Kapitel 7

Differentiation in normierten Räumen, insbesondere im $\mathbb{R}^n$

Die Differentialrechnung entstand im 17. Jahrhundert infolge von Fragen, die durch die Naturwissenschaften (insbesondere Physik und Astronomie) aufgeworfen wurden: Man interessierte sich bei der nichtgleichförmigen Bewegung eines Gegenstandes (Himmelskörpers) für die Größe und Richtung seiner momentanen Geschwindigkeit, für die Länge seiner Bahn etc. Aus der Optik kamen Fragen nach der Tangente an eine Kurve; denn für die Brechungsgesetze von Linsen benötigte man den Winkel zwischen dem Lichtstrahl und der Normalen der Linsenkontur. Die Normale steht aber senkrecht auf der Tangente.

Gottfried Wilhelm Leibniz
1646 - 1717

Isaac Newton
1643 - 1727

Innermathematisch betrachtet ergaben sich zwei Zugänge zur Differentialrechnung, nämlich einerseits über die Änderungsrate von Größen und zum anderen über die Verhältnisse infinitesimaler Größen. Die beiden großen Schöpfer der Differentialrechnung gingen unabhängig voneinander über jeweils einen dieser beiden Wege: Gottfried Wilhelm Leibniz, der Philosoph, Theologe, Diplomat, Rechtswissenschaftler und Autor der Monadologie, entwickelte die Theorie der Differentiale und ihrer Quotienten, Isaac Newton, der Physiker und Astronom, gelangte zur Ableitung über den Limes des Differenzenquotienten. Newton entwickelte eine Theorie der Fluxionen (das sind die Ableitungen), Leibniz schuf die heute noch verwendeten Bezeichnungen $\frac{dy}{dx}$ und $\int$ und arbeitete die Rechengesetze für die Ableitung heraus (Produktregel, Quotientenregel, Kettenregel).

7.1 Definitionen (Wiederholung der Schuldefinition) : Es sei D offen in $\mathbb{R}$. Eine Abbildung $f : D \to \mathbb{R}$ heißt **differenzierbar** bei $a \in D$, wenn $f'(a) \in \mathbb{R}$ so existiert, daß $\lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h)-f(a)}{h} = f'(a)$ gilt. $f'(a)$ heißt dann die **Ableitung** von f bei a.

7.2 Bemerkung : $l_a(h) := f'(a)h$ vermittelt eine lineare Abbildung von $\mathbb{R}$ nach $\mathbb{R}$. Ihr Graph ist die Gerade durch den Ursprung parallel zur Tangente an $f(x)$ im Punkte a. $l_a(h)$ erfüllt die Beziehung $\lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{|f(a+h)-f(a)-l_a(h)|}{|h|} = 0$ (*).
$l_a(h)$ ist stetig.
Wenn es umgekehrt eine lineare Abbildung $l(h) : \mathbb{R} \to \mathbb{R}$ gibt, die (*) erfüllt, so existiert die Abbildung $f'(a)$.

Die Definition der totalen Ableitung in normierten Räumen ist nach den Eigenschaften der Bemerkung 7.2 modelliert:

7.3 Definitionen : Notation: $\mathbb{K}$ sei entweder $\mathbb{R}$ oder $\mathbb{C}$. Es seien $(V, || \, ||)$, $(V', || \, ||')$ normierte $\mathbb{K}$-Vektorräume und D offen in V, geschrieben $D \Subset V$, und $a \in D$. $f : D \to V'$ heißt **differenzierbar** bei $a \in D$, wenn es ein stetiges $l_a \in \mathrm{Hom}_{\mathbb{K}}(V, V')$ so gibt, daß $\lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h)-f(a)-l_a(h)}{||h||} = 0$ (**). $l_a(h) =: Df_a$ heißt die **totale Ableitung** von f an der Stelle a oder das **totale Differential**.

7.4 Bemerkungen :

(i) Besonders wichtige Spezialfälle sind natürlich die Fälle, wenn V und V' Banachräume sind, insbesondere $\mathbb{R}^n$, $\mathbb{R}^m$. Auf letzteren Fall werden wir unsere Darstellung hier einschränken – mit der euklidischen Norm als zugrundeliegendem $\|\ \|$. Für den allgemeinen Fall der normierten Räume lassen sich jeweils analoge Ergebnisse erzielen.

(ii) $l \in \mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m) \to l$ ist stetig (bezüglich der l_2-Norm).
Denn: Wenn $M_{B,B'}(l) = (\alpha_{ij})\ , \quad \alpha_{ij} \in \mathbb{K}$, ist, so ist $p_i \circ M_{B,B'}(l(x)) = \eta_i = \sum_{j=1}^{m} \alpha_{ij}\xi_j$ stetig (nach 5.16 (vi)). Also ist l stetig.

