Version 25.7.2007 Juni 2007

MIIA – Differentialrechnung mehrerer reeller Veränderlicher – SoSe 2007

Kurzfassung

Martin Schottenloher

$\infty\,\infty\,\infty$

Kapitel X. Differenzierbare Abbildungen

Jetzt kommen wir zu den differenzierbaren Abbildungen in mehreren Veränderlichen, die teilweise schon im vorangehenden Paragraphen aufgetreten sind.

Der Begriff der totalen Differenzierbarkeit wird eingehend im Paragraphen 31 dargestellt und mit der partiellen Differenzierbarkeit verglichen. Danach werden in $\S\,32$ elementare Eigenschaften differenzierbarer Abbildungen und die maßgeblichen Regeln bereitgestellt.

In $\S\,33$ behandeln wir höhere Ableitungen und die Taylorsche Formel, und im Paragraf $\S\,34$ gehen wir auf eine Anwendung ein: Die Beschreibung des lokalen Verhaltens von dreimal stetig differenzierbaren Funktionen mittels der zweiten Ableitungen (Hessematrix) und die Folgerungen für die Existenz von lokalen Extrema. Im abschließenden Paragrafen $\S\,35$ stellen wir eine Reihe von Resultaten für konvergente Funktionenfolgen zusammen, die wir großenteils schon in der einen oder anderen Form kennen und die sich in der Vertauschung von Grenzprozessen manifestieren.

Der Umkehrsatz für differenzierbare Abbildungen und der Satz über implizite Funktionen werden erst im kommenden Semester im Rahmen der Vorlesung MIII behandelt.

§31 Der Begriff der Differenzierbarkeit

Im Kapitel VI wurde die Differenzierbarkeit von Funktionen f in einer Variablen mit Werten in $\mathbb{R}$ eingeführt und im letzten Kapitel wurde der Begriff ausgedehnt auf Funktionen in einer Variablen mit Werten in einem Banachraum E. In beiden Fällen ist f definiert auf einem Intervall $I \subset \mathbb{R}$ und die Differenzierbarkeit von $f : I \to \mathbb{R}$ bzw. $f : I \to E$ in einem Punkte $a \in I$ bedeutet, dass der Grenzwert

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

existiert. In anderer Form geschrieben ist die Differenzierbarkeit äquivalent zu

$$\lim_{h\to 0, h\neq 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0\,.$$

Dabei ist $f'(a) \in \mathbb{R}$ bzw. $f'(a) \in E$. Man beachte, dass der dritte Term im Zähler der letzten Formel als eine stetige $\mathbb{R}$-lineare Abbildung $h \mapsto f'(a)h = hf'(a)$ von $\mathbb{R}$ nach $\mathbb{R}$ bzw. von $\mathbb{R}$ nach E verstanden werden kann, also als Element von $\mathcal{L}(\mathbb{R}, E)$. In diesem Zusammenhang ist es hilfreich die folgende Aussage aus den Übungen heranzuziehen. Jede $\mathbb{R}$-lineare Abbildung $\phi : \mathbb{R} \to E$ in einen normierten Raum E ist von der Form $\phi(h) = hb$ mit einem Element $b \in E$. Denn es gilt für $b := \phi(1) :\ \phi(h) = \phi(h \cdot 1) = h\phi(1) = hb$. Also ist ϕ automatisch stetig: $\|\phi(h)\| = |h|\,\|b\|$, und daher $|||\phi||| = \|b\|$. Es gilt also $\mathcal{L}(\mathbb{R}, E) \cong E$.

Wir übertragen das Erfolgsmodell der Differenzierbarkeit in einer Variablen auf den Fall von mehreren Veränderlichen, indem wir diesen Aspekt der „starken“ Approximation durch eine stetige lineare Abbildung beibehalten:

(31.1) Definition: Sei $f : \Omega \to F$ eine Abbildung auf einer offenen Menge $\Omega \subset E$ eines normierten Raumes E über $\mathbb{R}$ mit Werten in einem weiteren normierten Raum F über $\mathbb{R}$.

1° f heißt *(total) differenzierbar im Punkte* $a \in \Omega$, wenn es eine stetige $\mathbb{R}$-lineare Abbildung $L : E \to F$, also $L = L(a) \in \mathcal{L}(E, F)$, gibt mit

$$\lim_{h\to 0, h\neq 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0\,.$$

2° f heißt *(total) differenzierbar (in Ω)*, wenn f in allen $a \in \Omega$ differenzierbar ist.

3° Schließlich heißt f *stetig differenzierbar*, wenn f differenzierbar ist, und die linearen Abbildungen $L(a) \in \mathcal{L}(E, F)$ in

$$\lim_{h\to 0, h\neq 0} \frac{f(a+h) - f(a) - L(a)(h)}{\|h\|} = 0$$

zu einer stetigen Abbildung $L : \Omega \to \mathcal{L}(E, F)\,,\ a \mapsto L(a)$, führen.[1]

f ist also immer dann differenzierbar im Punkte a, wenn $L \in \mathcal{L}(E, F)$ existiert, so dass für alle Folgen (h_n) in E mit $h_n \to 0$ und $h_n \neq 0$ (sowie $a + h_n \in \Omega$) auch die Folge

$$\left(\frac{f(a+h_n) - f(a) - L(h_n)}{\|h_n\|}\right)_{n\in\mathbb{N}}$$

gegen Null strebt, also

$$\lim_{n\to\infty} \frac{\|f(a+h_n) - f(a) - L(h_n)\|}{\|h_n\|} = 0$$

gilt.

Im folgenden sind E und F in der Regel normierte Vektorräume über $\mathbb{R}$, z.B. $E = \mathbb{R}^n$ und $F = \mathbb{R}^m$, und $\Omega \subset E$ bezeichnet eine offene Menge.

(31.2) Bemerkungen:

1° Die Differenzierbarkeit von f in einem Punkte $a \in \Omega$ ist so zu verstehen, dass f durch die stetige lineare Abbildung $L = L(a)$ approximiert wird. Es gilt, dass der Ausdruck

$$\frac{f(a+h) - f(a) - L(h)}{\|h\|}$$

[1] Es wird gleich in 31.2.5° gezeigt, dass im Falle der Existenz die lineare Abbildung $L = L(a)$ in 1° eindeutig bestimmt ist.

gegen Null strebt für $h \to 0$. Zu diesen Sachverhalt sagt man auch, dass $L(h)$ die Differenz $f(a+h) - f(a)$ *in erster Ordnung approximiert* (siehe auch 31.10).

2° Die Differenzierbarkeit von f in a ist unabhängig von den Normen auf E und F, sie hängt nur von den Topologien (also dem jeweiligen Konvergenzbegriff in E und F) ab. Wird zum Beispiel die Norm $\| \ \|$ auf E durch eine äquivalente Norm $\| \ \|'$ ersetzt, so gilt nach Definition der Äquivalenz von Normen

$$c\,\|y\| \leq \|y\|' \leq C\,\|y\|$$

für $c, C \in \mathbb{R}$. Also strebt

$$\frac{f(a+h) - f(a) - L(h)}{\|h\|}$$

genau dann gegen 0 für $\|h\| \to 0$, wenn das für

$$\frac{f(a+h) - f(a) - L(h)}{\|h\|'} \,,$$

$\|h\|' \to 0$, der Fall ist.

Das bedeutet, dass zum Beispiel für $E = \mathbb{R}^n$ und $F = \mathbb{R}^m$ alle Normen $\| \ \|_p$ in der Definition 31.1 hergenommen werden können. [29.6.07]

3° Im Falle $E = \mathbb{R}$ und $\Omega = \,]t_0, t_1[$ stimmt die Differenzierbarkeit von $f : \,]t_0, t_1[\,\to \mathbb{R}$ nach 16.1 mit der neuen Definition 31.1 überein, und dasselbe gilt für den Differenzierbarkeitsbegriff für $f : \,]t_0, t_1[\,\to E$ mit Werten in Banachräumen nach 25.3.

4° Wir betrachten hier nur $\mathbb{R}$-lineare Approximationen, insofern nur die reelle Differenzierbarkeit. Die komplexe Differenzierbarkeit, also $L = L(a)$ komplex linear, führt zu einer eng verwandten Theorie, der Funktionentheorie (siehe 30.19), die aber im Vergleich zur reellen Differenzierbarkeit einige wesentliche Unterschiede zu verzeichnen hat. Der komplexe Fall wird hier nicht weiter behandelt (siehe aber z.B. 31.14).

5° Ist f in a differenzierbar, so ist die stetige $\mathbb{R}$-lineare Abbildung $L \in \mathcal{L}(E, F)$, die

$$\lim_{h \to 0, h \neq 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0$$

erfüllt, eindeutig bestimmt.

(31.3) Definition: Sei $f : \Omega \to F$ in $a \in \Omega$ differenzierbar mit

$$\lim_{h \to 0, h \neq 0} \frac{f(a+h) - f(a) - L(h)}{\|h\|} = 0 \,.$$

Dann heißt die eindeutig bestimmte stetig lineare Abbildung $L = L(a) : E \to F$ die *Ableitung von f in a* und wird hier meistens folgendermaßen bezeichnet:

$$Df(a) = L(a) \,.$$

Andere Notationen, die gebräuchlich sind:

$$f'(a) = J_f(a) = \partial f(a) = Tf(a) \,.$$

Die Ableitung ist also nicht mehr einfach eine Zahl oder ein Vektor in F, sondern selbst eine Abbildung. In einer Variablen allerdings kann die Ableitung durch eine Zahl bzw. durch einen

Vektor repräsentiert werden.

Die Auswertung von $Df(a) \in \mathcal{L}(E, F)$ in $h \in E$ wird oft $Df(a)h$ oder $Df(a).h$ geschrieben. Also

$$L(h) = Df(a)(h) = Df(a)h = Df(a).h \, , \, h \in E,$$

und analog in Bezug auf die anderen Notationen. Die Notation $J_f(a)$ wird gelegentlich im Falle $E = \mathbb{R}^n$ und $F = \mathbb{R}^m$ bevorzugt, wenn die lineare Abbildung als Matrix verstanden wird, der so genannten *Jacobi-Matrix*, vgl. 31.7.

(31.4) Beispiele:

1° $u(x, y) := x + y \, , \, (x, y) \in \mathbb{R}$.

Es gilt dann für $a = (\xi, \eta)$, $h = (s, t)$ und $L(h) := s + t$:

$$u(a + h) - u(a) - L(h) = (\xi + s + \eta + t) - (\xi + \eta) - (s + t) = 0 \, ,$$

daher

$$\lim_{h \to 0} \frac{u(a + h) - u(a) - L(h)}{\|h\|} = 0 \, .$$

Also ist u differenzierbar mit konstanter Ableitung $Du(a) = L = u$ (vgl. 3°). Insbesondere ist u stetig differenzierbar, weil $a \mapsto Du(a)$ konstant ist.

2° $u(x, y) := x^2 + y^2 \, , \, (x, y) \in \mathbb{R}^2$.

Es gilt dann für $a = (\xi, \eta)$, $h = (s, t)$ und $L(h) := 2\xi s + 2\eta t$:

$$u(a + h) - u(a) - L(h) = (\xi + s)^2 + (\eta + t)^2 - (\xi^2 + \eta^2) - (2\xi s + 2\eta t) = s^2 + t^2 = \|h\|^2 \, ,$$

daher

$$\lim_{h \to 0} \frac{u(a + h) - u(a) - L(h)}{\|h\|} = 0 = \lim_{h \to 0} \|h\| = 0 \, .$$

Also ist u differenzierbar mit Ableitung $Du(a)(s, t) = 2\xi s + 2\eta t$ (vgl. 4°). u ist stetig differenzierbar.

