Standardbasis des $\mathbb{R}^n$ gegeben sind. Hierfür hat man daher besondere Bezeichnungen:

(9.3) **Definition.** Für $i \in \{1, \ldots, n\}$ schreibt man

$$f'(X_0; E_i) =: \partial_i f(X_0),$$

falls diese Ableitung existiert. $\partial_i f(X_0)$ heißt *i-te partielle Ableitung von f in X_0.*

Eine häufig verwendete Schreibweise ist auch

$$\partial_i f = \frac{\partial f}{\partial x_i}.$$

Wie partielle Ableitungen zu berechnen sind, ist klar: Für $X_0 = (x_1^0, \ldots, x_n^0)$ gilt

$$\begin{aligned} \partial_i f(X_0) &= f'(X_0; E_i) = \lim_{h \to 0} \frac{f(X_0 + hE_i) - f(X_0)}{h} \\ &= \lim_{h \to 0} \frac{f(x_1^0, \ldots, x_{i-1}^0, x_i^0 + h, x_{i+1}^0, \ldots, x_n^0) - f(x_1^0, \ldots, x_n^0)}{h}. \end{aligned}$$

Das ist die gewöhnliche Ableitung der Funktion

$$t \mapsto f(x_1^0, \ldots, x_{i-1}^0, t, x_{i+1}^0, \ldots, x_n^0)$$

an der Stelle x_i^0. Hieraus erklärt sich die Bezeichnung „partielle Ableitung“, und es ergibt sich eine einfache Berechnungsvorschrift.

Durch die partiellen Ableitungen läßt sich nun auch die Richtungsableitung allgemein und damit das Differential einer differenzierbaren Funktion leicht ausdrücken. Für $Y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ gilt wegen der Linearität des Differentials

$$\begin{aligned} f'(X_0; Y) &= Df_{X_0}(y_1 E_1 + \ldots + y_n E_n) \\ &= y_1 Df_{X_0}(E_1) + \ldots + y_n Df_{X_0}(E_n) \\ &= y_1 \partial_1 f(X_0) + \ldots + y_n \partial_n f(X_0) \\ &= \langle Y, (\partial_1 f(X_0), \ldots, \partial_n f(X_0)) \rangle. \end{aligned}$$

Daß man, wie hier, ein lineares Funktional auf dem $\mathbb{R}^n$ als Skalarprodukt mit einem festen Vektor darstellen kann, ist aus der Linearen Algebra wohlbekannt. Für den in diesem Fall auftretenden Vektor hat man eine besondere Bezeichnung.

(9.4) **Definition und Behauptung.** Existieren für $f : M \to \mathbb{R}$ (mit $M \subset \mathbb{R}^n$) die partiellen Ableitungen in X_0, so heißt der Vektor

$$\nabla f(X_0) := (\partial_1 f(X_0), \ldots, \partial_n f(X_0))$$

der *Gradient von* f *in* X_0. Ist f in X_0 differenzierbar, so gilt

$$f'(X_0; E) = Df_{X_0}(E) = \langle \nabla f(X_0), E \rangle \quad \text{für alle } E \in \mathbb{R}^n.$$

Man beachte, daß die Darstellung

$$\nabla f(X_0) = (\partial_1 f(X_0), \ldots, \partial_n f(X_0))$$

von einer speziellen Basis des $\mathbb{R}^n$ Gebrauch macht, daß der Gradient aber, wenn f in X_0 differenzierbar ist, nicht von der gewählten Basis abhängt. Der Gradient $\nabla f(X_0)$ ist ja durch die basisunabhängige Gleichung

$$\langle \nabla f(X_0), E \rangle = f'(X_0; E) \qquad \text{für } E \in \mathbb{R}^n$$

festgelegt.

Der Gradient hat eine einfache anschauliche Bedeutung: Nach der Cauchy-Schwarzschen Ungleichung (6.3) gilt für $\|E\| = 1$

$$f'(X_0; E) = \langle \nabla f(X_0), E \rangle \leq \|\nabla f(X_0)\|,$$

und das Gleichheitszeichen gilt (wie am Beweis von (6.3) abzulesen ist) genau dann, wenn

$$\nabla f(X_0) = \lambda E \qquad \text{mit einem } \lambda \geq 0$$

ist. Wir halten dies als Satz fest.