(iii) Daher kann für die Definition der totalen Ableitung im $\mathbb{R}^n$ die Forderung nach der Stetigkeit von l_a fallengelassen werden.

7.5 Definitionen : Die **Landauschen Symbole Klein-o** und **Groß-O**

werden folgendermaßen definiert: Wenn für zwei Funktionen $f,\ g$ $\lim\limits_{x\to a} \frac{f(x)}{g(x)} = 0$ ist, schreiben wir $f(x) = \mathrm{o}(g(x))$ und sagen: „$f(x)$ ist ein Klein-o von $g(x)$ für $x \to a$" oder „$f(x)$ ist von niedrigerer Ordnung als $g(x)$ für $x \to a$". Diese Bezeichnungsweise werden wir gleich im Zusammenhang mit der totalen Ableitung gebrauchen.

Für Informatiker wichtig ist auch die Notation Groß-O: Wenn $\frac{f(x)}{g(x)}$ für die Annäherung $x \to a$ beschränkt ist, d.h. wenn $C \in \mathbb{R}$ existiert mit $|\frac{f(x)}{g(x)}| \le C$ für $|x - a| < \delta$, so schreiben wir $f(x) = O(g(x))$ und sagen: „$f(x)$ ist ein Groß-O von $g(x)$ für x gegen a" oder „$f(x)$ ist höchstens von der Ordnung von $g(x)$ für x gegen a".

Edmund Landau
1877 - 1938

Mittels Groß-O wird z.B. die **Zeitkomplexität** von Algorithmen (etwa Sortieralgorithmen) ausgedrückt: Sei $f(n)$ die Anzahl der im Algorithmus benötigten Operationen für eine Anzahl n von zu verarbeitenden (zu sortierenden) Objekten und sei $f(n) = \mathrm{O}(g(n))$. Dann heißt die Komplexität des Algorithmus

(i) **konstant**, wenn $g(n) = 1$ oder $=$ const. ,

(ii) **linear**, wenn $g(n) = n$,

(iii) **quadratisch**, wenn $g(n) = n^2$,

(iv) **exponentiell**, wenn $g(n) = a^n$,

(v) **logarithmisch**, wenn $g(n) = \ln n$,

(vi) **linear-logarithmisch**, wenn $g(n) = n \ln n$,

(vii) **polynomial**, wenn $g(n) = n^r$,

(viii) **superexponentiell**, wenn $g(n) = n^n$ ist.

Ein Algorithmus heißt **schnell**, wenn seine Komplexität höchstens $\mathrm{O}(\ln^r n)$ ist.

7.6 Bemerkungen :

(i) Der Sortieralgorithmus quicksort (in der ANSI-C-Standardbibliothek <stdlib.h> als Funktion qsort vorhanden) ist zwar im worst case ein Algorithmus von quadratischer Komplexität, aber im Durchschnitt immerhin linear-logarithmisch.

(ii) Wenn f in a stetig ist, so ist $f(x) = f(a) + \mathrm{o}(1)$ für $x \to a$ (wegen $\lim\limits_{x \to a}(f(x) - f(a)) = 0$), d.h. f ist für $x \to a$ durch die Konstante $f(a)$ von geringerer Ordnung als 1 approximierbar.

(iii) f ist genau dann bei a differenzierbar, wenn $f(x) = f(a) + l_a(x-a) + \mathrm{o}(||x-a||)$ ist, d.h. f affin-linear von geringerer Ordnung als $||x-a||$ für $x \to a$ approximierbar ist.
Dieser Sachverhalt ist auf nebenstehender Graphik für den Spezialfall $f : \mathbb{R} \to \mathbb{R}$ veranschaulicht: Approximation der Kurve durch die Tangente im Punkt $(a, f(a))$. Es ist $l_a(x-a) = f'(a)(x-a) = Df_a(x-a)$.

(iv) f ist genau dann bei a differenzierbar, wenn

$$f(x) = f(a) + Df_a(x-a) + r(x-a)||x-a|| \quad mit \quad \lim_{x \to a} r(x-a) = 0$$

gilt. Zur Veranschaulichung der totalen Ableitung halten wir fest: Für $f : \mathbb{R} \to \mathbb{R}$ ist $\xi_2 = f(a) + Df_a(x-a)$ die Gleichung der **Tangente** im E^2 an die **Kurve** $\xi_2 = f(x)$ im Punkte $(a, f(a))$. Für $f : \mathbb{R}^2 \to \mathbb{R}$ ist $\xi_3 = f(a) + Df_a(x-a)$ die Gleichung der **Tangentialebene** im E^3 an die **Fläche** $\xi_3 = f(x)$ in $(a, f(a))$; $a, x \in \mathbb{R}^2$. Für $f : \mathbb{R}^n \to \mathbb{R}$ ist $\xi_{n+1} = f(a) + Df_a(x-a)$ die Gleichung der **Tangentialhyperebene** im E^{n+1} an die **Hyperfläche**[1] $\xi_{n+1} = f(x)$ im Punkte $(a, f(a))$; $a, x \in \mathbb{R}^n$.

