Kapitel 3: Differentiation

Wir beginnen mit einigen Vorbetrachtungen. In diesem Kapitel soll die Differentialrechnung für Funktionen von n reellen Veränderlichen behandelt werden. Unter einer reellen Funktion von n reellen Veränderlichen verstehen wir eine Abbildung $f : M \to \mathbb{R}$ mit $M \subset \mathbb{R}^n$. Für $X \in \mathbb{R}^n$, $X = (x_1, \dots, x_n)$ schreibt man üblicherweise

$$f(X) = f(x_1, \dots, x_n).$$

Man muß sich dabei stets darüber klar sein, daß diese Schreibweise die Festlegung einer Basis des $\mathbb{R}^n$ voraussetzt, auch wenn dies nicht immer explizit gesagt wird. In der Schreibweise $f(X)$ ist dagegen nicht auf eine spezielle Basis Bezug genommen; daher nennt man diese Schreibweise auch „koordinatenunabhängig“ oder „invariant“. Wenn wir in $\mathbb{R}^n$ eine andere Basis $(\tilde{E}_1, \dots, \tilde{E}_n)$ einführen, so können wir natürlich auch

$$f(X) = f\left(\sum_{i=1}^{n} \tilde{x}_i \tilde{E}_i\right) =: g(\tilde{x}_1, \dots, \tilde{x}_n)$$

schreiben, aber die Zuordnung $(\tilde{x}_1, \dots, \tilde{x}_n) \mapsto g(\tilde{x}_1, \dots, \tilde{x}_n)$ ist natürlich eine andere als die Zuordnung $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$. Aus diesen und anderen Gründen ist es zweckmäßig, so weit wie möglich koordinatenunabhängige Definitionen, Schreibweisen und Schlußweisen zu benutzen. Bei der analytischen Behandlung von Funktionen von mehreren Veränderlichen empfiehlt es sich also, in $f(X)$ immer das Argument X, also einen Punkt des Raumes, als das Wesentliche anzusehen und seine Beschreibung durch Koordinaten nur als Hilfsmittel zu betrachten.

Diese Betrachtungsweise führt uns auch dazu, an den Anfang der Differentialrechnung in höheren Dimensionen nicht die partiellen Ableitungen zu stellen. Was hierunter zu verstehen ist, ist schnell gesagt. Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine Funktion und $\overline{X} = (\overline{x}_1, \dots, \overline{x}_n) \in \mathbb{R}^n$. Als i-te partielle Ableitung von f an der Stelle $\overline{X}$ wird die Ableitung der Funktion

$$t \mapsto f(\overline{x}_1, \dots, \overline{x}_{i-1}, t, \overline{x}_{i+1}, \dots, \overline{x}_n)$$

an der Stelle $t = \overline{x}_i$ bezeichnet, falls sie existiert. Die i-te partielle Ableitung oder partielle Ableitung nach der i-ten Veränderlichen ist also die gewöhnliche Ableitung der Funktion, die sich ergibt, wenn alle Veränderlichen außer der i-ten festgehalten werden. Partielle Ableitungen beziehen sich also immer auf eine Basis. Abgesehen von diesem Schönheitsfehler ist es auch aus anderen Gründen nicht zweckmäßig, den Begriff der Differenzierbarkeit auf die Existenz der partiellen Ableitungen zu gründen. Dies soll jetzt durch Beispiele erläutert werden.

Wir betrachten Funktionen auf $\mathbb{R}^2$. In diesem Fall werden wir die Koordinaten (bezüglich der Standardbasis) eines Vektors X meist mit x, y statt x_1, x_2 bezeichnen. Sei nun $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

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

Die partiellen Ableitungen an der Stelle $(0,0)$, bezeichnet mit

$$\frac{\partial f}{\partial x}(0,0) \qquad \text{und} \qquad \frac{\partial f}{\partial y}(0,0),$$

existieren, denn die Funktionen $x \mapsto f(x,0) = 0$ und $y \mapsto f(0,y) = 0$ sind differenzierbar. Die Funktion f ist also an der Stelle (0,0) (und auch an jeder anderen Stelle) partiell differenzierbar. Sie ist aber an der Stelle (0,0) nicht stetig! In der Tat liegen in jeder Umgebung von (0,0) Punkte (a,a) mit $a \neq 0$, und es ist $f(a,a) = 1/2$, während doch $f(0,0) = 0$ ist. Nun sind wir aus Analysis I gewöhnt, daß jede differenzierbare Funktion auch stetig ist. Hier haben wir aber eine partiell differenzierbare Funktion von zwei Veränderlichen, die nicht stetig ist. Dies weist darauf hin, daß man unter der Differenzierbarkeit einer Funktion von mehreren Veränderlichen etwas anderes verstehen sollte als partielle Differenzierbarkeit.