(9.5) **Satz.** *Sei* f *in* X_0 *differenzierbar und* $\nabla f(X_0) \neq 0$. *Dann nimmt die Richtungsableitung von* f *in* X_0 *bezüglich Einheitsvektoren* E *genau für*

$$E = \frac{\nabla f(X_0)}{\|\nabla f(X_0)\|}$$

ihr Maximum an; dieses Maximum ist gleich $\|\nabla f(X_0)\|$.

Mit anderen Worten: Der Gradient weist in die Richtung stärksten Anstiegs der Funktion; seine Länge gibt die Größe dieses Anstiegs an.

Für differenzierbare reellwertige Funktionen gilt der Mittelwertsatz in Form einer Gleichung; er läßt sich bequem mit Hilfe des Gradienten formulieren.

(9.6) **Satz** (Mittelwertsatz). *Seien $X, Y \in M$ Punkte mit $[X, Y] \subset M$. Sei $f : M \to \mathbb{R}$ stetig in M und differenzierbar in $(1-t)X + tY$ für $t \in (0,1)$. Dann gibt es eine Zahl $\vartheta \in (0,1)$ mit*

$$f(Y) - f(X) = \langle \nabla f((1-\vartheta)X + \vartheta Y), Y - X\rangle.$$

Beweis. Setze $g(t) := f((1-t)X + tY)$ für $0 \leq t \leq 1$. Dann ist g stetig und nach der Kettenregel differenzierbar in $(0,1)$. Nach dem Mittelwertsatz aus Analysis I existiert daher eine Zahl $\vartheta \in (0,1)$ mit

$$\begin{aligned} f(Y) - f(X) &= g(1) - g(0) = g'(\vartheta) \\ &= \langle \nabla f((1-\vartheta)X + \vartheta Y), Y - X\rangle. \end{aligned}$$

■

Wir wollen nun auf den grundsätzlichen Zusammenhang zwischen Differenzierbarkeit und partieller Differenzierbarkeit eingehen. Dabei heißt eine reellwertige Funktion *partiell differenzierbar*, wenn ihre partiellen Ableitungen existieren. Für eine partiell differenzierbare Funktion $f : M \to \mathbb{R}$ ist also für jedes $i \in \{1, \ldots, n\}$ die Funktion

$$\partial_i f : M \to \mathbb{R} : X \mapsto \partial_i f(X)$$

erklärt, die man als *i-te partielle Ableitung* von f bezeichnet. Aus Differenzierbarkeit folgt partielle Differenzierbarkeit, nach Satz (9.2) sogar die Existenz aller Richtungsableitungen. Umgekehrt folgt aus der Existenz aller Richtungsableitungen nach einem Beispiel vom Anfang dieses Kapitels nicht einmal die Stetigkeit, also gewiß nicht die Differenzierbarkeit. Andererseits ist die partielle Differenzierbarkeit oft leicht nachprüfbar, und partielle Ableitungen sind in konkreten Fällen gut zu berechnen. Es sind daher Aussagen von Nutzen, die aus partieller Differenzierbarkeit und zusätzlichen Aussagen auf Differenzierbarkeit schließen lassen.

(9.7) **Satz.** *Ist f in einer Umgebung von X_0 partiell differenzierbar und sind die partiellen Ableitungen in X_0 stetig, so ist f in X_0 differenzierbar.*

Beweis. Sei U eine offene Kugel mit Mittelpunkt $X_0 = (x_1, \ldots, x_n)$, in der f partiell differenzierbar ist. Für $H = (h_1, \ldots, h_n) \in \mathbb{R}^n$ mit $X_0 + H \in U$ gilt nach dem Mittelwertsatz der Differentialrechnung

$$\begin{aligned} & f(X_0 + H) - f(X_0) \\ = & \sum_{i=1}^{n} [f(x_1, \ldots, x_{i-1}, x_i + h_i, x_{i+1} + h_{i+1}, \ldots, x_n + h_n) \end{aligned}$$