(v) Wenn Df_a existiert, so eindeutig.
Denn: Seien $l_a(h)$, $\bar{l}_a(h)$ zwei verschiedene lineare Abbildungen, die Definitionen 7.3 (**) erfüllen. Dann gilt $\lim\limits_{h\to 0} \frac{l_a(h)-\bar{l}_a(h)}{||h||} = 0$, also für $h = \lambda b$ mit $\lambda \in \mathbb{K}$, $b \in V$, $\lambda \neq 0$, $b \neq 0$ gilt $0 = \lim\limits_{\lambda\to 0} \frac{\lambda l_a(b)-\lambda\bar{l}_a(b)}{\lambda||b||} = \frac{1}{||b||}(l_a(b) - \bar{l}_a(b))$ und somit $l_a(b) = \bar{l}_a(b)$ für alle $b \in \mathbb{R}^n$. (Damit so geschlossen werden darf, wird D in 7.3 als offen vorausgesetzt.)

(vi) Wenn f eine konstante Funktion ist, d.h. wenn $f(x) = c$ für alle $x \in D$ gilt, so ist $Df_a = 0$ für alle $a \in D$.
Denn: $Df_a = 0$ erfüllt 7.3 (**).

(vii) Wenn f eine lineare Abbildung ist, so gilt $Df_a = f$.
Denn: $Df_a = f$ erfüllt 7.3 (**).

(viii) Wir betrachten die Funktionen $\sin z$ und $\cos z$ auf $\mathbb{C}$: $D(\sin z)_a(h) = h\cos a$, also $(\sin z)' = \cos z$, $D(\cos z)_a(h) = -h\sin a$, also $(\cos z)' = -\sin z$.
Denn: Mit Hilfe der Additionstheoreme und der Reihendarstellungen von sin und cos in 5.28 (iv) rechnet man leicht das Erfülltsein von 7.3 (**) nach.

(ix) Es gilt $D(z^n)_a(h) = na^{n-1}h$, also $(z^n)' = nz^{n-1}$.

(x) Es gilt $D(e^z)_a(h) = e^a h$, also $(e^z)' = e^z$.

7.7 Lemma : Sei $f : (V, ||\ ||) \to (V', ||\ ||')$ eine lineare Abbildung normierter Räume. Dann gilt: f ist genau dann stetig, wenn $\lambda \in \mathbb{R}$ so existiert, daß für alle $x \in V$ $||f(x)|| \leq \lambda ||x||$ gilt.

[1]Selbst der legendäre Perry Rhodan kennt den Begriff Hyperfläche!! Vgl. Perry Rhodan Lexikon, Arthur Moewig Verlag 1971, S. 109.

Slogan: Eine lineare Abbildung ist genau dann stetig, wenn sie beschränkt ist.

Beweis: Wenn die Bedingung erfüllt ist, so gilt $||f(x) - f(a)|| = ||f(x - a)|| \leq \lambda ||x - a||$, und somit ist f stetig in a.

Wenn umgekehrt f stetig ist, so existiert zu $\varepsilon = 1$ ein $\delta > 0$ so, daß $||x|| < \delta$ die Ungleichung $||f(x)|| < 1$ erfüllt. Setze $x := \frac{\delta y}{2||y||}$ für beliebiges $y \neq 0$. Dann gilt nämlich $||x|| = \frac{\delta}{2} < \delta$, also $||f(x)|| = \frac{\delta}{2||y||}||f(y)|| < 1$, d.h. $||f(y)|| < \frac{2}{\delta}||y||$. Wenn wir in der letzten Ungleichung $\leq$ statt $<$ schreiben, so gilt die Ungleichung für alle $y \in V$.

7.8 Satz (Kettenregel) : Es seien V, V', V'' normierte Vektorräume, $D \Subset V$, $f : D \to V'$ differenzierbar bei $a \in D$, $D' \Subset V'$ mit $f(D) \subset D'$, $g : D' \to V''$ differenzierbar bei $b := f(a)$. Dann ist $g \circ f$ bei a differenzierbar, und es gilt $D(g \circ f)_a = Dg_{f(a)} \circ Df_a$.

Beweis:

(i) Wir setzen $x := a + h$, $y := f(x)$, $h' := y - b$.

(ii) Da g als differenzierbar bei b vorausgesetzt ist, gilt
$g(y) = g(b) + Dg_b(y - b) + s(y - b)$ mit $s(y - b) = \mathrm{o}(||y - b||)$.