Nun bezieht sich partielle Differenzierbarkeit auf eine spezielle Basis. Man könnte daher von einer Funktion schärfer fordern, daß sie bezüglich jeder Basis partiell differenzierbar ist. Aber auch daraus würde nicht die Stetigkeit folgen. Dies wird belegt durch das Beispiel

$$f(x,y) := \begin{cases} \dfrac{x^2 y}{x^4+y^2} & \text{für } (x,y) \neq (0,0), \\ 0 & \text{für } (x,y) = (0,0). \end{cases}$$

In jeder Umgebung von $(0,0)$ liegen Punkte (a,a^2) mit $a \neq 0$, und hierfür ist $f(a,a^2) = \frac{1}{2}$. Also ist f nicht stetig in (0,0). Andererseits sind die Funktionen

$$y \mapsto f(0,y) = 0$$

und

$$t \mapsto f(t,\alpha t) = \frac{\alpha t}{t^2+\alpha^2}, \qquad \alpha \in \mathbb{R}$$

überall differenzierbar. Mit anderen Worten: Für jede Gerade G in $\mathbb{R}^2$ durch den Nullpunkt ist die Einschränkung von f auf G differenzierbar (genauer: für jeden Vektor $E \in \mathbb{R}^2$ ist die Funktion $t \mapsto f(tE)$ differenzierbar). Später werden wir hierfür sagen, daß alle Richtungsableitungen dieser Funktion existieren, insbesondere alle partiellen Ableitungen bei beliebiger Basis. Trotzdem ist die Funktion nicht stetig.

8 Differenzierbarkeit

Die Vorbetrachtungen machen klar, daß wir uns gut überlegen müssen, was eigentlich eine sinnvolle Definition der Differenzierbarkeit einer Funktion von mehreren Veränderlichen ist. Wir können uns hier durch den eindimensionalen Fall leiten lassen, müssen ihn aber inhaltlich richtig verstehen und dazu etwas uminterpretieren.

Der Grundgedanke der Differentialrechnung besteht darin, eine gegebene Abbildung von einem euklidischen Raum in einen anderen dadurch zu untersuchen, daß man die Abbildung in der Nähe eines betrachteten Punktes möglichst gut approximiert durch „möglichst einfache“ Abbildungen. Besonders einfach sind konstante Abbildungen und, nach diesem Trivialfall, lineare Abbildungen. Zusammensetzungen aus beiden nennt man affine Abbildungen. Eine Abbildung $F : \mathbb{R}^n \to \mathbb{R}^k$ heißt also *affin*, wenn

$$F(X) = L(X) + Z \qquad \text{für } X \in \mathbb{R}^n$$

ist, wobei $L : \mathbb{R}^n \to \mathbb{R}^k$ eine lineare Abbildung und $Z \in \mathbb{R}^k$ ein fester Vektor ist.

Betrachten wir nun zunächst den Fall einer Funktion $f : \mathbb{R} \to \mathbb{R}$. Eine affine Abbildung $g : \mathbb{R} \to \mathbb{R}$ ist von der Form

$$g(x) = cx + t \qquad \text{für } x \in \mathbb{R}$$

mit Konstanten c und t. Wir wollen nun zu gegebenem $x_0 \in \mathbb{R}$ eine affine Funktion g finden, die f bei x_0 „möglichst gut“ approximiert. Die erste Forderung an g ist natürlich, daß $g(x_0) = f(x_0)$ sein soll; dies wird erfüllt durch

$$g(x) = c(x - x_0) + f(x_0).$$

Schreiben wir wie üblich $x = x_0 + h$, so verschwindet also die Funktion

$$h \mapsto f(x_0 + h) - g(x_0 + h) = f(x_0 + h) - f(x_0) - ch$$

für $h = 0$. „Gute Approximation“ interpretieren wir nun so, daß die Differenz $f(x_0 + h) - f(x_0) - ch$ für $h \to 0$ schneller klein werden soll als h. Damit ist gemeint, daß sogar noch