Im übrigen sind sämtliche Polynome in zwei und auch in n Veränderlichen total differenzierbar mit stetiger Ableitung, die wiederum differenzierbar ist, wie wir in den Übungen sehen. Siehe auch 33.6.

3° Sei $f : E \to F$ eine stetige $\mathbb{R}$-lineare Abbildung, also $f \in \mathcal{L}(E, F)$.

Dann gilt $f(a + h) - f(a) = f(h)$ und es folgt $Df(a) = f$ konstant. f ist stetig differenzierbar. Dieses Beispiel verallgemeinert 1°.

4° Sei $B : E \times E \to F$ $\mathbb{R}$-bilinear und stetig.

Dann ist die zugehörige quadratische Form $q(x) := B(x, x) \, , \, x \in E = \Omega$, stetig differenzierbar mit $Dq(a)h = B(h, a) + B(a, h)$. Denn es ist $M = \sup\{\|B(v, w)\| : \|v\| \leq 1 \, , \, \|w\| \leq 1\} < \infty$ wegen der Stetigkeit von B, und es gilt $\|B(x, y)\| \leq M \, \|x\| \, \|y\|$. Also ist $Dq(a)$ stetige lineare Abbildung. Es ist jetzt

$$q(a+h) - q(a) - Dq(a)h = B(a, a) + B(a, h) + B(h, a) + B(h, h) - B(a, a) - B(a, h) - B(h, a) = B(h, h)$$

und daher

$$\frac{\|q(a + h) - q(a) - Dq(a)h\|}{\|h\|} \leq M \, \|h\| \longrightarrow 0$$

für $\|h\| \to 0$.

Im Falle $E = \mathbb{R}^n$ hat B vermöge der Standardbasis (e_k) von $\mathbb{R}^n$ die Darstellung durch die Matrix (g_{ik}) mit $g_{ik} := B(e_i, e_k) \in F$, das heißt $B(x,y) = \sum_{i,k=1}^n g_{ik} x^i y^k$ für $x = x^i e_i$, $y = y^k e_k$. Daher hat $Dq(a)$ die Basisdarstellung

$$Dq(a)h = \sum_{i,k=1}^n (g_{ik} a^i h^k + g_{ki} h^k a^i) = \sum_{k=1}^n \left(\sum_{i=1}^n (g_{ik} + g_{ki}) a^i \right) h^k ,$$

also hat man mit $q_k(a) := \sum_{i=1}^n (g_{ik} + g_{ki}) a^i \in F$ die folgende Darstellung der Ableitung:

$$Dq(a).h = \sum_{k=1}^n q_k(a) h^k .$$

Wenn noch $F = \mathbb{R}^m$ gilt, so hat die lineare Abbildung $Dq(a) : \mathbb{R}^n \to \mathbb{R}^m$ die Matrixdarstellung

$$Dq(a) = \left(q_k^j(a) \right)_{1 \le j \le m, 1 \le k \le n} ,$$

wobei $q_k(a) = \sum_{j=1}^m q_k^j(a) e_j \in \mathbb{R}^m$. Es gilt also

$$Dq(a).h = \sum_{j=1}^m \sum_{k=1}^n q_k^j(a) h^k e_j .$$

Gut bekannt ist der Fall $B(x,y) = \langle x, y \rangle$ mit $E = \mathbb{R}^n$ und $F = \mathbb{R}$. B ist symmetrisch und die g_{ik} sind die bekannten Kroneckersymbole: $g_{ik} = \delta_{ik}$. Es ist dann $q_k(a) = 2 \sum g_{ik} a^i = 2a^k$ und $Dq(a)h = \sum_{i,k=1}^n g_{ik} a^i h^k$.

Dieses Beispiel ist im Übrigen eine Verallgemeinerung von 2°.

Wie zuvor in Kapitel IX reduziert sich der vektorwertige Fall auf den skalaren Fall:

(31.5) Lemma: *Sei $f : \Omega \to F$ eine Abbildung auf der offenen Menge $\Omega \subset E$.*

1° *Sei f in $a \in \Omega$ differenzierbar. Für alle $\phi \in F' = \mathcal{L}(F, \mathbb{R})$ ist dann $\phi \circ f$ differenzierbar mit*

$$D(\phi \circ f)(a) = \phi \circ Df(a) ,$$

vgl.

$$E \xrightarrow{Df(a)} F \xrightarrow{\phi} \mathbb{R} .$$

Dies ist ein Spezialfall der Kettenregel (vgl. 32.3) unter Verwendung von $D\phi = \phi$ nach 31.4.3°.

2° Im Falle $F = \mathbb{R}^m$ und $f = (f^1, \dots, f^m)$ ist f schon differenzierbar in a, wenn das für alle skalaren $f^j : \Omega \to \mathbb{R}$, $j = 1, \dots, m$, gilt.[2] Für die Ableitung $Df(a) \in \mathcal{L}(E, \mathbb{R}^m) \cong \mathcal{L}(E, \mathbb{R})^m$ ist dann

$$Df(a) = \left(Df^1(a), \dots, Df^m(a) \right) .$$

Diese Aussage hat eine natürliche Verallgemeinerung auf den Fall eines Produktes $F = F_1 \times F_2 \times \ldots \times F_m$ von normierten Räumen F_j. $f : \Omega \to F$ ist in $a \in \Omega$ differenzierbar, wenn das für alle Komponenten $f^j : \Omega \to F_j$ gilt mit derselben Formel wie für den Fall von $F_j = \mathbb{R}$.

[2] 2° hat eine Entsprechung auch für den Fall eines unendlichdimensionalen Banachraumes F: *f in a differenzierbar $\iff$ $\phi \circ f$ in a differenzierbar $\forall\, \phi \in F'$*. Der Beweis dazu erfordert die Formel $\|v\| = \sup\{|\phi(v)| : \phi \in F' , \|\phi\| \le 1\}$ für $v \in F$, die aus dem Satz von Hahn-Banach folgt.

3° Im Falle $F = \mathbb{R}^m$ ist f in Ω genau dann stetig differenzierbar, wenn das für alle f^j zutrifft.

Beweis: 1° folgt aus der Stetigkeit von ϕ und aus

$$\frac{\phi \circ f(a+h) - \phi \circ f(a) - \phi \circ Df(a).h}{\|h\|} = \phi\left(\frac{f(a+h) - f(a) - Df(a).h}{\|h\|}\right).$$

2° folgt aus

$$\frac{f(a+h) - f(a) - Df(a).h}{\|h\|} = \left(\frac{f^1(a+h) - f^1(a) - Df^1(a).h}{\|h\|}, \dots, \frac{f^m(a+h) - f^m(a) - Df^m(a).h}{\|h\|}\right)$$

und der Tatsache, dass in $\mathbb{R}^m$ eine Folge (y_n) genau dann gegen 0 konvergiert, wenn alle Komponenten (y_n^j) , $j = 1, 2, \dots, m$, gegen 0 konvergieren.

3° ergibt sich direkt aus 1° und 2°. ■

(31.6) Satz: *Es sei* $f : \Omega \to F$ *differenzierbar im Punkte* $a \in \Omega$. *Dann existieren alle Richtungsableitungen*

$$L_v f(a) := \frac{d}{dt} f(a+tv)|_{t=0} = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t} = Df(a).v$$

für $v \in E$. *Im Falle* $E = \mathbb{R}^n$ *existieren also insbesondere die partiellen Ableitungen*

$$\frac{\partial f}{\partial x^k}(a) := \frac{d}{dt} f(a + te_k)|_{t=0} \in F\,,\ k = 1, \dots, n\,.$$

Diese Festlegung der partiellen Ableitung verallgemeinert die in 30.10 getroffene Definition auf den vektorwertigen Fall.

Die partielle Differenzierbarkeit erweist sich als notwendige Bedingung für die totale Differenzierbarkeit.

(31.7) Matrixdarstellung: Im Falle von $E = \mathbb{R}^n$ und $F = \mathbb{R}^m$ hat die Ableitung $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ einer in a (total) differenzierbaren Abbildung $f : \Omega \to \mathbb{R}^m$, $f = (f^1, \dots, f^m)$, die folgende Darstellung als Matrix:

$$Df(a) = \left(\partial_k f^j(a)\right) = \left(\frac{\partial f^j}{\partial x^k}(a)\right).$$

Es gilt also

$$Df(a)h = \sum_{j=1}^{m} \sum_{k=1}^{n} \partial_k f^j(a) h^k\, e_j\,.$$

Die Matrix $\left(\partial_k f^j(a)\right)$ heißt *Jacobi-Matrix* und wird auch mit $J_f(a)$ bezeichnet.

Im Falle $m = 1$ ist

$$Df(a)h = \sum_{k=1}^{n} \partial_k f(a) h^k = \langle \nabla f(a), h \rangle$$

und $J_f(a) = (\partial_1 f(a), \dots, \partial_n f(a))$. Die Jacobi-Matrix erweist sich hier also als der bereits bekannte Gradient $\nabla f(a)$ von f in a. [3.7.07]

Wenn also für $f = (f^1, \dots, f^m) : \Omega \to \mathbb{R}^m$ die Funktionen f^j partiell differenzierbar in a sind, so ist die gerade definierte Matrix der Kandidat für die (totale) Ableitung von f in a, vgl. 31.13.

(31.8) Beispiel: Die Funktion

$$u(x,y) := \begin{cases} \frac{xy}{x^2+y^2}, & \text{wenn } (x,y) \neq 0 \text{ und} \\ 0 & \text{sonst.} \end{cases}$$

ist partiell differenzierbar, aber nicht differenzierbar in 0.

Diese Funktion ist nicht einmal stetig, denn für $(x,y) = (t,t)$ ist $u(t,t) = \frac{1}{2}$, auch im Grenzwert $t \to 0$. Für $(x,y) = (t,2t)$ gilt $u(t,2t) = \frac{2}{5}$, auch im Grenzwert $t \to 0$.

Daraus hätte man schon ableiten können, dass f in 0 nicht differenzierbar ist, wie das folgende einfache Kriterium ergibt.

(31.9) Satz: Eine in $a \in \Omega$ total differenzierbare Abbildung $f : \Omega \to F$ ist in a stetig.

In der Tat kann die totale Differenzierbarkeit von f in a in der Form

$$f(a+h) = f(a) + Df(a).h + r(h) \quad \text{mit} \quad \lim_{h \to 0, h \neq 0} \frac{r(h)}{\|h\|} = 0$$

geschrieben werden. Daher gilt

$$\|f(a+h) - f(a)\| \leq \left(|||Df(a)||| + \frac{r(h)}{\|h\|} \right) \|h\| \ \longrightarrow 0$$

für $h \to 0$.

(31.10) Bemerkung: Die Bedingung

$$\lim_{h \to 0, h \neq 0} \frac{r(h)}{\|h\|} = 0\,, \ \text{ also } \ \frac{r(h)}{\|h\|} \ \longrightarrow \ 0\,,$$

wird gelegentlich durch $o(\|h\|)$ (in Worten „klein-o von Norm-h“) abgekürzt. Man kann dann die Differenzierbarkeit durch

$$f(a+h) = f(a) + Df(a).h + o(\|h\|)$$

ausdrücken.

Wir formulieren jetzt ein weiteres wichtiges Kriterium (hinreichend für die totale Differenzierbarkeit):

(31.11) Satz: *Im Falle $E = \mathbb{R}^n$ sei die Funktion $u : \Omega \to \mathbb{R}$ stetig partiell differenzierbar (im Sinne der Definition 30.10). Dann ist u auch total differenzierbar mit*

$$Du(a).h = \langle \nabla u(a), h \rangle = \sum_{k=1}^{n} \frac{\partial u}{\partial x^k}(a) h^k\,,$$

wobei $h = (h^1, \ldots, h^n) \in \mathbb{R}^n$. u ist dann stetig differenzierbar.