$$\begin{aligned}
& -f(x_1,\ldots,x_{i-1},x_i,x_{i+1}+h_{i+1},\ldots,x_n+h_n)] \\
= & \sum_{i=1}^{n} h_i \partial_i f(x_1,\ldots,x_{i-1},x_i+c_ih_i,x_{i+1}+h_{i+1},\ldots,x_n+h_n)
\end{aligned}$$

mit geeigneten $c_i \in (0,1)$, also

$$\begin{aligned}
& \frac{1}{\|H\|}\left|f(X_0+H)-f(X_0)-\langle\nabla f(X_0),H\rangle\right| \\
= & \frac{1}{\|H\|}\left|\sum_{i=1}^{n} h_i[\partial_i f(x_1,\ldots,x_{i-1},x_i+c_ih_i,\ldots,x_n+h_n)-\partial_i f(X_0)]\right| \\
\leq & \sum_{i=1}^{n}\left|\partial_i f(x_1,\ldots,x_{i-1},x_i+c_ih_i,x_{i+1}+h_{i+1},\ldots,x_n+h_n)-\partial_i f(X_0)\right|.
\end{aligned}$$

Da nach Voraussetzung jede partielle Ableitung $\partial_i f$ in X_0 stetig ist, ist der Limes der rechten Seite für $\|H\| \to 0$ gleich 0. Daraus folgt die Behauptung. ■

Nach dieser Behandlung reellwertiger Funktionen kehren wir nun zu allgemeinen differenzierbaren Abbildungen in den $\mathbb{R}^k$ zurück. Wir verwenden jetzt Koordinatenfunktionen und deren partielle Ableitungen.

Gegeben seien im folgenden, wenn nichts anderes gesagt ist, eine offene Menge $M \subset \mathbb{R}^n$, ein Punkt $X_0 \in M$ und eine Abbildung $F: M \to \mathbb{R}^k$. Für $i = 1,\ldots,k$ sei $f_i: M \to \mathbb{R}$ die i-te Koordinatenfunktion von F bezüglich der Standardbasis $E'_1,\ldots,E'_k$ des $\mathbb{R}^k$, also

$$F(X) = (f_1(X),\ldots,f_k(X)) = \sum_{i=1}^{k} f_i(X)E'_i \qquad \text{für } X \in M.$$

Mit $X = (x_1,\ldots,x_n)$, $Y = (y_1,\ldots,y_k)$ lautet also die Gleichung $Y = F(X)$ in Koordinatenschreibweise

$$\begin{aligned}
y_1 &= f_1(x_1,\ldots,x_n), \\
&\vdots \\
y_k &= f_k(x_1,\ldots,x_n).
\end{aligned}$$

In der älteren Lehrbuch-Literatur findet man Abbildungen $F: M \to \mathbb{R}^k$ daher auch unter dem Stichwort „Funktionensysteme“ behandelt.

Differenzierbarkeit einer Abbildung ist, analog wie bei der Stetigkeit, gleichwertig mit Differenzierbarkeit der Koordinatenfunktionen.

(9.8) **Satz.** *Die Abbildung F ist genau dann differenzierbar in X_0, wenn alle Koordinatenfunktionen f_i in X_0 differenzierbar sind ($i = 1, \ldots, k$). Ist das der Fall, so gilt*

$$DF_{X_0}(H) = \sum_{i=1}^{k} (Df_i)_{X_0}(H) E'_i \qquad \text{für } H \in \mathbb{R}^n.$$

Beweis. Sei $L : \mathbb{R}^n \to \mathbb{R}^k$ eine lineare Abbildung. Nach (7.5) gilt

$$\begin{aligned} & \lim_{H \to 0} \frac{F(X_0 + H) - F(X_0) - L(H)}{\|H\|} = 0 \\ \Leftrightarrow \quad & \lim_{H \to 0} \frac{f_i(X_0 + H) - f_i(X_0) - (p_i \circ L)(H)}{\|H\|} = 0 \qquad \text{für } i = 1, \ldots, k. \end{aligned}$$