(iii) Aus der Differenzierbarkeit von f bei a folgt $y = b + Df_a(x-a) + r(x-a)$ mit $r(x - a) = \mathrm{o}(||x - a||)$.

(iv) Also gilt $g \circ f(x) = g \circ f(a) + Dg_b(Df_a(x - a) + r(x - a)) + s(y - b) = g \circ f(a) + Dg_b \circ Df_a(x - a) + Dg_b(r(x - a)) + s(y - b)$.

(v) Zu zeigen bleibt also: $Dg_b \circ r(x - a) + s(y - b) = \mathrm{o}(||x - a||)$.

(vi) Da Dg_b linear stetig ist, gilt nach 7.7 $||Dg_b \circ r(x - a)|| < \lambda ||r(x - a)||$, also $Dg_b \circ r(x - a) = \mathrm{o}(||x - a||)$.

(vii) Zu zeigen bleibt also noch: $s(y - b) = \mathrm{o}(||x - a||)$.

(viii) Nun gilt doch $\lim\limits_{x\to a} \frac{||y-b||}{||x-a||} = \lim\limits_{x\to a} \frac{||Df_a(x-a)+r(x-a)||}{||x-a||} \leq \lim\limits_{x\to a} \frac{||Df_a(x-a)||}{||x-a||} + 0 = \lambda$.

(ix) Daher gilt $\lim\limits_{x\to a} \frac{||s(y-b)||}{||x-a||} = \lim\limits_{x\to a} \frac{s(||y-b||)}{||y-b||} \lim\limits_{x\to a} \frac{||y-b||}{||x-a||} = 0 \cdot \lambda = 0$, q.e.d.

7.9 Satz : Sei $D \Subset \mathbb{K}^n$. Dann ist $f : D \to \mathbb{K}^m$ genau dann differenzierbar bei a, wenn für alle $i = 1, 2, \ldots, m$ die **i-te Koordinatenfunktion** $f^i :=$

$p_i \circ f$ differenzierbar bei a ist. (Vgl. die analoge Aussage 5.16 (v) über Stetigkeit.) Es gilt dann $p_i \circ Df_a = D(p_i \circ f)_a$, d.h. totale Ableitung und Projektion kommutieren. In anderer Schreibweise: $Df_a = (D(p_1 \circ f)_a, \ldots, D(p_m \circ f)_a)^t$.
Beweis: Sei f differenzierbar bei a. Da die Projektion p_i linear, also differenzierbar ist, ist nach 7.8 auch $p_i \circ f$ bei a differenzierbar.

Sei nun umgekehrt $p_i \circ f$ für alle i bei a differenzierbar. Dann gilt doch $0 = \lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{p_i\circ f(a+h) - p_i \circ f(a) - D(p_i\circ f)_a}{||h||} = \lim\limits_{\substack{h\to 0\\ h\neq 0}} p_i \circ \frac{f(a+h)-f(a)-l(h)}{||h||}$ für alle $i = 1, \ldots, m$, wobei $l(h) := (D(p_1 \circ f)_a(h), \ldots, D(p_m \circ f)_a(h))^t$ sei. Also ist $\lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h)-f(a)-l(h)}{||h||} = 0$ und somit $l = Df_a$. Aus 7.8 folgt die restliche Behauptung.

Analog zu den von der Schule her bekannten Rechenregeln läßt sich zeigen:

7.10 Satz : Sei V normierter $\mathbb{K}$-Vektorraum und $D \Subset V$. f, g seien bei $a \in D$ differenzierbare Abbildungen $D \to \mathbb{K}$. Dann sind $\lambda f + \mu g$ $(\lambda, \mu \in \mathbb{K})$, $f \cdot g$ sowie, im Falle $g(a) \neq 0$, auch $\frac{f}{g}$ bei a differenzierbar, und es gelten folgende Beziehungen:

(i) $D(\lambda f + \mu g)_a = \lambda Df_a + \mu Dg_a$ **Linearität**,

(ii) $D(f \cdot g)_a = g(a)Df_a + f(a)Dg_a$ **Produktregel**,

(iii) $D(\frac{f}{g})_a = \frac{g(a)Df_a - f(a)Dg_a}{g(a)^2}$ **Quotientenregel**.

7.11 Einige Ableitungsfunktionen : $(m = n = 1, \mathbb{K} = \mathbb{R})$
Da $x = \exp(\ln x)$ gilt, ist nach 7.8 $1 = (\exp(\ln x))' = (\ln x)' \exp(\ln x)$, also gilt

$$(\ln x)' = x^{-1} \quad .$$

(Genau genommen haben wir hier nur gezeigt: Wenn $\ln x$ differenzierbar ist, so ist die Ableitungsfunktion x^{-1}. Die Differenzierbarkeit von $\ln x$ werden wir später – in Kapitel 11 – unabhängig von diesem Resultat zeigen.).