Beweis:

$$u(a+h) - u(a) = \sum_{k=1}^{n} \left(u(a + h^1 e_1 + \ldots + h^k e_k) - u(a + h^1 e_1 + \ldots + h^{k-1} e_{k-1}) \right)\,.$$

Auf jeden Summanden wird der elementare Mittelwertsatz 17.3 angewandt mit dem Ergebnis

$$u(a+h)-u(a)=\sum_{k=1}^{n}\left(\partial_k u(a+h^1e_1+\ldots+h^{k-1}e_{k-1}+\theta_k h^k e_k)\right)h^k\,,$$

wobei $\theta_k\in[0,1]$. Damit folgt die Behauptung aus $\lim_{h\to 0}\partial_k u(a+h^1e_1+\ldots+h^{k-1}e_{k-1}+\theta_k h^k e_k)=\partial_k u(a)$. ($\partial_k u$ ist in a stetig.) ■

(31.12) Bemerkung: Partielle Ableitungen definiert man im Falle $E=\mathbb{R}^n$ auch für Abbildungen $f:\Omega\to F$ mit Werten in einem allgemeinen normierten Raum F (insbesondere für $E=\mathbb{R}^m$) durch

$$\frac{\partial f}{\partial x^k}(a)=\partial_k f(a)=\frac{d}{dt}f(a+te_k)|_{t=0}\,,$$

(vgl. 31.6) und man erhält die zu 31.11 analoge Aussage:

$$f\ \text{stetig partiell differenzierbar}\iff f\ \text{stetig differenzierbar}\,.$$

Und es gilt die entsprechende Formel

$$Df(a).h=\frac{d}{dt}f(a+th)|_{t=0}=\sum_{k=1}^{n}\partial_k f(a)h^k\,.$$

Mit $\mathcal{C}^1(\Omega,F)$ wird die Menge der stetig differenzierbaren Abbildungen $\Omega\to F$ bezeichnet (die natürlich wieder ein $\mathbb{R}$-Vektorraum ist, wie wir gleich im nächsten Paragraphen sehen werden).

(31.13) Differenzierbarkeitstest: Bei der Überprüfung, ob eine Abbildung $g:\Omega\to F$ in n Veränderlichen (stetig) differenzierbar ist, kann man sukzessive folgendermaßen vorgehen:

Testfrage A (Notwendige Bedingung): Ist g stetig in $a\in\Omega$?

Testfrage B (Notwendige Bedingung): Wenn ja, ist g partiell differenzierbar in a?

Testfrage C (Hinreichende Bedingung): Wenn auch das zutrifft, sind die partiellen Ableitungen stetig in a?

Testfrage D (Notwendige und hinreichende Bedingung im Falle $F=\mathbb{R}^m$): Wenn g in a partiell differenzierbar ist, erfüllt die Matrix $L=(\partial_k g^j(a))$ die Differenzierbarkeitsbedingung, d.h. $g(a+h)=g(a)+L(h)+o(\|h\|)$?

(31.14) Beispiele zu diesen Tests:

1° $u(x,y)=x\,|y|$ ist nicht partiell differenzierbar in 0.

2° $u(x,y)=\frac{\cos x}{\sin y}$ ist stetig differenzierbar in $\Omega=\{(x,y)\mid\forall\,k\in\mathbb{Z}:y\neq\pi k\}$, weil u dort stetig partiell differenzierbar ist.

3° $f(x,y)=e^x(\cos y,\sin y)=e^z$, $(x,y)\in\mathbb{R}^2$, ist stetig differenzierbar in $\Omega=\mathbb{R}^2\cong\mathbb{C}$, weil f dort stetig partiell differenzierbar ist. Mit $f=(u,v)$, $u=e^x\cos y$, $v=e^x\sin y$ ist

$$J_f=\begin{pmatrix}u&-v\\v&u\end{pmatrix}$$

Es gilt mit $h = (s,t) = s + it$ als Notation: $Df(z).h = (us - vt, vs + ut)$ und in komplexer Schreibweise $Df(z).h = (us - vt) + i(vs + ut) = (u + iv) \cdot (s + it) = e^z \cdot h$, wobei $\cdot$ die komplexe Multiplikation bezeichnet. Die totale Ableitung $Df(z) : \mathbb{R}^2 \to \mathbb{R}^2$, ist also in diesem Falle komplex linear, und die Funktion f ist holomorph (siehe 30.19).

4° $u(x,y) = (x^2 + y^2)\sin((x^2 + y^2)^{-\frac{1}{2}})$, $(x,y) \neq (0,0)$, hat eine stetige Fortsetzung nach 0 (durch $u(0) := 0$), die (total) differenzierbar aber nicht stetig partiell differenzierbar ist.

§32 Erste Eigenschaften und fundamentale Regeln

Wie für die Stetigkeit für allgemeine topologische Räume und wie für die Differenzierbarkeit von Funktionen auf Intervallen gelten auch für die totale Differenzierbarkeit die üblichen Permanenzeigenschaften: Unter Addition, Multiplikation und Hintereinanderausführung bleibt die Differenzierbarkeit erhalten.

Mit E, F werden in diesem Paragrafen wieder normierte Räume über $\mathbb{R}$ bezeichnet, z.B. $E = \mathbb{R}^n, F = \mathbb{R}^m$, jeweils mit der euklidischen Norm, und mit $\Omega \subset E$ offene Teilmengen von E. In vielen Fällen ist es sinnvoll, F als vollständig, also als Banachraum vorauszusetzen.

(32.1) Additionsregel: Es seien $f, g : \Omega \to F$ differenzierbar im Punkte $a \in \Omega$. Dann ist auch $f + g : \Omega \to F$ differenzierbar in a, und es gilt

$$D(f+g)(a) = Df(a) + Dg(a)\,,$$

kurz:

$$(f+g)' = f' + g'\,.$$

(32.2) Produktregel: Es seien $f : \Omega \to F$ und $g : \Omega \to G$ differenzierbar in $a \in \Omega$, wobei G ein weiterer normierter Raum über $\mathbb{R}$ ist. Dann ist im Falle $F = \mathbb{R}$ auch das Produkt $fg : \Omega \to G$, in a differenzierbar, und es gilt

$$D(fg)(a) = Df(a)g(a) + f(a)Dg(a)\,,$$

kurz:

$$(fg)' = f'g + fg'\,.$$

$\mathcal{C}^1(\Omega, F)$ ist also ein Vektorraum über $\mathbb{R}$.

Allgemeiner gilt für eine stetige bilineare Abbildung $B : F \times G \to H$ in einen weiteren normierten Raum über $\mathbb{R}$, dass auch $h := B(f,g) : \Omega \to H$, $x \mapsto B(f(x), g(x))$ in a differenzierbar ist mit entsprechender Formel für die Ableitung:

$$Dh(a) = B(Df(a), g(a)) + B(f(a), Dg(a))$$

im Sinne von $Dh(a)v = B(Df(a)v, g(a)) + B(f(a), Dg(a)v)$, $v \in E$.
Kurz:

$$h' = B(f', g) + B(f, g')\,.$$

Beweis: Im ersten Fall wird $r(h) := fg(a+h) - fg(a) - \big((Df(a).h)g(a) + f(a)Dg(a).h\big)$ in der Form

$$r(h) = \big(f(a+h) - f(a)\big)g(a+h) + f(a)\big(g(a+h) - g(a)\big) - (Df(a).h)g(a) - f(a)Dg(a).h$$

also

$$r(h) = \big(f(a+h) - f(a) - Df(a).h\big)g(a+h) + f(a)\big(g(a+h) - g(a) - Dg(a).h\big) + (Df(a).h)\big(g(a+h) - g(a)\big)$$

geschrieben. Zu $M := \max\{2\,\|g(a)\|\,, \|f(a)\|\,, |||Df(a)|||\}$ gibt es $r > 0$, so dass $\|g(a+h)\| \leq M$ ist für $\|h\| < r$. Die Dreiecksungleichung liefert daher

$$\frac{\|r(h)\|}{\|h\|} \leq M \frac{\|f(a+h) - f(a) - Df(a).h\| + \|(g(a+h) - g(a) - Dg(a).h\| + \|g(a+h) - g(a)\|}{\|h\|}\,,$$

woraus

$$\lim_{h \to 0} \frac{\|r(h)\|}{\|h\|} = 0$$

unmittelbar folgt.

Der allgemeine Fall hat einen analogen Beweis. Er lässt sich allerdings auch mit Hilfe der Ableitung einer stetigen bilinearen Abbildung $B : F \times G \to H$ (analog zu 31.4.4°) aus der Kettenregel herleiten. ■

(32.3) Kettenregel: Es seien $f : \Omega \to F$ in $a \in \Omega$ und $g : \Sigma \to G$ in $f(a) = b \in \Sigma$ differenzierbar ($\Sigma \subset F$ offen), wobei noch $f(\Omega) \subset \Sigma$ vorausgesetzt sei. Dann ist auch $g \circ f : \Omega \to G$ in a differenzierbar, und es gilt

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a)\,,$$

kurz:

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

Beweis: Es sei $h \in E$. Setze $k := f(a+h) - f(a)$. Falls es $\delta > 0$ geben sollte mit $f(a+th) - f(a) = 0$ für alle $t \in \mathbb{R}\,,\ |t| < \delta$, so ist $Df(a).h = 0$ und

$$g \circ f(a+h) - g \circ f(a) - Dg(b) \circ Df(a).h = 0\,.$$

Ansonsten sei h für kleine $\|h\|$ so gewählt, dass $k = f(a+h) - f(a) \neq 0$ ist. Dann gilt für

$$r(h) := g(f(a+h)) - g(f(a)) - Dg(b) \circ Df(a).h = g(b+k) - g(b) - Dg(b).k + Dg(b).(k - Df(a).h)\ :$$

$$\frac{\|r(h)\|}{\|h\|} = \frac{\|g(b+k) - g(b) - Dg(b).k + Dg(b).(f(a+h - f(a) - Df(a).h)\|}{\|h\|}\,,$$

also

$$\frac{\|r(h)\|}{\|h\|} \leq \frac{\|g(b+k) - g(b) - Dg(b).k\|}{\|k\|}\frac{\|k\|}{\|h\|} + \frac{\|Dg(b).(f(a+h - f(a) - Df(a).h)\|}{\|h\|}\,.$$

Wegen

$$\frac{\|k\|}{\|h\|} = \frac{\|f(a+h) - f(a) - Df(a).h + Df(a).h\|}{\|h\|} \leq \frac{\|f(a+h) - f(a) - Df(a).h\|}{\|h\|} + |||Df(a)|||$$

ist $\frac{\|k\|}{\|h\|}$ beschränkt durch ein $M > 0$, wenn $\|h\| < r$ für ein geeignetes $r > 0$, und wir erhalten

$$\frac{\|r(h)\|}{\|h\|} \leq \frac{\|(g(b+k) - g(b) - Dg(b).k)\|}{\|k\|} M + |||Dg(b)|||\, \frac{\|f(a+h) - f(a) - Df(a).h\|}{\|h\|}\,,$$

und dieser Ausdruck konvergiert offensichtlich gegen 0 für $h \to 0$ (denn aus $h \to 0$ folgt $k \to 0$, weil f stetig ist). ■

(32.4) Folgerung: Für eine Funktion $u : \Omega \to \mathbb{R}$ und eine Kurve $\gamma : [t_0, t_1] = I \to \Omega$ gilt:

1° Ist γ in $t \in I$ differenzierbar und ist u in $\gamma(t)$ differenzierbar, so ist $u \circ \gamma$ in t differenzierbar mit

$$D(u \circ \gamma)(t) = Du(\gamma(t)).\dot{\gamma}(t)\,.$$

2° Ist γ in $t \in I$ differenzierbar und ist im Falle $E = \mathbb{R}^n$ die Funktion u in $\gamma(t)$ stetig partiell differenzierbar, so hat man wegen 31.11 dieselbe Konklusion mit der Darstellung

$$D(u \circ \gamma)(t) = \langle \nabla u(\gamma(t)), \dot{\gamma}(t) \rangle = \sum_{k=1}^{n} \partial_k u(\gamma(t)) \dot{\gamma}^k(t) \,.$$

Damit ist der angekündigte Beweis der in $\S\, 30$ zitierten Kettenregel 30.11 nachgeholt.