Daraus folgen die Behauptungen unmittelbar. ■

An der Gleichung in Satz (9.8) können wir nun sofort ablesen, durch welche Matrix das Differential von F in X_0 bezüglich der Standardbasen beschrieben wird. Sei $E_1, \ldots, E_n$ die Standardbasis von $\mathbb{R}^n$. Ist $L : \mathbb{R}^n \to \mathbb{R}^k$ eine lineare Abbildung, so ist die ihr bezüglich der Standardbasen zugeordnete Matrix $(a_{ij})_{\substack{i=1,\ldots,k \\ j=1,\ldots,n}}$ definiert durch

$$L(E_j) = \sum_{i=1}^{k} a_{ij} E'_i, \qquad j = 1, \ldots, n.$$

Setzen wir nun in der Gleichung in (9.8) speziell $H = E_j$ ein, so erhalten wir wegen $Df_{X_0}(E_j) = \partial_j f(X_0)$ die Gleichungen

$$DF_{X_0}(E_j) = \sum_{i=1}^{k} \partial_j f_i(X_0) E'_i,$$

also $a_{ij} = \partial_j f_i(X_0)$. Wir halten das fest und definieren:

(9.9) **Satz und Definition.** *Sei $F : M \to \mathbb{R}^k$ mit $M \subset \mathbb{R}^n$ differenzierbar in $X_0 \in M$. Dann wird das Differential DF_{X_0} bezüglich der Standardbasen in $\mathbb{R}^n$ und $\mathbb{R}^k$ beschrieben durch die $k \times n$-Matrix*

$$JF(X_0) := \begin{pmatrix} \partial_1 f_1(X_0) & \cdots & \partial_n f_1(X_0) \\ \vdots & & \vdots \\ \partial_1 f_k(X_0) & \cdots & \partial_n f_k(X_0) \end{pmatrix}.$$

Sie heißt die *Funktionalmatrix* oder *Jacobische Matrix* der Abbildung F an der Stelle X_0. Im Fall $k = n$ wird die Determinante $\det JF(X_0)$ der Funktionalmatrix als *Funktionaldeterminante* von F in X_0 bezeichnet.

Schreibweise. Für die Funktionaldeterminante von F (im Fall $n = k$) findet man auch oft die Bezeichnung

$$\frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}.$$

Beschreibt man lineare Abbildungen (nach Einführung von Basen in den beteiligten Räumen) durch Matrizen, so wird einer Komposition von Abbildungen als Matrix bekanntlich das Matrizenprodukt der Matrizen der einzelnen Abbildungen zugeordnet. Mit der Bezeichnung aus (9.9) lautet die Kettenregel (8.4) daher jetzt folgendermaßen: Die Funktionalmatrix einer Komposition ist gleich dem Matrizenprodukt (in der richtigen Reihenfolge) der Funktionalmatrizen der einzelnen Abbildungen.

(9.10) **Satz** (Kettenregel in Koordinatenschreibweise). *Seien $M \subset \mathbb{R}^n$ und $N \subset \mathbb{R}^k$ offen, seien $F : M \to N$ und $G : N \to \mathbb{R}^m$ Abbildungen, sei F differenzierbar in X_0 und G differenzierbar in $F(X_0)$. Dann gilt*

$$J(G \circ F)(X_0) = JG(F(X_0)) \cdot JF(X_0)$$

(*Matrizenprodukt*), *also mit $G \circ F =: H$*

$$\partial_j h_i(X_0) = \sum_{r=1}^{k} \partial_r g_i(F(X_0)) \partial_j f_r(X_0)$$

für $i = 1, \ldots, m$ und $j = 1, \ldots, n$.

Etwas übersichtlicher ist vielleicht die Schreibweise

$$\frac{\partial h_i}{\partial x_j}(X_0) = \sum_{r=1}^{k} \frac{\partial g_i}{\partial y_r}(Y_0) \frac{\partial f_r}{\partial x_j}(X_0), \qquad Y_0 = F(X_0).$$

Man kann auch gelegentlich Schreibweisen wie

$$g_i(y_1(x_1, \ldots, x_n), \ldots, y_k(x_1, \ldots, x_n)) = \tilde{g}_i(x_1, \ldots, x_n)$$

und dann die Kettenregel in der Form

$$\frac{\partial \tilde{g}_i}{\partial x_j} = \sum_{r=1}^{k} \frac{\partial g_i}{\partial y_r} \frac{\partial y_r}{\partial x_j}$$

finden. Diese ungenaue Bezeichnungsweise wollen wir aber lieber vermeiden (zumal wenn dann noch die „Erklärung“ hinzugefügt wird, auf der rechten Seite sei

y_r erst eine „unabhängige Veränderliche“ und dann eine „abhängige Veränderliche“).