Da $\ln a^x = x \ln a$ gilt, ist $\ln a = (\ln a^x)' = a^{-x}(a^x)'$, also gilt

$$(a^x)' = a^x \ln a \quad .$$

Da $x = \sin(\arcsin x)$ gilt, ist $1 = (\arcsin x)' \cos(\arcsin x)$, also gilt

$$(\arcsin x)' = (1 - x^2)^{-\frac{1}{2}} \quad .$$

Ebenso läßt sich zeigen:

$(\arccos x)' = -(1-x^2)^{-\frac{1}{2}}$,

$(\arctan x)' = (1-x^2)^{-1}$, $(\operatorname{arccot} x)' = -(1+x^2)^{-1}$,

$(\sinh x)' = \cosh x$, $(\cosh x)' = \sinh x$,

$(\tanh x)' = \cosh^{-2} x$, $(\coth x)' = -\sinh^{-2}$,

$(\operatorname{arsinh} x)' = (x^2+1)^{-\frac{1}{2}}$, $(\operatorname{arcosh} x)' = \pm(x^2-1)^{-\frac{1}{2}}$,

$(\operatorname{artanh} x)' = (1-x^2)^{-1}$, $(\operatorname{arcoth} x)' = (1-x^2)^{-1}$.

Wir lernen jetzt für **Skalarfelder** (d.h. für reellwertige Funktionen) eine weitere Art von Ableitungen kennen:

7.12 Definitionen : Seien $D \Subset \mathbb{R}$, $f : D \to \mathbb{R}$, $a = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t \in D$ und es sei $D_i f(a) := \lim\limits_{\substack{\xi \to 0 \\ \xi \neq 0}} \frac{f(\alpha_1, \alpha_2, \ldots, \alpha_i + \xi, \ldots, \alpha_n) - f(\alpha_1, \ldots, \alpha_n)}{\xi}$ existent. Dann heißt $D_i f(a) =: f_{\xi_i}|_{x=a} =: \frac{\partial f}{\partial \xi_i}(a)$ die **i-te partielle Ableitung** von $f(x) = f(\xi_1, \xi_2, \ldots, \xi_n)$ **an der Stelle** $x = a$.

Wenn $D_i f(a)$ für alle $a \in D$ existiert, so ist $D_i f : a \to \mathbb{R}$ eine Abbildung von D nach $\mathbb{R}$, und es kann möglicherweise die **partielle Ableitung zweiter Ordnung** $D_{i,j} f(a) := D_j(D_i f)(a)$ gebildet werden. Analog werden **partiellen Ableitungen q-ter Ordnung** $D_{i_1, i_2, \ldots, i_q} f(a)$ definiert. Wenn alle partiellen Ableitungen q-ter Ordnung auf einer nichtleeren offenen Menge D existieren und stetig sind, so heißt f eine **$\mathbf{C^{(q)}}$-Funktion**, geschrieben $f \in C^{(q)}$.

7.13 Bemerkungen :

(i) Die i-te partielle Ableitung ist die gewöhnliche Ableitung (Def.7.1) einer modifizierten Funktion $g : \mathbb{R} \to \mathbb{R}$, nämlich von $g(x) := f(\alpha_1, \ldots, \alpha_{i-1}, x, \alpha_{i+1}, \ldots, \alpha_n)$. Dann gilt nämlich $D_i f(a) = g'(x)|_{x=\alpha_i}$. $D_i f(a)$ ist also die Steigung der Tangente in i-Richtung an die Hyperfläche $\xi_{n+1} = f(\xi_1, \xi_2, \ldots, \xi_n)$ im Punkte a.

(ii) Zur Veranschaulichung der Vorgänge sei zunächst noch ausführlicher auf die graphische und arithmetische Darstellung von Funktionen von $\mathbb{R}^n$ nach $\mathbb{R}^m$ eingegangen:

7.14 Satz : Wenn $D \Subset V$ und $f : D \to V'$ differenzierbar bei a ist, so ist f stetig bei a.

Beweis: f ist genau dann differenzierbar bei a, wenn ein stetiges $l_a(h) \in \mathrm{Hom}(V, V')$ existiert mit $\lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{f(a+h)-f(a)-l_a(h)}{||h||} = 0$.

Also existiert ein $\delta' > 0$ so, daß $||f(a+h)-f(a)-l_a(h)|| < ||h||$ gilt, sofern $||h|| < \delta'$ ist. Die Linearität von $l_a(h)$ impliziert wegen 7.7 $||l_a(h)|| \leq \lambda||h||$.

Daher gilt für $||h|| < \delta'$: $||f(a+h) - f(a)|| \leq ||f(a+h) - f(a) - l_a(h)|| + ||l_a(h)|| < ||h|| + \lambda||h|| = \mu||h||$.