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0) - ch}{h} = 0 \tag{$*$}$$

sein soll. Hiermit sind wir genau bei der Definition der Differenzierbarkeit aus Analysis I angelangt: $(*)$ ist äquivalent damit, daß f bei x_0 differenzierbar und daß $f'(x_0) = c$ ist. Die obigen Überlegungen legen es aber jetzt nahe, nicht die Zahl c, sondern die durch sie definierte lineare Abbildung $L : h \mapsto ch$ in den

Mittelpunkt zu stellen, also zu sagen: Die Funktion f heißt differenzierbar in x_0, wenn es eine lineare Funktion $L : \mathbb{R} \to \mathbb{R}$ gibt mit

$$\lim_{h\to 0} \frac{f(x_0+h) - f(x_0) - L(h)}{h} = 0.$$

In dieser Form läßt sich die Definition sofort auf Abbildungen zwischen höherdimensionalen Räumen übertragen.

(8.1) **Definition und Behauptung.** Sei $M \subset \mathbb{R}^n$ offen, sei $F : M \to \mathbb{R}^k$ eine Abbildung, sei $X_0 \in M$. Die Abbildung F heißt *differenzierbar in* X_0, wenn es eine lineare Abbildung $L : \mathbb{R}^n \to \mathbb{R}^k$ gibt mit

$$\lim_{H\to 0} \frac{F(X_0+H) - F(X_0) - L(H)}{\|H\|} = 0.$$

Gibt es eine solche lineare Abbildung, so ist sie eindeutig bestimmt; sie heißt dann das *Differential von* F *in* X_0 und wird mit DF_{X_0} bezeichnet. F heißt *differenzierbar*, wenn F differenzierbar in X ist für alle $X \in M$.

Beweis der Behauptung. Zu zeigen ist, daß es höchstens eine lineare Abbildung mit der genannten Eigenschaft gibt. Seien also $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}^k$ zwei lineare Abbildungen dieser Art. Dann gilt

$$\lim_{H\to 0} \frac{L_1(H) - L_2(H)}{\|H\|} = 0.$$

Für jeden Vektor $X \in \mathbb{R}^n \setminus \{0\}$ folgt insbesondere

$$0 = \lim_{t\searrow 0} \frac{L_1(tX) - L_2(tX)}{\|tX\|} = \frac{L_1(X) - L_2(X)}{\|X\|},$$

also $L_1(X) = L_2(X)$; somit ist $L_1 = L_2$. ■

Hinweis. In der älteren Literatur wird Differenzierbarkeit im Sinne von Definition (8.1) oft als „vollständige“ oder „totale“ Differenzierbarkeit bezeichnet.

In Definition (8.1) dürfen n und k beliebige natürliche Zahlen sein. Ist eine dieser Zahlen gleich 1, so kann man sich die Bedeutung des Differentials gut veranschaulichen.

Sei zunächst $k = 1$ und, der Einfachheit halber, $n = 2$. In diesem Fall empfiehlt es sich, von der Funktion $f : M \to \mathbb{R}$ (mit $M \subset \mathbb{R}^2$) den Graphen

$$\text{Graph } f := \{(x,y,z) \in \mathbb{R}^3 \mid (x,y) \in M,\ z = f(x,y)\}$$

zu betrachten. Sei $X_0 \in M$ und Df_{X_0} das Differential von f in X_0. Dann ist also $Df_{X_0} : \mathbb{R}^2 \to \mathbb{R}$ eine lineare Funktion. Differenzierbarkeit von f in X_0 kann auch so formuliert werden, daß