(32.5) Folgerung: Ist $u : \Omega \to \mathbb{R}$ differenzierbar in $a \in \Omega \subset \mathbb{R}^n = E$, so bestimmt $\nabla u(a)$ im Falle $\nabla u(a) \neq 0$ *die Richtung des stärksten Anstiegs von* u *bei* a *in folgendem Sinne:* Für alle $v \in \mathbb{R}^n$ mit $\|v\| = 1$ erfüllt die Richtungsableitung

$$|L_v u(a)| \leq \|\nabla u(a)\|$$

und für

$$w := \frac{\nabla u(a)}{\|\nabla u(a)\|}$$

gilt

$$L_w u(a) = \|\nabla u(a)\| \;.$$

Denn es gilt stets $L_v u(a) = \langle \nabla u(a), v \rangle$, also $|L_v u(a)| \leq \|\nabla u(a)\| \, \|v\| = \|\nabla u(a)\|$, wenn $\|v\| = 1$, und es ist $L_w u(a) = \|\nabla u(a)\|$.

Inwiefern ist $L_v u(a)$ als Anstieg zu verstehen? Es geht um den *Graphen* Γ der Funktion u, $\Gamma := \{(x, u(x)) \in \Omega \times \mathbb{R} : x \in \Omega\} \subset \mathbb{R}^{n+1}$, den man sich im Falle $n = 2$ als eine Fläche im $\mathbb{R}^3$ über Ω vorstellen kann (vgl. die Abbildungen unten). Und es geht zu $a \in \Omega$ um die *Tangentialvektoren* an Γ in $(a, u(a)) \in \Gamma$, das sind die Geschwindigkeitsvektoren $\dot{\beta}(0)$ von differenzierbaren Kurven $\beta : [-r, r] \to \Gamma$ mit $\beta(0) = (a, u(a))$. Diese Kurven in Γ kann man auch mit Hilfe von differenzierbaren Kurven $\alpha : [-r, r] \to \Omega$ mit $\alpha(0) = a$ beschreiben als $\beta(t) = (\alpha(t), u \circ \alpha(t))$. Es gilt $\dot{\beta}(0) = (\dot{\alpha}(0), \langle \nabla u(a), \dot{\alpha}(0) \rangle)$. Die Gesamtheit aller dieser Tangentialvektoren in $\mathbb{R}^{n+1}$ wird daher wegen $\langle \nabla u(a), e_j \rangle = \partial_j u(a)$ durch die Vektoren $(e_j, \partial_j u(a))\,,\ j = 1, 2, \ldots, n$, als Vektorraum aufgespannt. Dieser Untervektorraum aller Tangentialvektoren an Γ in $(a, u(a))$ wird *Tangentialvektorraum* (*Tangentialebene* im Falle $n = 2$) genannt und mit $T_a \Gamma$ bezeichnet. Die gerade durchgeführte Beschreibung der Tangentialvektoren zeigt, dass $T_a \Gamma$ das orthogonale Komplement von $(\nabla u(a), -1)$ ist, $T_a \Gamma = (\nabla u(a), -1)^{\perp}$, denn es gilt

$$\left\langle \dot{\beta}(0), (\nabla u(a), -1) \right\rangle = \langle \dot{\alpha}(0), \nabla u(a) \rangle - \langle \nabla u(a), \dot{\alpha}(0) \rangle) = 0 \,.$$

Das passt zu unserer Erfahrung im Falle $n = 1$:

$$T_a \Gamma = (f'(a), -1)^{\perp} = \{(x, y) \in \mathbb{R}^2 : y = f'(a)x\} \,,$$

wobei $u = f$, und liest sich im Falle $n = 2$:

$$T_a \Gamma = \{(x, y, z) \in \mathbb{R}^3 : z = \partial_1 u(a)x + \partial_2 u(a)y\} \,.$$

Der *Anstieg in Richtung* $v \in \mathbb{R}^n$, $\|v\| = 1$, von dem die Rede ist, ist die Steigung der Kurve β_v, die durch $\alpha(t) = a + tv$ induziert wird, also die Steigung von $\beta_v(t) = (a + tv, u(a + tv))$ in Relation zu v für $t = 0$. Oder, anders ausgedrückt, die Steigung von $\beta_v(t)$ in der durch v und $(0, \ldots, 0, 1) \in \mathbb{R}^{n+1}$ bestimmten Ebene E_v (siehe Abbildung 3). Das ist aber gerade die letzte Komponente von $\dot{\beta}_v(0)$, das Skalarprodukt $\langle \nabla u(a), v \rangle = L_v u(a)$.

In den folgenden drei Abbildungen wird der Fall $n = 2$ illustriert, wobei ohne Einschränkung der Allgemeinheit $a = 0, u(a) = 0$ und $\Omega =]-1,1[\times]-1,1[\subset \mathbb{R}^2$ angenommen wird. Zudem wird $\nabla u(a) \neq 0$ vorausgesetzt. Schließlich sind die (x, y)-Koordinaten noch so gewählt, dass $\nabla u(a)$ in Richtung der y-Achse zeigt. Der allgemeine Fall ergibt sich dann durch eine geeignete Drehung des Raumes und damit von $\nabla u(a)$ und $T_a\Gamma$ um die z-Achse.

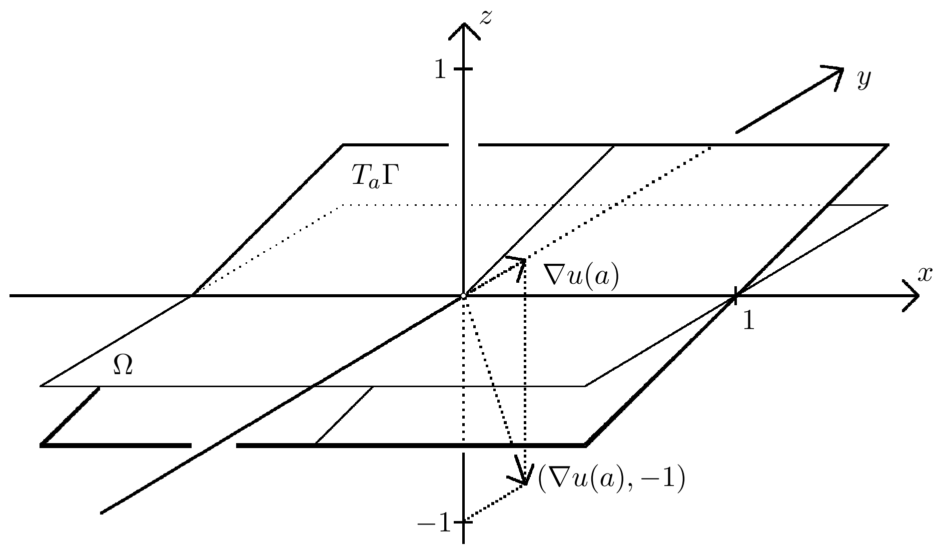

Abb. 1: Die Tangentialebene $T_a\Gamma$ und ihre Festlegung durch den Gradientenvektor $\nabla u(a)$ über dem Definitionsbereich $\Omega =]-1,1[\times]-1,1[$

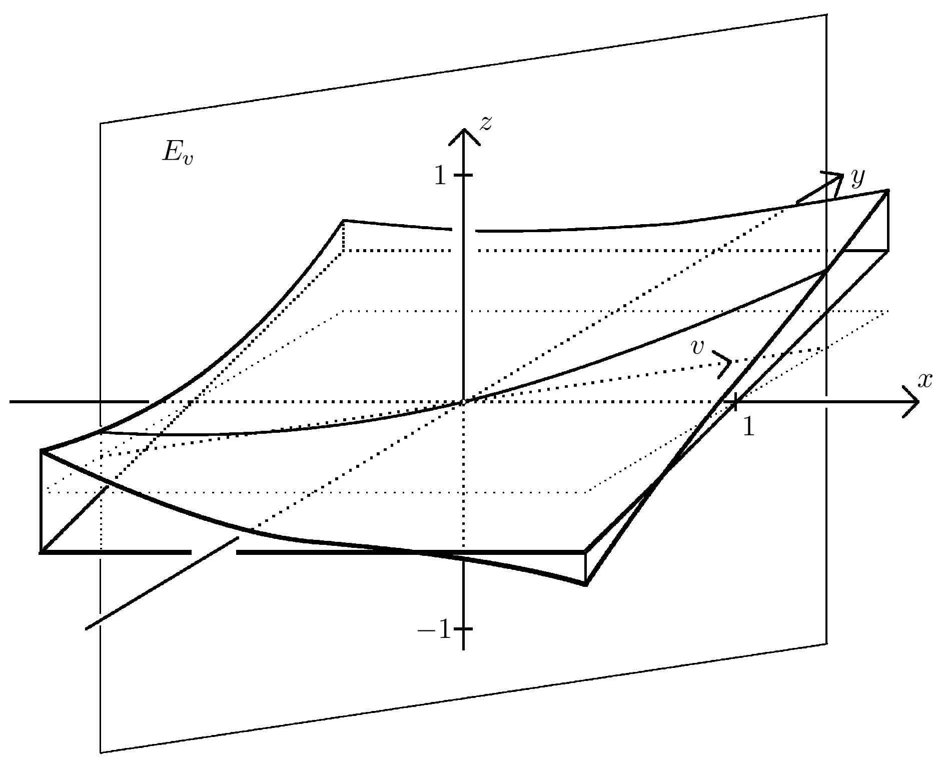

Abb. 2: Die Fläche $\Gamma \subset \mathbb{R}^3$, die durch v gegebene Ebene E_v und die Spur der Kurve β_v als Schnitt der Fläche Γ und der Ebene E_v

Durch Weglassen von allen Daten außerhalb der Ebene E_v erhalten wir die Größe $L_v u(a) = \langle \nabla u(a), v\rangle$ als Tangentensteigung m (Anstieg) von der Funktion $z = \beta_v(t)$ in $t = 0$:

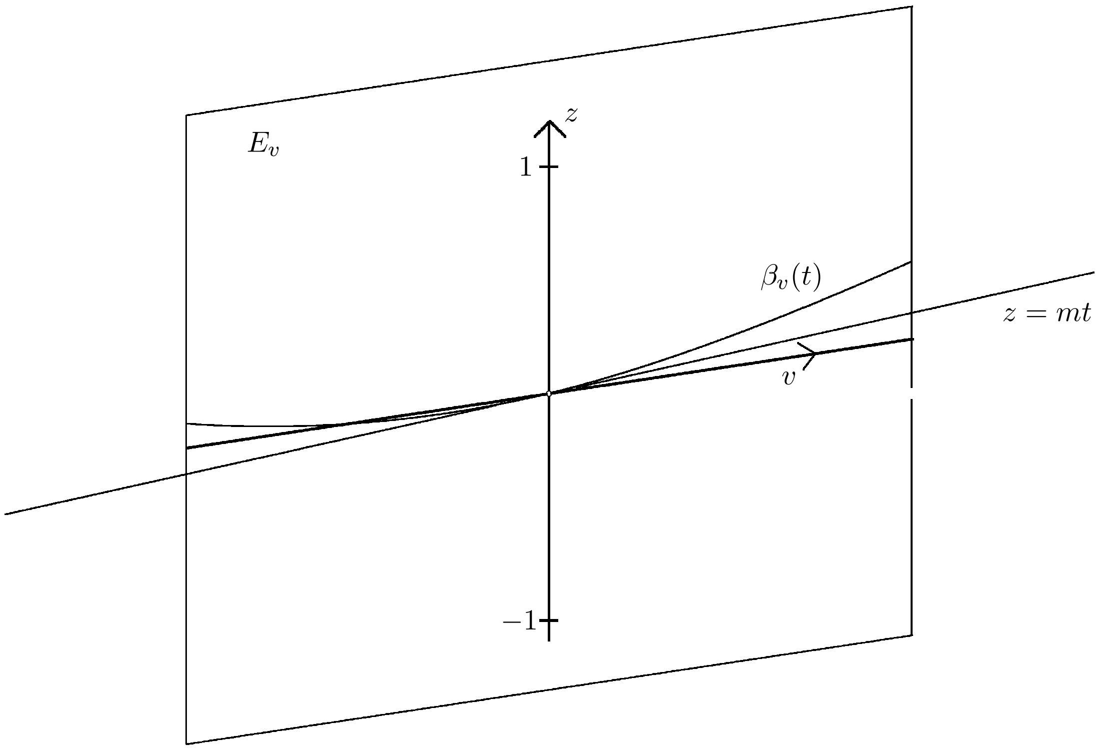

Abb. 3: Die Spur der Kurve β_v in der Ebene E_v
und die Tangente $z = mt$ mit der Steigung $m = L_v u(a) = \langle \nabla u(a), v\rangle$

Die Steigung $m = L_v u(a)$ ist der Anstieg in Richtung $v \in \mathbb{R}^2$, $\|v\| = 1$, und 35.5 besagt, dass im Falle $\nabla u(a) \neq 0$ der Anstieg am größten ist in Richtung des Einheitsvektors

$$v = \frac{\nabla u(a)}{\|\nabla u(a)\|}\,.$$

(32.6) Folgerung: Ist $u : \Omega \to \mathbb{R}$ differenzierbar in $a \in \Omega \subset \mathbb{R}^n = E$ mit $\nabla u(a) = 0$ (a heißt dann ein *kritischer Punkt*), so verschwinden alle Richtungsableitungen: $L_v u(a) = 0$. Diese Situation liegt insbesondere dann vor, wenn u in a ein lokales Extremum hat.

Wir haben damit $\nabla u(a) = 0$, das heißt $\partial_k u(a) = 0$ für alle $k = 1, \ldots, n$, als notwendige Bedingung dafür, dass eine in a differenzierbare Funktion $u : \Omega \to \mathbb{R}$ in a ein lokales Extremum hat. Hinreichende Kriterien beschreiben wir später mit Hilfe höherer Ableitungen.

Von den ersten Eigenschaften differenzierbarer Abbildungen gehen wir hier auf verschiedene Versionen des Mittelwertsatzes und elementare Folgerungen ein.

(32.7) Satz: (Mittelwertsatz) Für eine differenzierbare Funktion $u : \Omega \to \mathbb{R}$ seien $a \in \Omega$ und $v \in E$ mit $a + tv \in \Omega$ für alle $t \in [0, 1]$. Dann gibt es $\xi \in [0, 1]$ mit

$$u(a + v) - u(a) = Du(a + \xi v).v\,.$$

Denn $\varphi(t) := u(a + tv)$ ist differenzierbar auf $[0, 1]$, so dass nach dem elementaren Mittelwertsatz 17.3 ein $\xi \in [0, 1]$ mit $\varphi(1) - \varphi(0) = \varphi'(\xi)$ existiert. Und $\varphi'(t) = Du(a + tv).v$.

(32.8) Satz: (Schrankensatz) *Für eine differenzierbare Abbildung $f : \Omega \to E$ folgt für $a \in \Omega$ und $v \in E$ mit $a + tv \in \Omega$ für alle $t \in [0,1]$ aus 27.19 die Abschätzung*

$$\|f(a+v) - f(a)\| \le \sup\{|||Df(a+tv)||| : t \in [0,1]\}\, \|v\| \ ,$$

und man findet zu jedem $a \in \Omega$ eine Umgebung $U \in \Omega$, so dass

$$\|f(x) - f(a)\| \le \sup\{|||Df(y)||| : y \in U\}\, \|x - a\|$$

für $x \in U$.

(32.9) Folgerung: *Eine differenzierbare Abbildung $f : \Omega \to F$ mit $Df = 0$ ist lokal konstant* [3]. *f ist konstant, wenn Ω zusammenhängend ist.*

Beweis: Zu $a \in \Omega$ ergibt sich $f(a) = f(x)$ für x in einer Umgebung $U \subset \Omega$ von a. Das bedeutet, dass f lokal konstant ist.

Sei wieder $a \in \Omega$. Die Menge $Z := \{x \in \Omega : f(x) = f(a)\}$ ist nicht leer, sie enthält ja a, und sie ist abgeschlossen, weil f stetig ist ($Z = f^{-1}(f(a))$). Z ist auch offen, weil mit jedem $b \in Z$ auch eine ganze Umgebung zu Z gehört, wie zu Beginn des Beweises gezeigt wird. ■

(32.10) Folgerung: Zwei differenzierbare Abbildungen $f, g : \Omega \to F$ auf einem Gebiet $\Omega \subset E$ unterscheiden sich nur durch eine Konstante, wenn $Df = Dg$ gilt.

(32.11) Folgerung: Eine differenzierbare Abbildung ist lokal lipschitzstetig, wenn die Ableitung Df lokal beschränkt ist.

(32.12) Folgerung: Also ist eine stetig differenzierbare Abbildung lokal lipschitzstetig. [10.7.07]

(32.13) Mittelwertsatz in Integralform: $f : \Omega \to F$ sei stetig differenzierbar mit Werten in einem Banachraum F. Für $a \in \Omega$ und $v \in E$ gelte $a + tv \in \Omega$ für alle $t \in [0,1]$. Dann ist

$$f(a+v) - f(a) = \int_0^1 Df(a+tv)\, dt.v \, .$$

Beweis: Nach dem Hauptsatz 27.18 (für den wir allerdings den Mittelwertsatz 27.19 gebraucht haben) gilt

$$f(a+v) - f(a) = \int_0^1 \frac{d}{dt} f(a+tv)\, dt = \int_0^1 Df(a+tv)\, dt.v \, .$$

■

Hier zeigt sich der Vorteil einer vektorwertigen Integrationstheorie. Aus 32.13 lassen sich die vorangehenden Mittelwertsätze im wesentlichen folgern. Der elegante Beweis von 32.13 benötigt in unserer Darstellung den vektorwertigen Hauptsatz 27.18 der Differential- und Integralrechnung. Zum Beweis von 27.18 haben wir die folgende Konsequenz aus dem vektorwertigen Mittelwertsatz 27.19 verwendet:

(32.14) Spezialfall von 32.9: *Eine stetig differenzierbare Funktion $H : [a,b] \to E$ mit $\dot{H} = 0$ ist konstant.*

---

[3] Eine Eigenschaft gilt in einem topologischen Raum X *lokal*, wenn es zu jedem Punkt aus X eine Umgebung $U \subset X$ gibt, so dass die Eigenschaft für U bzw. für alle $x \in U$ gilt. So bedeutet *lokal konstant* für eine Abbildung f auf X, dass es zu jedem Punkt aus X eine Umgebung gibt, so dass $f|_U$ konstant ist. Analog für lokal beschränkt. Eine Raum heißt lokal (kurven-) zusammenhängend, wenn es zu jedem Punkt aus X eine Umgebung gibt, die (kurven-) zusammenhängend ist.

Dieses Resultat kann ohne Rückgriff auf die Mittelwertsätze aus $\S\,32$ und $\S\,27$ bewiesen werden, z.B. ähnlich wie 27.19, aber elementarer oder ganz einfach mit Hilfe des Satzes von Hahn-Banach, aus dem folgt, dass es zu zwei $v, w \in E$ mit $v \neq w$ stets $\mu \in E'$ mit $\mu(v) \neq \mu(w)$ gibt: Wäre H nicht konstant, so gäbe es $t \in [a,b]$ und $\mu \in E'$, so dass für $\phi := \mu \circ H$ gilt: $\phi(a) \neq \phi(t)$. Aber aus $\dot{H} = 0$ folgt $\dot{\phi} = 0$ und daher ist ϕ nach dem Mittelwertsatz 17.3 konstant.

Jetzt kann der Hauptsatz mit Hilfe von 32.14 als bewiesen gelten. Es folgt 32.13 und daraus unmittelbar 32.8 für stetig differenzierbare f.

§33 Höhere Ableitungen

In diesem Paragrafen werden die partiellen Ableitungen höherer Ordnung eingeführt und der Satz von Taylor wird bewiesen.

Im Folgenden ist Ω eine offene Teilmenge aus dem normierten Raum $E = \mathbb{R}^n$ und F ist wieder ein normierter Raum (z.B. $F = \mathbb{R}^m$).

(33.1) Satz: (Satz von Schwarz)[4] *Es sei $f : \Omega \to F$ eine Abbildung auf einer offenen Menge $\Omega \subset \mathbb{R}^n$, für die die partiellen Ableitungen in i-Richtung und in k-Richtung existieren. Die partielle Ableitung $\partial_k u : \Omega \to F$ sei außerdem in i-Richtung partiell differenzierbar mit stetiger Ableitung $\partial_i \partial_k u : \Omega \to F$. Dann existiert auch $\partial_k \partial_i u$ und es gilt*

$$\partial_k \partial_i u = \partial_i \partial_k u\,.$$

Als wichtige Konsequenz ergibt sich: Ist u eine partiell differenzierbare Funktion und sind alle partiellen Ableitungen $\partial_j f$ stetig partiell differenzierbar, so gilt $\partial_k \partial_i u = \partial_i \partial_k u\,.$

(33.2) Beispiel: Die Funktion

$$u(x,y) := \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, & \text{wenn } (x,y) \neq 0 \text{ und} \\ 0 & \text{sonst.} \end{cases}$$

ist partiell differenzierbar, die partiellen Ableitungen sind wieder partiell differenzierbar, aber in 0 unterscheiden sich $\partial_1 \partial_2 u$ und $\partial_2 \partial_1 u$.

$$\partial_2 u(x,y) = u_y(x,y) = \frac{(x^3 - 3xy^2)(x^2+y^2) - (x^3 y - xy^3)2y}{(x^2+y^2)^2} = \frac{x^5 - 4x^3y^2 - xy^4}{(x^2+y^2)^2}$$

und $u_y(0,0) = 0$. Also

$$u_{yx}(0,0) = \partial_1 \partial_2 u(0,0) = \lim_{x \to 0} \frac{1}{x} u_y(x,0) = \lim_{x \to 0} \frac{1}{x} \frac{x^5}{x^4} = 1\,.$$

Analog (das Vertauschen von x und y erzeugt einen Faktor -1):

$$\partial_1 u(x,y) = u_x(x,y) = -\frac{y^5 - 4y^3x^2 - x^4y}{(x^2+y^2)^2}$$

und $u_x(0,0) = 0$. Also

$$u_{xy}(0,0) = \partial_2 \partial_1 u(0,0) = \lim_{y \to 0} \frac{1}{y} \left(-\frac{y^5}{y^4} \right) = -1\,.$$

[4] Vgl. 30.15 und die Anwendung in 30.14.