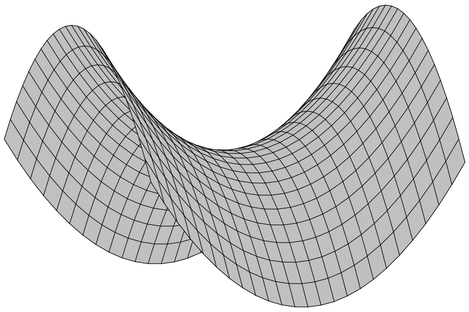
Setze $\delta := \min\{\delta', \frac{\varepsilon}{\mu}\}$. Dann ist $||f(a+h) - f(a)|| < \varepsilon$, wenn $||h|| < 0$. Also ist f bei a stetig.

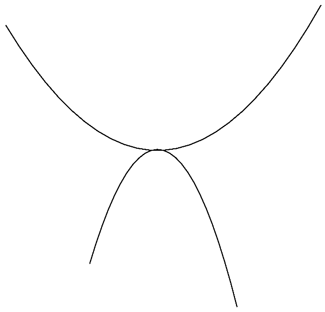
7.15 Definitionen : Sei $D \Subset V$, $f : D \to \mathbb{R}$ und $a \in D$ so, daß $f(a) \geq f(x)$ für alle $x \in K_\varepsilon(a) \subset D$. Dann sagen wir: f hat in a ein **lokales** oder **relatives Maximum** (analog: **lokales Minimum**). Relative Maxima oder Minima heißen **relative Extrema**.

7.16 Satz (notwendige Bedingung für Extrema) : Sei $D \Subset \mathbb{R}^n$, $f : D \to \mathbb{R}$. Wenn f in $a \in D$ ein relatives Extremum hat und wenn $D_i f(a)$ existiert, so ist $D_i f(a) = 0$.
Beweis: Wegen 7.13 (i) ist dies nichts anderes als die von der Schule her bekannte notwendige Bedingung für relative Extrema einer reellwertigen Funktion einer reellen Veränderlichen.

7.17 Bemerkungen :

(i) $D_i f(a) = 0$ ist nur eine notwendige Bedingung für ein relatives Extremum.
Denn: z.B. $\zeta = \xi^2 - \eta^2$ liefert $D_1 f(\xi, \eta) = 2\xi$, $D_2 f(\xi, \eta) = -2\eta$. Also in $(0, 0)$ ist $D_1 f(0, 0) = D_2 f(0.0) = 0$. Aber $(0, 0)$ ist kein relatives Extremum. (Hier: in ξ-Richtung Minimum, in η-Richtung Maximum.) Dies ist ein sog. **Sattelpunkt**.





(ii) Wenn $D \Subset \mathbb{R}^n$ kompakt ist, so nimmt nach 5.19 jede stetige Funktion $f : D \to \mathbb{R}$ ihr absolutes Maximum bzw. Minimum an. Liegt dieses Extremum aber auf dem Rand von D, so ist dort **nicht** notwendig $D_i f = 0$.
Beispiel: $f = \mathrm{id} : [0,1] \to \mathbb{R}$ hat absolutes Maximum bei $x = 1$, aber $f^{'}(1) = 1$.

7.18 Satz (Mittelwertsatz für reellwertige Funktionen einer reellen Veränderlichen) : Sei $f : [a,b] \to \mathbb{R}$ stetig und $f^{'}$ existiere für jedes $x \in]a,b[$. Dann existiert $c \in]a,b[$ mit $f(b) - f(a) = f^{'}(c)(b-a)$.
Beweis: Setze $m := \frac{f(b)-f(a)}{b-a}$ und $F(x) := f(b) - f(x) - m(b-x)$. $F(x)$ ist stetig auf dem kompakten Intervall $[a,b]$, also hat F nach 5.19 ein Maximum x_1 und ein Minimum x_2 auf $[a,b]$. Es gilt $F(a) = F(b) = 0$.
1. Fall: $F(x_1) > 0$. Dann ist $x_1 \in]a,b[$, also nach 7.16 $F^{'}(x_1) = 0$, d.h. $f^{'}(x_1) = m$. Setze $c := x_1$.
2. Fall: $F(x_2) < 0$. Schließe analog zum 1. Fall. Nimm $c := x_2$.
3. Fall: $F(x_1) \leq 0 \wedge F(x_2 \geq 0)$. Dann ist $F(x_1) = F(x_2) = 0$, also $F|_{[a,b]} \equiv 0$, also $F^{'}(x) = 0$ für $x \in]a,b[$, also $f^{'}(x) = m$ für alle $x \in]a,b[$. Setze $c := x$. □

7.19 Bemerkungen :

(i) Der Spezialfall $f(a) = f(b)$ in 7.18 heißt **Satz von Rolle**.