$$f(X) = f(X_0) + Df_{X_0}(X - X_0) + R(X_0, X)$$

mit

$$\lim_{X \to X_0} \frac{R(X_0, X)}{\|X - X_0\|} = 0 \tag{$*$}$$

gilt. Die durch

$$g(X) := f(X_0) + Df_{X_0}(X - X_0), \quad X \in \mathbb{R}^2,$$

definierte Funktion g ist affin, und ihr Graph

$$\text{Graph } g = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \mathbb{R}^2,\ z = g(x, y)\}$$

ist eine Ebene durch den Punkt $(X_0, f(X_0))$. Die Beziehung $(*)$, also

$$\lim_{X \to X_0} \frac{f(X) - g(X)}{\|X - X_0\|} = 0$$

besagt, daß der Graph von f bei Annäherung von X an X^0 sich dem Graphen von g besonders gut anschmiegt. Der Graph von g wird daher auch als *Tangentialebene* des Graphen von f im Punkt $(X_0, f(X_0))$ bezeichnet. Differenzierbarkeit einer reellen Funktion kann also auch als Existenz einer Tangentialebene an den Graphen interpretiert werden. Das Differential beschreibt die Stellung dieser Tangentialebene.

Betrachten wir jetzt den Fall $n = 1$ und $k \geq 2$. Die Menge $M \subset \mathbb{R}$ wollen wir etwa als Intervall annehmen. Die Abbildung $F : M \to \mathbb{R}^k$ bezeichnet man als eine *parametrisierte Kurve*. (Eine physikalische Interpretation im Fall $k = 3$ wäre etwa, daß F die Bahn eines Massepunktes in Abhängigkeit von der Zeit beschreibt.) Sei $t_0 \in M$ und F differenzierbar in t_0. Das Differential $DF_{t_0} : \mathbb{R} \to \mathbb{R}^k$ ist eine lineare Abbildung von $\mathbb{R}$ in $\mathbb{R}^k$, also gilt $DF_{t_0}(h) = hDF_{t_0}(1)$. Wir setzen

$$DF_{t_0}(1) =: F'(t_0);$$

das ist also ein Vektor in $\mathbb{R}^k$. Die Limesgleichung in der Definition (8.1) der Differenzierbarkeit lautet dann

$$\lim_{h \to 0} \frac{F(t_0 + h) - F(t_0) - F'(t_0)h}{|h|} = 0,$$

was äquivalent ist mit

$$\lim_{h \to 0} \frac{F(t_0 + h) - F(t_0)}{h} = F'(t_0).$$

Damit ist eine anschauliche Deutung des Vektors $F'(t_0)$ gefunden. Wir bezeichnen ihn als *Ableitung* der Abbildung F oder als *Tangentenvektor* der parametrisierten Kurve F in t_0. Ist f_i die i-te Koordinatenfunktion von F, also

$$F(t) = (f_1(t), \ldots, f_k(t)) \qquad \text{für } t \in M,$$

so ist wegen (7.5)

$$F'(t_0) = (f_1'(t_0), \ldots, f_k'(t_0)).$$

Man erhält den Ableitungs- oder Tangentenvektor also durch koordinatenweises Differenzieren.

Wir kehren zum allgemeinen Fall zurück und wollen zunächst einige grundlegende Aussagen über differenzierbare Abbildungen zusammenstellen. Danach untersuchen wir insbesondere reellwertige Funktionen, die wir anschliessend in Gestalt von Koordinatenfunktionen wieder für die Behandlung allgemeiner differenzierbarer Abbildungen nutzbar machen.

Die folgenden Sätze zeigen, daß der durch (8.1) eingeführte Differenzierbarkeitsbegriff, anders als die Forderung der partiellen Differenzierbarkeit, die aus Analysis I bekannten Implikationen von Differenzierbarkeit auszudehnen gestattet.