■

Die partiellen Ableitungen u_x, u_y sind stetig, also ist u sogar total differenzierbar und stetig differenzierbar. Die „zweiten" partiellen Ableitungen sind allerdings nicht stetig, was schon aus dem Satz von Schwarz folgt.

(33.3) Definition: Eine Abbildung $f : \Omega \to F$ heißt p*-fach* oder auch p*-mal partiell differenzierbar*, wenn für alle q-Tupel $(k_1, \dots, k_q)$, $q \le p$, $q, p \in \mathbb{N}$ von Indizes $k_j \in \{1, 2, \dots, n\}$ jeweils die partielle Ableitung

$$\partial_{k_q} \dots \partial_{k_1} f$$

existiert. Diese ist durch

$$\partial_{k_q} \dots \partial_{k_1} f := \partial_{k_q} \left(\partial_{k_{q-1}} \dots \partial_{k_1}\right) f = \frac{\partial}{\partial x^{k_q}} \partial_{k_{q-1}} \dots \partial_{k_1} f$$

rekursiv definiert. $\partial_{k_q} \dots \partial_{k_1} f$ heißt *partielle Ableitung der Ordnung* q. Der Vollständigkeit halber setzen wir $\partial_\emptyset f = f$, so dass jede Funktion 0-fach partiell differenzierbar ist.

f ist p*-fach stetig partiell differenzierbar*, wenn f p-fach partiell differenzierbar ist und alle partiellen Ableitungen $\partial_{k_q} \dots \partial_{k_1} f$ der Ordnung $q \le p$ stetig sind. Eine 0-fach stetig partiell differenzierbare Abbildung f ist demzufolge stetig.

Mit $\mathcal{C}^p(\Omega, F)$ wird der $\mathbb{R}$-Vektorraum der p-fach stetig partiell differenzierbaren Abbildungen $f : \Omega \to F$ bezeichnet (klar, dass es sich um einen Vektorraum handelt, nach der Produktregel ergibt sich für den Fall $F = \mathbb{R}$ sogar, dass $\mathcal{C}^p(\Omega, F)$ in natürlicher Weise eine $\mathbb{R}$-Algebra ist. Insbesondere ist $\mathcal{C}^0(\Omega, F) = \mathcal{C}(\Omega, F)$.

Ferner wird eine Abbildung f *beliebig oft* oder auch *unendlich oft stetig partiell differenzierbar* genannt, wenn sie für alle $p \in \mathbb{N}$ p-fach stetig partiell differenzierbar ist. Der entsprechende Funktionenraum ist $\mathcal{C}^\infty(\Omega, F)$.

Im Falle $F = \mathbb{R}$ (oder gelegentlich $F = \mathbb{C}$) schreibt man auch $\mathcal{C}^p(\Omega)$ statt $\mathcal{C}^p(\Omega, \mathbb{R})$ bzw. $\mathcal{C}^p(\Omega, \mathbb{C})$.

(33.4) Folgerung: Für eine p-fach stetig partiell differenzierbare Abbildung f ist nach dem Satz von Schwarz die Reihenfolge der Differentiation ohne Bedeutung: Für jede Permutation $k_{\pi(j)}$ der k_j in $(k_1, \dots, k_p)$ ist

$$\partial_{k_{\pi(q)}} \dots \partial_{k_{\pi(1)}} f = \partial_{k_q} \dots \partial_{k_1} f\,.$$

(33.5) Notation: Für $\alpha, \beta \in \mathbb{N}^n$ mit $\alpha = (\alpha_1, \dots, \alpha_n)$ und $\beta = (\beta_1, \dots, \beta_n)$ setzen wir:

$|\alpha| = \sum_{k=1}^n \alpha_j$ (Ordnung von α)

$\alpha \le \beta \;:\Longleftrightarrow\; \forall\, j \le n : \alpha_j \le \beta_j$

$\alpha\,! := \alpha_1!\alpha_2! \dots \alpha_n!$

$\binom{\alpha}{\beta} := \binom{\alpha_1}{\beta_1}\binom{\alpha_2}{\beta_2} \dots \binom{\alpha_n}{\beta_n}$

$x^\alpha := \prod_1^n (x^j)^{\alpha_j} = (x^1)^{\alpha_1}(x^2)^{\alpha_2} \dots (x^n)^{\alpha_n}$ für $x = (x^1, \dots, x^n)$ als n-Tupel oder als Satz von Unbestimmten.

$c_\alpha x^\alpha$ mit $c_\alpha \in F$ ist ein *Monom* und eine endliche Summe $P(x) = \sum_{\alpha \in A} c_\alpha x^\alpha$, $A \subset \mathbb{N}^n$, ist ein *Polynom*.

$\deg P := \sup\{k \mid \exists\, \alpha \in A : |\alpha| = k \,\&\, c_\alpha \neq 0\}$. $\deg 0 := -\infty$.

$\partial_j^q f := (\partial_j)^q f$, also $\partial_j^q f := \partial_j \dots \partial_j f$ q-fach.

$\partial_\alpha f := \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f = \partial^\alpha f$ für $f \in \mathcal{C}^p(\Omega, F)$ und $|\alpha| \le p$.

(33.6) Satz: Jedes Polynom $P = \sum_{\alpha \in A} c_\alpha x^\alpha$ ist beliebig oft stetig partiell differenzierbar. Es gilt:

1°
$$\partial_j x^\alpha = \alpha_j x^{\alpha - e_j} \,,$$
wenn $\alpha_j \geq 1$, und daher
$$\partial_j P = \sum_{\alpha \in A, \alpha_j \geq 1} \alpha_j c_\alpha x^{\alpha - e_j} \,.$$

2°
$$\partial^\beta P = \beta\,! \sum_{\alpha \geq \beta} \binom{\alpha}{\beta} c_\alpha x^{\alpha - \beta} \,.$$

(33.7) Definition: (Differentialpolynome) Zu einem Polynom $P = \sum_{\alpha \in A} c_\alpha x^\alpha \neq 0$ mit reellen Koeffizienten, d.h. $c_\alpha \in \mathbb{R}$ (gelegentlich auch in $\mathbb{C}$), wird
$$P(\partial) := \sum_{\alpha \in A} c_\alpha \partial^\alpha$$
gesetzt; das soll für Abbildungen $f \in \mathcal{C}^p(\Omega, F)$, $d := \deg P \leq p$, die folgende Ableitungsoperation bedeuten:
$$P(\partial) f := \sum_{\alpha \in A} c_\alpha \partial^\alpha f \,.$$
$P(\partial)$ ist dann wohldefiniert als $\mathbb{R}$-linearer Operator
$$P(\partial) : \mathcal{C}^p(\Omega, F) \longrightarrow \mathcal{C}^{p-d}(\Omega, F) \,,$$
und wird *Differentialpolynom* oder *Differentialoperator* genannt (genauer: $P(\partial)$ ist ein polynomoialer Differentialoperator der Ordnung d mit konstanten Koeffizienten). [13.7.07]

Ganz allgemein werden in der Mathematik und auch in der Physik viele dynamische Prozesse durch eine partielle Differentialgleichung der Form
$$P(\partial) f = g$$
beschrieben mit einem Polynom P. Dabei wird zu einer Vorgabe g und weiteren Bedingungen wie z.B. Anfangs- oder Randbedingungen eine Funktion f gesucht, die die Bedingungen erfüllt und der Differentialgleichung genügt.

(33.8) Beispiele: Die folgenden Differentialoperatoren sind typische Beispiele, die in Geometrie und Physik auftreten:

1° $P(x) = \sum_{k=1}^n x^k$, $P(\partial) = \partial_1 + \ldots + \partial_n$.

2° $P(x) = \langle x, x \rangle = \sum_1^n (x^k)^2$. $P(\partial) = \partial_1^2 + \ldots + \partial_n^2$. Das ist der *Laplace-Operator*, der auch mit Δ bezeichnet wird.
$$\Delta = \partial_1^2 + \ldots + \partial_n^2 = \frac{\partial^2}{\partial (x^1)^2} + \ldots + \frac{\partial^2}{\partial (x^n)^2} \,.$$

Die Lösungen $u : \Omega \to \mathbb{R}$ der homogenen Differentialgleichung
$$\Delta u = 0$$
heißen die *harmonischen Funktionen*. Die inhomogene Gleichung
$$\Delta f = g$$

wird auch die *Poisson-Gleichung* genannt.

3° $P(t,x) = \sum (x^k)^2 - t$ liefert den Operator $P(\partial) = \Delta - \frac{\partial}{\partial t}$, der die so genannte *Wärmeleitgleichung*

$$\left(\Delta - \frac{\partial}{\partial t}\right) u = g$$

beschreibt.

4° Analog führt $P(t,x) = \sum (x^k)^2 - t^2$ zum Wellenoperator

$$P(\partial) = \Delta - \frac{\partial^2}{\partial t^2} \,.$$

5° Und ebenso $P(t,x) = \sum (x^k)^2 + it$ zum Schrödinger-Operator

$$P(\partial) = \Delta + i\frac{\partial}{\partial t} \,.$$

(33.9) Leibnizregel: Für $f, g \in \mathcal{C}^p(\Omega)$ und $\alpha \in \mathbb{N}^n$ mit $|\alpha| \leq p$ gilt die Formel

$$\partial^\alpha (fg) = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \partial^\beta f \partial^{\alpha - \beta} g \,.$$

Sei ferner P ein Polynom vom Grad d in n Unbestimmten. Dann gilt im Falle $p \geq d$ die Formel

$$P(\partial)(fg) = \sum_{\alpha \in \mathbb{N}^n} \frac{1}{\alpha!} \big(P_\alpha(\partial) f\big) \partial^\alpha g \,,$$

wobei $P_\alpha := \partial^\alpha P$.

Zum Beweis der Taylorformel (33.12) ist das folgende Resultat von Nutzen, das durch Induktion bewiesen werden kann:

(33.10) Lemma:

$$\frac{1}{k!} \frac{d^k}{dt^k} f(a + tx)|_{t=\tau} = \frac{1}{k!} \sum_{1 \leq j_1, \ldots, j_k \leq n} \partial_{j_1} \ldots \partial_{j_k} f(a + \tau x) x^{j_1} \ldots x^{j_k} = \sum_{|\alpha| = k} \frac{\partial^\alpha f(a + \tau x)}{\alpha\,!} x^\alpha \,.$$

Die erste Gleichung folgt unmittelbar aus der Kettenregel, während die zweite Gleichung den Vergleich von zwei verschiedenen Aufzählungen bedeutet. Dieser Vergleich ist bereits im Fall $k = 2$ aufschlussreich:

$$\sum_{i,j} \partial_i \partial_j f(a + \tau x) x^i x^j = \sum_{j=1}^n \partial_j^2 f(a + \tau x)(x^j)^2 + 2 \sum_{i<j} \partial_i \partial_j f(a + \tau x) x^i x^j = 2 \sum_{|\alpha| = 2} \frac{\partial^\alpha f(a + \tau x)}{\alpha\,!} x^\alpha \,.$$

(33.11) Definition: (p-Jet) Es sei f p-mal stetig partiell differenzierbar auf Ω. Der p-*Jet* oder das p-*te Taylorpolynom von* f *bei* $a \in \Omega$ ist das Polynom

$$j_a^p f(x) := \sum_{|\alpha| \leq p} \frac{\partial^\alpha f(a)}{\alpha!} \, x^\alpha \,.$$

(33.12) Satz: (Satz von Taylor) *Sei* Ω *offene Teilmenge des* $\mathbb{R}^n$ *und* $f \in \mathcal{C}^{p+1}(\Omega, \mathbb{R})$. *Wenn für* $a \in \Omega$ *und* $x \in \mathbb{R}^n$ *das Geradenstück zwischen* a *und* $a + x$ *ganz in* Ω *enthalten ist, so gibt es ein* $\tau \in [0, 1]$ *mit*

$$f(a + x) = j_a^p f(x) + \sum_{|\alpha| = p+1} \frac{\partial^\alpha f(a + \tau x)}{\alpha\,!} x^\alpha \,.$$

Zusatz: Für das Restglied

$$R(x) = \sum_{|\alpha|=p+1} \frac{\partial^\alpha f(a+\tau x)}{\alpha\,!} x^\alpha$$

gilt

$$\lim \frac{1}{\|h\|^p} R(h) = 0\,.$$

Das folgt direkt aus dem Satz von Taylor 19.5 in einer Veränderlichen unter Anwendung von 33.10.