Im Fall $m = n = 1$ reduziert sich die Kettenregel in Koordinatenschreibweise auf eine einzige Gleichung, die wir für

$$h(t) = g(f_1(t), \ldots, f_k(t))$$

in der Form

$$\frac{dh}{dt} = \frac{\partial g}{\partial x_1}\frac{df_1}{dt} + \ldots + \frac{\partial g}{\partial x_k}\frac{df_k}{dt}$$

schreiben können. Hat hier df_i/dt (und damit auch dh/dt) das Argument t, so muß $\partial g/\partial x_i$ das Argument $(f_1(t), \ldots, f_k(t))$ haben.

10 Höhere Ableitungen, Taylorformel, lokale Extrema

Im folgenden seien stets eine offene Menge $M \subset \mathbb{R}^n$, ein Punkt $X_0 \in M$ und eine reellwertige Funktion $f : M \to \mathbb{R}$ gegeben.

In diesem Abschnitt betrachten wir höhere partielle Ableitungen reellwertiger Funktionen. Es liegt auf der Hand, wie sie zu definieren sind. Existiert die i-te partielle Ableitung $\partial_i f$ von f in einer Umgebung von X_0 und existiert die k-te partielle Ableitung von $\partial_i f$ in X_0, so wird sie mit

$$\partial_k \partial_i f(X_0) \quad \text{oder} \quad \frac{\partial^2 f}{\partial x_k \partial x_i}(X_0)$$

bezeichnet. Entsprechend wird allgemein die partielle Ableitung

$$\partial_{i_r} \ldots \partial_{i_1} f = \frac{\partial^r f}{\partial x_{i_r} \ldots \partial x_{i_1}}$$

rekursiv definiert. Man bezeichnet sie als *partielle Ableitung r-ter Ordnung*. Existieren alle partiellen Ableitungen r-ter Ordnung in X_0, so heißt f *r-mal partiell differenzierbar* in X_0, und wenn dies für alle $X_0 \in M$ gilt, heißt f *r-mal partiell differenzierbar*. Ist f r-mal partiell differenzierbar und sind alle partiellen Ableitungen r-ter Ordnung der Funktion f stetig, so heißt f *r-mal stetig differenzierbar* (im Fall $r = 1$ kurz *stetig differenzierbar*). Die Menge der auf M r-mal stetig differenzierbaren reellen Funktionen wird mit $C^r(M)$ bezeichnet. Unter $C^0(M)$ wird die Menge der stetigen reellen Funktionen auf M verstanden. Ein Element von $C^r(M)$ heißt auch *Funktion der Klasse C^r* auf M. Es gilt also

$$C^0(M) \supset C^1(M) \supset C^2(M) \supset \ldots,$$

und zwar jeweils mit strikter Inklusion.

SCHREIBWEISE. Wird mehrmals nach derselben Veränderlichen differenziert, so verwendet man abkürzende Schreibweisen wie z.B.

$$\frac{\partial^2 f}{\partial x \partial x} =: \frac{\partial^2 f}{\partial x^2},$$

$$\underbrace{\partial_i \cdots \partial_i}_{s\text{-mal}} \underbrace{\partial_k \cdots \partial_k}_{r\text{-mal}} f = \frac{\partial^{r+s} f}{\partial x_i^s \partial x_k^r}.$$

Das folgende Beispiel zeigt, daß es ohne zusätzliche Voraussetzungen nicht gleichgültig ist, in welcher Reihenfolge die partiellen Differentiationen ausgeführt werden.