(8.2) **Satz.** *Ist $F : M \to \mathbb{R}^k$ in X_0 differenzierbar, so ist F in X_0 stetig.*

Beweis. Sei F in X_0 differenzierbar. Zu gegebenem $\epsilon \in \mathbb{R}^+$ existiert dann ein $\delta_1 \in \mathbb{R}^+$ mit

$$\frac{\|F(X_0 + H) - F(X_0) - DF_{X_0}(H)\|}{\|H\|} < \frac{\epsilon}{2}$$

für alle H mit $X_0 + H \in M$ und $0 < \|H\| < \delta_1$. Da lineare Abbildungen gleichmäßig stetig sind, existiert ein $\delta_2 \in \mathbb{R}^+$ mit $\|DF_{X_0}(H)\| < \epsilon/2$ für alle $H \in \mathbb{R}^n$ mit $\|H\| < \delta_2$. Für alle $X_0 + H \in M$ mit $\|H\| < \delta := \min\{\delta_1, \delta_2, 1\}$ gilt dann

$$\begin{aligned}
& \|F(X_0 + H) - F(X_0)\| \\
\leq\ & \|F(X_0 + H) - F(X_0) - DF_{X_0}(H)\| + \|DF_{X_0}(H)\| \\
\leq\ & \frac{\epsilon}{2}\|H\| + \frac{\epsilon}{2} \leq \epsilon.
\end{aligned}$$

■

(8.3) **Satz.** *Sind die Abbildungen $F : M \to \mathbb{R}^k$ und $G : M \to \mathbb{R}^k$ differenzierbar in X_0, so auch $F + G$, und es gilt*

$$D(F + G)_{X_0} = DF_{X_0} + DG_{X_0}.$$

Beweis. Die Behauptung ergibt sich aus

$$\begin{aligned} & (F+G)(X_0+H)-(F+G)(X_0)-(DF_{X_0}+DG_{X_0})(H) \\ = \; & [F(X_0+H)-F(X_0)-DF_{X_0}(H)]+[G(X_0+H)-G(X_0)-DG_{X_0}(H)] \end{aligned}$$

nach Division durch $\|H\|$ und Limesbildung $H \to 0$. ■

Die wichtige Kettenregel besagt, daß die Komposition differenzierbarer Abbildungen differenzierbar ist und daß das Differential der Komposition die Komposition der Differentiale ist.

(8.4) **Satz** (Kettenregel). *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 $X_0 \in M$, sei F differenzierbar in X_0 und G differenzierbar in $F(X_0)$. Dann ist $G \circ F$ differenzierbar in X_0, und es gilt*

$$D(G \circ F)_{X_0} = DG_{F(X_0)} \circ DF_{X_0}.$$

Beweis. Wir setzen $F(X_0) =: Y_0$ und $F(X_0+H)-F(X_0) = Z_H$ für $H \in \mathbb{R}^n \setminus \{0\}$ und $X_0+H \in M$. Für diese H gilt

$$\begin{aligned} & \frac{1}{\|H\|}\left[(G\circ F)(X_0+H)-(G\circ F)(X_0)-DG_{F(X_0)}\circ DF_{X_0}(H)\right] \\ = \; & \frac{1}{\|H\|}[G(Y_0+Z_H)-G(Y_0)-DG_{Y_0}(Z_H)] \\ + \; & \frac{1}{\|H\|}[DG_{Y_0}(F(X_0+H)-F(X_0))-DG_{Y_0}(DF_{X_0}(H))] \\ = \; & \frac{1}{\|H\|}[G(Y_0+Z_H)-G(Y_0)-DG_{Y_0}(Z_H)] \\ + \; & DG_{Y_0}\left(\frac{1}{\|H\|}(F(X_0+H)-F(X_0)-DF_{X_0}(H))\right), \end{aligned}$$

wobei zuletzt benutzt wurde, daß das Differential eine lineare Abbildung ist. Wegen der Differenzierbarkeit von F in X_0 und der Stetigkeit der linearen Abbildung DG_{Y_0} gilt

$$\lim_{H\to 0} DG_{Y_0}\left(\frac{1}{\|H\|}(F(X_0+H)-F(X_0)-DF_{X_0}(H))\right) = 0.$$

Für den ersten Summanden haben wir

$$\begin{aligned} & \frac{1}{\|H\|}[G(Y_0+Z_H)-G(Y_0)-DG_{Y_0}(Z_H)] \\ = \; & \begin{cases} 0, & \text{falls } Z_H = 0 \\ \dfrac{\|Z_H\|}{\|H\|}\dfrac{1}{\|Z_H\|}[G(Y_0+Z_H)-G(Y_0)-DG_{Y_0}(Z_H)] & \text{sonst.} \end{cases} \end{aligned}$$

Aus der Differenzierbarkeit von G in Y_0, der Stetigkeit von F in X_0 und der Beschränktheit von $\|Z_H\|/\|H\|$ in einer Umgebung von 0 (die aus der Differenzierbarkeit von F in X_0 folgt) ergibt sich jetzt

$$\lim_{H\to 0} \frac{1}{\|H\|} \left[G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)\right] = 0.$$

Daraus folgt die Behauptung. ■

Bei konkreten Berechnungen wird man im allgemeinen die Kettenregel in Koordinatenschreibweise benutzen; hierauf kommen wir später zurück.