Für $p+1$-fach stetige partiell differenzierbare Abbildungen auf einer offenen Menge Ω des $\mathbb{R}^n$ mit Werten in einem Banachraum F gibt es ebenfalls eine Taylorformel bei der allerdings für das Restglied die Integralform vorzuziehen ist, ansonsten erhält man im allgemeinen nur Abschätzungen in Verallgemeinerung zum Schrankensatz 32.8.

Totale Differentiation höherer Ordnung:

Es gibt auch eine Theorie der höheren totalen Differentiation. Eine (total) differenzierbare Abbildung $f : \Omega \to F$ heißt demnach *zweifach (zweimal) differenzierbar*, wenn die Ableitung $Df : \Omega \to \mathcal{L}(E, F)$ differenzierbar ist, und man erhält die zweite Ableitung

$$D(Df) : \Omega \to \mathcal{L}(E, \mathcal{L}(E, F))\,.$$

(33.13) Satz: Für zweimal differenzierbare Abbildungen $f : \Omega \to F$ und $a \in \Omega$ gilt für alle $h, k \in E$:

$$D(Df)(a)(h)(k) = D(Df)(a)(k)(h)\,.$$

Diese Aussage lässt sich besser verstehen, wenn man die Isomorphie von $\mathcal{L}(E, \mathcal{L}(E, F))$ mit dem normierten Vektorraum der stetigen bilinearen Abbildungen $E \times E \to F$, den wir mir $\mathcal{L}_2(E, F)$ bezeichnen, in Rechnung stellt:

$$\iota : \mathcal{L}(E, \mathcal{L}(E, F)) \to \mathcal{L}_2(E, F)\,,\ A \mapsto ((v, w) \to (A(v))(w)\,,\ (v, w) \in E \times E\,,$$

ist Isomorphismus. Als bilineare Abbildung schreiben wir dann

$$D^2 f(a) := \iota(D(Df)(a))$$

für die zweite Ableitung, und die Aussage 33.13 bedeutet gerade, dass $D^2 f(a)$ symmetrisch ist:

$$D^2 f(a)(v, w) = D^2 f(a)(w, v)\,,\ (v, w) \in E \times E\,.$$

Analog für die totale Differentiation der Ordnung $p \in \mathbb{N}_3$.

§34 Das lokale Verhalten einer Funktion

In diesem kurzen Paragrafen wird der Satz von Taylor angewandt, um das lokale Verhalten einer zweimal stetig partiell differenzierbaren Funktion auf das des zugehörigen 2-Jets zurückzuführen und auf diese Weise hinreichende Kriterien vorzustellen, wann bei kritischen Punkten lokale Extrema vorliegen.

Es sei also $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}$ eine zweimal stetig partiell differenzierbare Funktion. Bei gilt dann nach dem Satz von Taylor

$$f(a+x) = f(a) + \langle \nabla f(a), x \rangle + \frac{1}{2} \sum_{1 \le j,k \le n} \partial_j \partial_k f(a) x^j x^k + R(x)$$

für $x \in B(0,r)$, wenn $B(a,r) \subset \Omega$. Dabei gilt $\|x\|^{-2} |R(x)| \to 0$ für $x \to 0$.

(34.1) Definition: Die Matrix

$$H_f(a) = H(a) = H := \big(\partial_j \partial_k f(a)\big)_{1 \le j,k \le n}$$

heißt die *Hessematrix* von f in a. Die Quadrik

$$x_{n+1} = f(a) + \langle \nabla f(a), x \rangle + \frac{1}{2} x^T H x =: q(x)$$

(also die Teilmenge $Q = \{(q(x), x_{n+1}) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$) heißt die *Schmiegequadrik* von f in a.

Das Ziel des Paragrafen ist es, das lokale Verhalten von f in einem kritischen Punkt $a \in \Omega$ durch das Verhalten der Schmiegequadrik von f in a zu beschreiben, und diese Vergleich dazu zu benutzen, Kriterien dafür zu erstellen, dass ein lokales Extremum vorliegt. Es wird also der Graph von f bei a mit der Schmiegequadrik verglichen.

(34.2) Beispiele:

1° $f(x,y) = x^2 + y^2$ hat die Hessematrix $H = \operatorname{diag}(2,2)$. Die Schmiegequadrik von f im Nullpunkt $a = 0 \in \mathbb{R}^2$ wird in diesem Falle durch die Funktion selbst gegeben, es ist das *(elliptische) Paraboloid*

$$z = x^2 + y^2 .$$

2° Ähnlich für $f(x,y) = x^2 - y^2$: $H = \operatorname{diag}(2,-2)$ mit Schmiegequadrik das *(hyperbolische) Paraboloid*

$$z = x^2 - y^2 .$$

Man stelle eine Skizze dieser beiden Fälle dar, wie auch von $f(x,y) = x^2$.

(34.3) Definition: Eine Funktion $f : \Omega \to \mathbb{R}$ heißt bei $a \in \Omega$ *positiv definit* bzw. *negativ definit* wenn es eine Umgebung $U \subset \Omega$ von a und eine Konstante $\eta > 0$ so gibt, dass

$$\forall\, h \in \mathbb{R}^n \backslash \{0\} :\ a + h \in U \implies f(a+h) - f(a) > \eta\, \|h\|^2$$

bzw.

$$\forall\, h \in \mathbb{R}^n \backslash \{0\} :\ a + h \in U \implies f(a+h) - f(a) < -\eta\, \|h\|^2 .$$

Ein Punkt $a \in \Omega$ heißt *Sattelpunkt* von f, wenn es eine orthogonale Zerlegung $R^n = V \oplus W$ in Untervektorräume V, W von $\mathbb{R}^n$ mit $V \neq 0 \neq W$ so gibt, dass $f|_{(a+V)\cap\Omega}$ positiv definit bei a und $f|_{(a+W)\cap\Omega}$ negativ definit bei a ist.

Unser Beispiel $34.2.1°$ ist positiv definit bei 0, das Beispiel $34.2.2°$ hat einen Sattelpunkt bei 0. $f(x,y) = x^2$ ist nirgends lokal positiv definit und hat keinen Sattelpunkt.

Wenn f bei a positiv definit ist, so hat f dort ein lokales Minimum. Wenn also f auch noch differenzierbar in Ω ist, so ist a ein kritischer Punkt: $\nabla f(a) = 0$. In diesem Falle ist die Taylorentwicklung bis zur Ordnung 2:

$$f(a+x) = f(a) + \frac{1}{2}x^T Hx + R(x)$$

und es gilt

$$\frac{R(x)}{\|x\|} \to 0\,,\ x \to 0\,.$$

(34.4) Bemerkung: Aus der Linearen Algebra ist bekannt, dass eine symmetrische $n \times n$-Matrix $H \in \mathbb{R}^{n\times n}$ positiv definit heißt, wenn sie

$$\forall\, x \in \mathbb{R}^n \backslash \{0\} : x^T Hx > 0$$

erfüllt. Eine solche Matrix hat lauter positive Eigenwerte und hat in geeigneten Koordinaten, die durch eine orthogonale Transformation aus den Standardkoordinaten hervorgehen, nach dem Spektralsatz Diagonalgestalt, also die Form $H = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ mit den Eigenwerten λ_j von H. Daher ist H genau dann positiv definit, wenn die Funktion $x \mapsto x^T Hx =: q(x)$ bei 0 im Sinne der Definition 34.3 bei 0 lokal positiv definit ist. Denn

$$q(x) = \sum_{j=1}^{n} \lambda_j (x^j)^2 > \lambda\,\|h\|^2$$

für $h \neq 0$, wenn $\lambda > 0$ kleiner als $\min\{\lambda_j : 1 \leq j \leq n\}$ gewählt wird.

Analog für negativ definit. [17.7.07]

(34.5) Satz: *Es sei $f : \Omega \to \mathbb{R}$ eine dreimal stetig partiell differenzierbare Funktion und a ein kritischer Punkt von f. Außerdem sei die Hessematrix $H = H_f(a)$ invertierbar ($\det H \neq 0$). Dann hat f bei a dasselbe lokale Verhalten wie die Hessematrix H bei 0 in folgendem Sinne:*

Ist H positiv (negativ) definit, so gilt dasselbe für f bei a, und hat die Hesseform $q(x) = x^T Hx$ den Sattelpunkt 0 (H ist echt indefinit), so auch f bei a mit derselben Zerlegung.

Beweis: Da $\det H \neq 0$ ist, hat $\mathbb{R}^n$ eine Zerlegung $\mathbb{R}^n = V \oplus W$ in orthogonale Untervektorräume, so dass $H|_V$ positiv definit und $H|_W$ negativ definit ist, wobei auch die Fälle $V = 0$ oder $W = 0$ möglich sind. Zu zeigen ist, dass dann auch $f|_{(a+V)\cap\Omega}$ bei a lokal positiv definit und $f|_{(a+W)\cap\Omega}$ negativ definit ist. Daraus ergeben sich sämtliche Behauptungen des Satzes.

Es gibt $\lambda > 0$, so dass $x^T Hx > \lambda\,\|x\|^2$ für alle $x \in V\backslash\{0\}$. Und es ist nach dem Satz von Taylor $f(a+x) = f(a) + \frac{1}{2}x^T Hx + R(x)$ für kleine $\|x\|$, wobei

$$\psi(x) := \frac{R(x)}{\|x\|^2} \to 0\,.$$

Zu $\eta < \frac{1}{2}\lambda$, $\eta > 0$, gibt es daher $\varepsilon > 0$, so dass $|\psi(x)| < \frac{1}{2}\lambda - \eta$ für alle $x \in V$, $\|x\| < \varepsilon$, gilt, und es folgt für alle $v \in V$, $\|v\| = 1$,

$$\frac{1}{2}v^T Hv + \psi(x) > \frac{1}{2}\lambda + \psi(x) > \eta\,.$$

Mit $x \in V$, $x \neq 0$, und $v = \frac{x}{\|x\|}$ ergibt sich schließlich

$$f(a+x) - f(a) = \frac{1}{2}x^T Hx + \|x\|^2\,\psi(x) > \eta\,\|x\|^2\ .$$

$f|_{(a+V)\cap\Omega}$ ist also lokal positiv definit bei 0. ∎

(34.6) Spezialfall n=2: Sei $\det H \neq 0$. Wir haben die folgenden Kriterien:

1° $\det H > 0 \implies f$ *hat bei* a *ein lokales Extremum.* Denn $\det H > 0$ bedeutet, dass die beiden Eigenwerte gleiches Vorzeichen haben. f hat also bei a eine lokales Extremum, und zwar ein Minimum, wenn H positiv definit ist, und ein Maximum, wenn H negativ definit ist.