BEISPIEL. Sei $f : \mathbb{R}^2 \to \mathbb{R}$ erklärt durch

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

Die partiellen Ableitungen zweiter Ordnung existieren, und man erhält

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1, \qquad \frac{\partial^2 f}{\partial x \partial y}(0,0) = 1.$$

Derartiges kann jedoch nicht passieren, wenn die partiellen Ableitungen zweiter Ordnung noch stetig sind, wie der folgende Satz zeigt.

(10.1) **Satz** (Vertauschbarkeit der Differentiationsreihenfolge). *Ist $r \geq 2$ und $f \in C^r(M)$, so sind die partiellen Ableitungen von f bis zur r-ten Ordnung unabhängig von der Reihenfolge der Differentiationen, d.h. es gilt*

$$\partial_{i_1} \cdots \partial_{i_r} f = \partial_{i_{\sigma(1)}} \cdots \partial_{i_{\sigma(r)}} f$$

für jede Permutation σ der Zahlen $1, \ldots, r$.

Beweis. Da jede Permutation der Zahlen $1, \ldots, r$ Produkt von solchen Permutationen ist, bei denen nur zwei benachbarte Zahlen vertauscht werden, genügt es, den Beweis für den Fall $r = 2$ zu führen. Außerdem bedeutet es keine Einschränkung der Allgemeinheit, $n = 2$ anzunehmen, weil bei den in Betracht kommenden Differentiationen alle bis auf zwei Veränderliche festgehalten werden.

Sei $X = (x_1, x_2) \in M$ und $U \subset M$ eine offene Kugel mit Mittelpunkt X. Im folgenden sei $H = (h_1, h_2)$ so klein, daß $X + H \in U$ ist. Dann gilt

$$\begin{aligned}
& f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2) \\
= \; & \varphi(x_1 + h_1) - \varphi(x_1) \qquad \text{mit } \varphi(t) := f(t, x_2 + h_2) - f(t, x_2) \\
= \; & \varphi'(x_1 + c_1 h_1) h_1 \qquad \text{mit } c_1 \in (0,1) \text{ (Mittelwertsatz)} \\
= \; & [\partial_1 f(x_1 + c_1 h_1, x_2 + h_2) - \partial_1 f(x_1 + c_1 h_1, x_2)] h_1 \\
= \; & [\psi(x_2 + h_2) - \psi(x_2)] h_1 \qquad \text{mit } \psi(t) := \partial_1 f(x_1 + c_1 h_1, t) \\
= \; & \psi'(x_2 + c_2 h_2) h_1 h_2 \qquad \text{mit } c_2 \in (0,1) \text{ (Mittelwertsatz)} \\
= \; & \partial_2 \partial_1 f(x_1 + c_1 h_1, x_2 + c_2 h_2) h_1 h_2 .
\end{aligned}$$

Analog findet man

$$\begin{aligned}
& f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2) \\
= \; & \partial_1 \partial_2 f(x_1 + d_1 h_1, x_2 + d_2 h_2) h_1 h_2
\end{aligned}$$

mit passenden $d_1, d_2 \in (0, 1)$. Für $h_1 h_2 \neq 0$ folgt also

$$\partial_2 \partial_1 f(x_1 + c_1 h_1, x_2 + c_2 h_2) = \partial_1 \partial_2 f(x_1 + d_1 h_1, x_2 + d_2 h_2).$$

Man beachte, daß c_i und d_i hier von h_i abhängen ($i = 1, 2$), da der Mittelwertsatz auf dem Intervall $(x_1, x_1 + h_1)$ bzw. $(x_2, x_2 + h_2)$ angewandt wurde. Es gilt jedoch $c_i, d_i \in (0, 1)$ und daher mit $h_i \to 0$ auch $c_i h_i \to 0$ und $d_i h_i \to 0$. Wegen der vorausgesetzten Stetigkeit der partiellen Ableitungen zweiter Ordnung ergibt sich mit $(h_1, h_2) \to (0, 0)$ die Behauptung

$$\partial_2 \partial_1 f(x_1, x_2) = \partial_1 \partial_2 f(x_1, x_2). \qquad \blacksquare$$