(ii) Sei $f : [a,b] \to \mathbb{R}$ stetig und in $]a,b[$ überall differenzierbar. Sei für zwei feste $m, M \in \mathbb{R}$ $m \leq f^{'}(x) \leq M$ für alle $x \in]a,b[$. Dann gilt $m(y-x) \leq f(y) - f(x) \leq M(y-x)$ für alle x, y mit $a \leq x \leq y \leq b$.
Klar, wegen 7.18.

(iii) Kennzeichnung der konstanten Funktionen.
Wenn $f : [a,b] \to \mathbb{R}$ stetig und in $]a,b[$ differenzierbar ist und wenn $\bigwedge_{x\in]a,b[} f'(x) = 0$ ist, so ist f konstant.
Folgt aus (ii). Setze $m = M = 0$!

(iv) Sei $c \in \mathbb{R}$ und sei $f : \mathbb{R} \to \mathbb{R}$ differenzierbar mit $f'(x) = cf(x)$ für alle $x \in \mathbb{R}$. Dann ist $f(x) = f(0)e^{cx}$ für alle $x \in \mathbb{R}$.
Denn: Setze $F(x) := f(x)e^{-cx}$. Dann ist $F'(x) = f'(x)e^{-cx} - cf(x)e^{-cx} = e^{-cx}(f'(x) - cf(x)) = 0$. Also ist nach (iii) $F(x) = \text{const}$. Da $F(0) = f(0)$ ist, gilt $F(x) = f(0)$ für alle x, also $f(x) = f(0)e^{cx}$ für alle $x \in \mathbb{R}$.

(v) Die Differentialgleichung von (iv) ist ein Modell für gewisse Wachstumsprozesse: Sei z.B. $f(x)$ die Größe einer Population zum Zeitpunkt x. $\frac{f'(x)}{f(x)}$ wird auch als **Wachstumsrate** zum Zeitpunkt x bezeichnet. Man nimmt gern eine konstante Wachstumsrate an (z.B. Wirtschaftswachstum!), z.B. $c\%$ pro Jahr. Also nach (iv) $f(x) = f(0)e^{\frac{cx}{100}}$.
Gesucht ist die sog. **Verdoppelungszeit**. Zur Zeit $x = 0$ sei die Größe der Population $f(0)$. Für welches x ist $f(x) = 2f(0)$?
$2f(0) = f(0)e^{\frac{cx}{100}}$. $cx = 100\ln 2 \approx 69{,}31$. Also z.B. bei 7% Wachstum erfolgt eine Verdoppelung in 10 Jahren!

(vi) e^x ist diejenige Funktion $f : \mathbb{R} \to \mathbb{R}$, für die $f'(x) = f(x)$ und $f(0) = 1$ gilt.

(vii) $f : [a,b] \to \mathbb{R}$ stetig, in $]a,b[$ differenzierbar, $\bigwedge_{x\in]a,b[} f'(x) > 0$ (bzw. ≥ 0 bzw. < 0 bzw. ≤ 0) $\to$ f ist in $[a,b]$ streng monoton wachsend (bzw. monoton wachsend bzw. streng monoton fallend bzw. monoton fallend).
Denn: Sei $f'(x) > 0$. Annahme: f **nicht** streng monoton wachsend, d.h. $\exists x_1 < x_2 : f(x_1) \geq f(x_2) \to \exists c \in]a,b[: f'(c) = \frac{f(x_2)-f(x_1)}{x_2-x_1} \leq 0$ (wegen 7.18).

(viii) $f :]a,b[\to \mathbb{R}$ differenzierbar, f zweimal differenzierbar in $x_1 \in]a,b[$, $f'(x_1) = 0$, $f''(x_1) > 0$ (bzw. $f''(x_1) < 0$) $\to$ f hat in x_1 ein isoliertes lokales Minimum (bzw. Maximum).
Denn: Sei $f''(x_1) > 0$, $f''(x_1) = \lim_{x\to x_1} \frac{f'(x)-f'(x_1)}{x-x_1} > 0$. Daraus folgt $\exists \delta > 0 : \{0 < |x - x_1| < \delta \to \frac{f'(x)-f'(x_1)}{x-x_1} > 0\}$, also $\frac{f'(x)}{x-x_1} > 0$ für $|x - x_1| < \delta$.
Nach (vii) ist also $f(x)$ streng monoton wachsend in $]x_1, x_1+\delta[$. Ebenso

folgt: $f(x)$ ist streng monoton fallend in $]x_1 - \delta, x_1[$, also hat f in x_1 ein isoliertes lokales Minimum. $\square$

7.20 Satz : Sei $D \overset{\circ}{\subset} \mathbb{R}^n$, $f : D \to \mathbb{R}$ eine $C^{(k)}$-Funktion. Dann ist $D_{i_k \ldots i_2 i_1} f = D_{i_{\pi(k)} \ldots i_{\pi(2)} i_{\pi(1)}} f$ für jede Permutation $\pi \in S_k$.
Beweis: Mit Hilfe des Mittelwertsatzes. Siehe O. Forster: Analysis 2, p.41.