2° $\det H < 0 \implies f$ *hat bei* a *einen Sattelpunkt.* Denn $\det H < 0$ bedeutet, dass die Eigenwerte verschiedene Vorzeichen haben und die Eigenraumzerlegung von $\mathbb{R}^2$ in zwei orthogonale eindimensionale Unterräume ist die Zerlegung, die zeigt, dass f lokal auf dem einen Unterraum positiv definit und auf dem anderen negativ definit ist.

Für $n > 2$ erhält man ähnliche Kriterien.

(34.7) Beispiele:

1° 34.2 liefert die direktesten Fälle.

2° $f(x,y) := \big(3x + 4y + \sin xy\big)\big((2x+y) - \cos x\,(1 - \cos x)\big)$ für $(x,y) \in \mathbb{R}^2$. 0 ist kritischer Punkt und $f(0) = 0$. Die Taylorentwicklung ist $f(x,y) = (3x+4y)(2x+y) + R(x,y)$, da die Terme $\sin xy$ und $1 - \cos x$ von der Ordnung 2 in 0 verschwinden. Die lineare Transformation $(x,y) \mapsto (3x+4y, 2x+y) = (x', y')$ führt zunächst zur quadratischen Form $x'y'$ und die weitere lineare Transformation $(x', y') \mapsto (x'+y', x'-y') = (\xi, \eta)$ zu $\frac{1}{4}(\xi^2 - \eta^2)$. f hat also in 0 einen Sattelpunkt.

3° *Affensattel*: $f(x,y) = 3x^2y - y^3$. Dann $\nabla f(0) = 0$ und $H_f(0) = 0$. In Polarkoordinaten $(x,y) = r(\cos t, \sin t) = re^{it}$ gilt $f = r^3 \sin 3t$ (f ist Imaginärteil von $z \mapsto z^3$). Also ist f für $r > 0$ im Bereich $0 < t < \frac{\pi}{3}$ positiv, im Bereich $\frac{\pi}{3} < t < 2\frac{\pi}{3}$ negativ und so weiter im Wechsel. Es liegt also kein Sattelpunkt vor.

Das letzte Beispiel zeigt, dass im Falle eine ausgearteten Hessematrix ($\det H = 0$) im allgemeinen keine Resultate wie in 34.5 zu erwarten sind.

§35 Funktionenfolgen. Vertauschung von Grenzprozessen

Zum Schluss des Kapitels und der Vorlesung MIIA wird ein fundamentales Hintergrundthema zusammengeführt und weiterentwickelt, das fast überall in der Analysis eine wichtige Rolle spielt. Es geht um die Vertauschung von Grenzprozessen, die sich beispielsweise durch Permanenzeigenschaften von konvergenten Funktionefolgen formulieren lässt.

Wohlbekannt:

(35.1) Satz: *Sei* (f_n) *eine Folge stetiger Abbildungen* $f_n : X \to Y$ *zwischen den metrischen Räumen* X, Y*, die gleichmäßig gegen eine Abbildung* $f : X \to Y$ *konvergiert. Dann ist* f *stetig.*

(35.2) Bemerkungen: 1° Die Stetigkeit bedeutet: $x_k \to x \Longrightarrow f(x_k) \to f(x)$, also

$$\lim_{k\to\infty} \lim_{n\to\infty} f_n(x_k) = f(x) = \lim_{n\to\infty} f_n(\lim_{k\to\infty} x_k)\,,$$

und daher

$$\lim_{k\to\infty} \lim_{n\to\infty} f_n(x_k) = \lim_{n\to\infty} \lim_{k\to\infty} f_n(x_k)\,.$$

Insofern handelt es sich bei dem Satz um Vertauschung von Grenzprozessen.

2° Es genügt, vorauszusetzen, dass (f_n) lokal gleichmäßig konvergiert, dass es also zu jedem Punkt $a \in X$ eine Umgebung U gibt, so dass $(f_n|_U)$ gleichmäßig gegen $f|_U$ konvergiert, oder auch dass (f_n) kompakt konvergiert, d.h. auf jeder kompakten Teilmenge $K \subset X$ gleichmäßig konvergiert.

(35.3) Folgerung: Für X kompakt und $F = Y$ ein Banachraum ist $\mathcal{C}(X, F)$ mit der Supremumsnorm wieder eine Banachraum.

(35.4) Satz: *Sei (f_n) eine Folge von stetigen Funktionen $f_n : [a, b] \to F$ mit Werten in dem Banachraum F, die gleichmäßig gegen $f : [a, b] \to F$ konvergieren möge. Dann gilt*

$$\int_a^b f(t)\,dt = \lim_{n\to\infty} \int_a^b f_n(t)\,dt\,.$$

Ein entsprechendes Resultat ist auch richtig für $d\mu$ mit μ von beschränkter Variation und für sprungstetige Funktionen $f_n \in \mathcal{S}([a,b]\,, F)$ anstelle von stetigen Funktionen, wenn nur $f_n \to f$ gleichmäßig, d.h. in der Supremumsnorm gilt.

Diese Aussagen gelten nach Definiton der entsprechenden Integrale. Mit einer Zerlegungsfolge (Z^k) des Intervalls $[a, b]$ mit $\Delta(Z^k) \to 0$ und Zwischenvektoren $\tau^n \in \mathbb{Z}_*^k$ gilt mit $S_n^k := S(f_n, Z^k, \tau^k)$ die Identität

$$\lim_{n\to\infty} \lim_{k\to\infty} S_n^k = \lim_{n\to\infty} \int_a^b f_n(t)\,dt = \int_a^b f\,dt = \lim_{k\to\infty} S(\lim_{n\to\infty} f_n, Z^k, \tau^k) = \lim_{k\to\infty} \lim_{n\to\infty} S_n^k$$

wegen

$$S(\lim_{n\to\infty} f_n, Z^k, \tau^k) = \sum_{l=1}^{m^k} \lim_{n\to\infty} f_n(\tau_l^k)(t_l^k - t_{l-1}^k) = \lim_{n\to\infty} S_n^k\,,$$

es handelt sich also wieder um eine Vertauschung von Grenzprozessen.

Nach der Stetigkeit und der Integrierbarkeit kommen wir jetzt zur Differenzierbarkeit und stellen ein neues Resultat dar.

(35.5) Satz: *Sei (f_n) eine Folge von stetig differenzierbaren Abbildungen $f_n : \Omega \to F$, wobei E ein normierter Raum und F ein Banachraum sei. Ist dann die Folge (f_n) punktweise konvergent gegen $f : \Omega \to F$ und die Folge (Df_n) der Ableitungen, $Df_n : \Omega \to \mathcal{L}(E, F)$, gleichmäßig konvergent gegen $L : \Omega \to \mathcal{L}(E, F)$, dann ist auch f stetig differenzierbar und es gilt $Df = L$.*

Kurz:

$$f_n \to f \;\&\; f_n' \to L \text{ glm. } \Longrightarrow f' = \lim f_n'\,.$$

(35.6) Bemerkungen:

1° Auch hier genügt es, vorauszusetzen, dass (Df_n) lokal gleichmäßig gegen L konvergiert.

2° Unter den angegebenen Voraussetzungen konvergiert dann auch (f_n) gleichmäßig gegen f.

3° Die gleichmäßige Konvergenz einer Funktionenfolge aus differenzierbaren Funktionen hat in der Regel keine differenzierbare Funktion als Grenzfunktion.

4° Der Satz entspricht einer weiteren Vertauschung von Grenzprozessen. Zum Beispiel im Falle einer Variablen:

$$\lim_{h\to 0} \lim_{n\to\infty} \frac{f_n(a+h) - f_n(a)}{h} = f'(a) = \lim_{n\to\infty} f_n'(a) = \lim_{n\to\infty} \lim_{h\to 0} \frac{f_n(a+h) - f_n(a)}{h}\,.$$

(35.7) Folgerung: (Differentiation unter dem Integral, Parameterintegral) *Sei* $g : [t_0, t_1] \times \Omega \to F$ *eine Familie von Abbildungen* $(g_t)_{t \in [t_0,t_1]}$ *auf einer offenen Menge* $\Omega \subset E$ *in einem normierten Raum* E *mit Werten in einem Banachraum* F*:* $g_t(x) := g(t, x)$*. Es sei* g *stetig und für jedes feste* $t \in [t_0, t_1] =: I$ *sei* $g_t : \Omega \to F$ *differenzierbar mit Ableitung* $D_x g(t, x) := Dg_t(x)$*. Sei außerdem noch* $D_x g : I \times \Omega \to F$ *stetig.*

Dann ist die Abbildung

$$f(x) := \int_I g(t, x)\, dt\,,\ x \in \Omega\,,$$

stetig differenzierbar und es gilt die Formel

$$Df(x) = \int_I D_x(t, x)\, dt\,.$$

Beweis: Sei (Z^n) eine Zerlegungsfolge von I mit $\Delta(Z^n) \to 0$, und sei $\tau^n \in \mathbb{Z}^n_*$. Setze

$$f_n(x) := S(g(\ ,x), Z^n, \tau^n) = \sum_{k=1}^{m^n} g(\tau^n_k, x)(t^n_k - t^n_{k-1})\,.$$

Dann gilt für jedes $x \in \Omega$:

$$f(x) = \int_I g(t, x)\, dt = \lim_{n\to\infty} S(g(\ ,x), Z^n, \tau^n) = \lim_{n\to\infty} f_n(x)$$

sowie

$$L(x) := \int_I D_x g(t, x)\, dt = \lim_{n\to\infty} S(D_x g(\ ,x), Z^n, \tau^n)\,.$$

Es ist

$$Df_n(x) = D(\sum_{k=1}^{m^n} g(\tau^n_k, x)(t^n_k - t^n_{k-1})) = \sum_{k=1}^{m^n} D_x g(\tau^n_k, x)(t^n_k - t^n_{k-1})\,.$$

Daher konvergiert $\big(Df_n(x)\big)$ gegen $L(x)$ für jedes $x \in \Omega$. Als Abschätzung hat man

$$|||Df_n(x) - L(x)||| \le \sup\{|||D_x(t, x)||| : t \in I\}\Delta(Z^n)\,,$$

und wegen der Stetigkeit von $D_x g$ und der Kompaktheit von I gibt es zu jedem $a \in \Omega$ eine Umgebung $U \subset \Omega$ mit

$$\sup\{|||D_x(t, x)||| : t \in I, x \in U\} =: M < \infty\,.$$

Also gilt

$$|||Df_n(x) - L(x)||| \le M\Delta(Z^n) \to 0$$

gleichmäßig in $x \in U$, das heißt $\big(Df_n|_U\big)$ konvergiert gleichmäßig gegen L. Daher folgt die Behauptung aus dem Satz 35.5. ■

Damit ist auch der Hilfssatz 30.18 bewiesen. [20.7.07]

(35.8) Beispiel: Als Gegenbeispiel zu dieser Aussage dient die folgende Funktion, die in den Ferienübungen behandelt wird:

$$g(t, x) := \begin{cases} \frac{tx^3}{(t^2+x^2)^2}, & \text{wenn } (t, x) \ne 0 \text{ und} \\ 0 & \text{sonst.} \end{cases}$$

Die Funktion $f(x) := \int_0^1 g(t, x)\, dt\,,\ x \in \mathbb{R}$, ist differenzierbar, aber $f'(0) \ne \int_0^1 g_x(t, 0)dt$.