Den Fall $n = 1$ der Kettenregel (8.4), der häufig vorkommen wird, schreiben wir noch mit Verwendung des Ableitungsvektors.

(8.5) **Korollar.** *Seien $I \subset \mathbb{R}$ und $N \subset \mathbb{R}^k$ offen, seien $F : I \to N$ und $G : N \to \mathbb{R}^m$ Abbildungen. Ist F differenzierbar in $t_0 \in I$ und G differenzierbar in $F(t_0)$, so gilt*

$$(G \circ F)'(t_0) = DG_{F(t_0)}(F'(t_0)).$$

Beweis. Nach (8.4) ist

$$D(G \circ F)_{t_0} = DG_{F(t_0)} \circ DF_{t_0};$$

für $h \in \mathbb{R}$ folgt also

$$\begin{aligned} h(G \circ F)'(t_0) &= D(G \circ F)_{t_0}(h) = DG_{F(t_0)}(DF_{t_0}(h)) \\ &= DG_{F(t_0)}(hF'(t_0)) = hDG_{F(t_0)}(F'(t_0)). \end{aligned}$$

■

Hiermit läßt sich auch die Bedeutung des Differentials noch etwas einsichtiger machen. Mit den Bezeichnungen aus (8.5) ist $F : I \to N$ (wobei I ein Intervall sei) eine parametrisierte Kurve durch $F(t_0) =: X_0$, und $F'(t_0)$ ist ihr Tangentenvektor an dieser Stelle. Umgekehrt ist jeder Vektor $T \in \mathbb{R}^k$ Tangentenvektor einer passenden parametrisierten Kurve durch X_0; zum Beispiel kann man $F(t) := X_0 + tT$ für $t \in \mathbb{R}$ setzen. Nun ist die Komposition $G \circ F$ eine parametrisierte Kurve in $\mathbb{R}^m$ durch $F(X_0)$. Nach (8.5) gilt

$$(G \circ F)'(t_0) = DG_{X_0}(F'(t_0)).$$

Man erhält also den Tangentenvektor der Bildkurve $G \circ F$, indem man auf den Tangentenvektor der ursprünglichen parametrisierten Kurve F das Differential der Abbildung G an der Stelle X_0 anwendet.

Einer der nützlichsten Sätze aus Analysis I über differenzierbare Funktionen ist der Mittelwertsatz. Er besagt, daß man für eine differenzierbare Funktion $f : [x, y] \to \mathbb{R}$ die Differenz $f(y) - f(x)$ darstellen kann durch

$$f(y) - f(x) = f'(z)(y - x),$$

wobei $z \in (x, y)$ eine passende Zwischenstelle ist. Ausgenutzt wird dies meist in folgender Weise. Weiß man, daß für alle $z \in (x, y)$ die Abschätzung $|f'(z)| \leq c$ mit einer Konstanten c gilt, so folgt

$$|f(y) - f(x)| \leq c|y - x|.$$

Mit anderen Worten, die Differenz der Funktionswerte bei x und y ist klein, wenn im ganzen Intervall der Betrag der Ableitung klein ist. In dieser Form läßt sich der Mittelwertsatz auf differenzierbare Abbildungen F von $\mathbb{R}^n$ in den $\mathbb{R}^k$ übertragen, nicht jedoch in der vorherigen Form einer Gleichung. Qualitativ besagt diese Verallgemeinerung des Mittelwertsatzes, daß der Abstand $\|F(Y) - F(X)\|$ nicht groß ist, wenn das Differential von F längs der Verbindungsstrecke von X und Y nicht groß ist. Die Größe des Differentials wird dabei im Sinne der in (7.1) eingeführten Norm linearer Abbildungen gemessen. Mit

$$[X, Y] := \{(1 - \lambda)X + \lambda Y \mid 0 \leq \lambda \leq 1\}$$

bezeichnen wir die Verbindungsstrecke der Punkte $X, Y \in \mathbb{R}^n$.