Einer der wichtigsten Sätze aus Analysis I über mehrfach differenzierbare Funktionen war die Taylorformel. Wir wollen nun untersuchen, was sie für Funktionen von n Veränderlichen liefert, wenn man diese auf Geraden durch einen betrachteten Punkt einschränkt. Zuerst erinnern wir uns an die Taylorformel aus Analysis I; dabei können wir uns auf einen Spezialfall beschränken. Sei $g : [-\epsilon, h] \to \mathbb{R}$ mit $\epsilon, h > 0$ eine $(k + 1)$-mal differenzierbare Funktion ($k \in \mathbb{N}$). Dann gilt

$$g(h) = g(0) + \frac{1}{1!} g'(0) h + \frac{1}{2!} g''(0) h^2 + \ldots + \frac{1}{k!} g^{(k)}(0) h^k + \frac{1}{(k+1)!} g^{(k+1)}(c) h^{k+1}$$

mit einer passenden Zwischenstelle $c \in (0, h)$.

Wir betrachten nun eine Funktion $f : M \to \mathbb{R}$ mit offenem Definitionsbereich $M \subset \mathbb{R}^n$ und einen Punkt $X_0 \in M$ und machen die folgende Voraussetzung.

Voraussetzung. $k \in \mathbb{N}$, $f \in C^k(M)$, die partiellen Ableitungen k-ter Ordnung von f sind in M differenzierbar, $H \in \mathbb{R}^n$ ist ein Vektor mit $[X_0, X_0 + H] \subset M$.

Nun setzen wir

$$g(t) := f(X_0 + tH) \qquad \text{für } t \in [-\epsilon, 1]$$

wobei $\epsilon > 0$ so gewählt sei, daß $[X_0, X_0 - \epsilon H] \subset M$ ist. Da f differenzierbar ist (denn die partiellen Ableitungen sind noch differenzierbar, also stetig, somit ist f nach (9.7) differenzierbar), ist g nach der Kettenregel differenzierbar, und nach (9.10) gilt

$$g'(t) = \sum_{i=1}^{n} \partial_i f(X_0 + tH) h_i,$$

wenn $H = (h_1, \ldots, h_n)$ ist. Nach Voraussetzung sind die partiellen Ableitungen $\partial_i f$ $(i = 1, \ldots, n)$ noch differenzierbare Funktionen, daher können wir abermals die Kettenregel anwenden und erhalten

$$g''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \partial_j \partial_i f(X_0 + tH) h_i h_j.$$

So fortfahrend, erhalten wir schließlich

$$g^{(r)}(t) = \sum_{i_1, \ldots, i_r = 1}^{n} \partial_{i_1} \cdots \partial_{i_r} f(X_0 + tH) h_{i_1} \cdots h_{i_r}$$

für $r = 1, \ldots, k+1$. Auf g können wir die oben zitierte Taylorformel anwenden (mit $h = 1$). Es gibt also eine Zahl $c \in (0, 1)$ mit

$$g(1) = \sum_{r=0}^{k} \frac{1}{r!} g^{(r)}(0) + \frac{1}{(k+1)!} g^{(k+1)}(c).$$

Setzen wir hier die oben berechneten Ableitungen von g ein, so erhalten wir den folgenden Satz.

(10.2) **Satz** (Taylorformel). *Sei $M \subset \mathbb{R}^n$ offen, $X_0 \in M$, $H \in \mathbb{R}^n$ mit $[X_0, X_0 + H] \subset M$, $k \in \mathbb{N}$, $f \in C^k(M)$, und die partiellen Ableitungen k-ter Ordnung von f seien in M differenzierbar. Dann gibt es eine Zahl $c \in (0, 1)$ mit*

$$\begin{aligned} f(X_0 + H) &= f(X_0) + \sum_{r=1}^{k} \frac{1}{r!} \sum_{i_1, \ldots, i_r = 1}^{n} \partial_{i_1} \cdots \partial_{i_r} f(X_0) h_{i_1} \cdots h_{i_r} \\ &\quad + \frac{1}{(k+1)!} \sum_{i_1, \ldots, i_{k+1} = 1}^{n} \partial_{i_1} \cdots \partial_{i_{k+1}} f(X_0 + cH) h_{i_1} \cdots h_{i_{k+1}}. \end{aligned}$$