(8.6) **Satz.** *Sei $M \subset \mathbb{R}^n$ offen, seien $X, Y \in M$ Punkte mit $[X, Y] \subset M$. Die Abbildung $F : M \to \mathbb{R}^k$ sei stetig in M und differenzierbar in $(1 - \lambda)X + \lambda Y$ für alle $\lambda \in (0, 1)$. Gilt*

$$\|DF_Z\| \leq c \quad \text{für } Z = (1 - \lambda)X + \lambda Y \text{ mit } \lambda \in (0, 1),$$

mit einer Konstanten c, so gilt

$$\|F(Y) - F(X)\| \leq c\|Y - X\|.$$

Beweis. Zunächst behandeln wir einen Spezialfall. Sei $G : [0, 1] \to \mathbb{R}^k$ eine stetige Abbildung derart, daß die Einschränkung $G|(0, 1)$ differenzierbar ist und daß

$$\|G'(t)\| \leq c \qquad \text{für alle } t \in (0, 1)$$

gilt. Wähle $\epsilon \in \mathbb{R}^+$ und setze

$$A := \{t \in [0, 1] \mid \|G(t) - G(0)\| \leq (c + \epsilon)t + \epsilon\}.$$

Da G in 0 stetig ist, enthält A jedenfalls ein Intervall $[0, \tau]$ mit $\tau > 0$. Setze $s := \sup A$; dann ist also $0 < s \leq 1$. Da G in s stetig ist, gilt

$$\|G(s) - G(0)\| \leq (c + \epsilon)s + \epsilon,$$

also $s \in A$. Angenommen, es wäre $s < 1$. Da G in s differenzierbar ist, gibt es ein $h > 0$ mit $s + h \leq 1$ und

$$\left\| \frac{G(s+h) - G(s)}{h} - G'(s) \right\| \leq \epsilon,$$

also mit

$$\left\| \frac{G(s+h) - G(s)}{h} \right\| \leq \|G'(s)\| + \epsilon \leq c + \epsilon.$$

Es folgt

$$\begin{aligned} \|G(s+h) - G(0)\| &\leq \|G(s+h) - G(s)\| + \|G(s) - G(0)\| \\ &\leq (c+\epsilon)h + (c+\epsilon)s + \epsilon \\ &= (c+\epsilon)(s+h) + \epsilon, \end{aligned}$$

also $s + h \in A$, im Widerspruch zur Definition von s. Damit ist $s = 1$ bewiesen; es gilt also

$$\|G(1) - G(0)\| \leq c + 2\epsilon.$$

Da $\epsilon \in \mathbb{R}^+$ beliebig war, folgt

$$\|G(1) - G(0)\| \leq c.$$

Nun sei F wie in (8.6) vorausgesetzt. Wir definieren $K : [0,1] \to \mathbb{R}^n$ durch

$$K(t) := (1-t)X + tY \qquad \text{für } t \in [0,1].$$

Trivialerweise ist K differenzierbar und $K'(t) = Y - X$. Nach der Kettenregel (8.4) ist $F \circ K$ differenzierbar in $t \in (0,1)$, und es gilt

$$D(F \circ K)_t = DF_{K(t)} \circ DK_t,$$

also

$$(F \circ K)'(t) = DF_{K(t)}(K'(t)) = DF_{K(t)}(Y - X).$$

Nach (7.1) folgt

$$\|(F \circ K)'(t)\| = \|DF_{K(t)}(Y - X)\| \leq \|DF_{K(t)}\| \|Y - X\| \leq c\|Y - X\|.$$

Nach dem bereits Bewiesenen (mit $G := F \circ K$ und $c\|Y - X\|$ statt c) folgt

$$\|F(Y) - F(X)\| = \|(F \circ K)(1) - (F \circ K)(0)\| \leq c\|Y - X\|. \qquad \blacksquare$$