7.21 Satz : Sei $D \overset{\circ}{\subset} \mathbb{R}^n$ und $a \in D$. Dann gelten:

(i) Wenn $f : D \to \mathbb{R}$ differenzierbar bei a ist, so existiert $D_i f(a)$ für alle i; genauer: es ist $D_i f(a) = Df_a(e_i)$ (e_i ist i-ter Standardvektor von $\mathbb{R}^n$).
Beweis: Wenn f differenzierbar bei a ist, so gilt $\lim\limits_{\substack{h \to 0 \\ h \neq 0}} \frac{f(a+h) - f(a) - Df_a(h)}{||h||} = 0$, also speziell für $h = \lambda e_i$ $\lim\limits_{\substack{\lambda \to 0 \\ \lambda \neq 0}} \frac{f(\alpha_1, \ldots, \alpha_{i-1}, \alpha_i + \lambda, \alpha_{i+1}, \ldots, \alpha_n) - f(\alpha_1, \ldots, \alpha_n)}{\lambda ||e_i||} = \lim\limits_{\substack{\lambda \to 0 \\ \lambda \neq 0}} \frac{\lambda Df_a(e_i)}{\lambda ||e_i||} = Df_a(e_i)$ und der erste Limes ist doch $D_i f(a)$. Umgekehrt:

(ii) Wenn $f : D \to \mathbb{R}$ von der Klasse $C^{(1)}$ ist. so ist f differenzierbar bei a; genauer: Es gilt $Df_a(x) = \sum\limits_{i=1}^{n} D_i f(a) \xi_i$, wobei $x = (\xi_1, \ldots, \xi_n)^t$ ist.
Beweis: Beweis mit Hilfe des Mittelwertsatzes. Siehe O. Forster: Analysis 2, p.47f.

7.22 Folgerungen :

(i) $f : D \to \mathbb{R}^m$ differenzierbar bei $a \in D \overset{\circ}{\subset} \mathbb{R}^n \to \bigwedge\limits_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} D_i f^j(a)$ existiert.

(ii) $\bigwedge\limits_{j=1,\ldots,m} f^j \epsilon C^{(1)}$ in $D \overset{\circ}{\subset} \mathbb{R}^n \to Df_a$ existiert für alle $a \in D$, nach 7.9.

(iii) $f \epsilon C^{(1)}$ in $D \overset{\circ}{\subset} \mathbb{R}^n \to f \epsilon C^{(0)}$ in D.

7.23 Satz : $D \overset{\circ}{\subset} \mathbb{R}^n$, $f : D \to \mathbb{R}^m C^{(1)} \to Df_a(x) = f'(a)x$ für alle $a \in D$, wobei $f'(a) := \begin{pmatrix} D_1 f^1(a) & \ldots & D_n f^1(a) \\ \ldots & & \ldots \\ D_1 f^m(a) & \ldots & D_n f^m(a) \end{pmatrix} \in M_{m,n}(\mathbb{R})$ die sog. **Jacobi-Matrix** oder auch **Funktionalmatrix** ist.

7.24 Satz (Kettenregel für partielle Ableitungen) : Es seien $D \overset{\circ}{\subset} \mathbb{R}^n$, $a \in D$; $g_1, \ldots, g_m : D \to \mathbb{R}$ $C^{(1)}$-Funktionen; $U \overset{\circ}{\subset} \mathbb{R}^m$, $(g_1(a), \ldots,$

$g_m(a)) \in U$; $f : U \to \mathbb{R}$ $C^{(1)}$-Funktion, und $F : D \to \mathbb{R}$ sei durch $F(x) := f(g_1(x), \ldots, g_m(x))$ definiert.
Dann ist $D_i F(a) = \sum D_j f(g_1(a), \ldots, g_m(a)) D_i g_j(a)$.
Beweis: Mit $g := (g_1, \ldots, g_m)$ ist $F = f \circ g$, also $DF_a(x) = (Df_{g(a)} \circ Dg_a)(x) = f'(g(a))g'(a)x$ wegen 7.8 und 7.23
$= \begin{pmatrix} D_1 f(g(a)) & \ldots & D_m f(g(a)) \end{pmatrix} \begin{pmatrix} D_1 g_1(a) & \ldots & D_n g_1(a) \\ \ldots & & \ldots \\ D_1 g_m(a) & \ldots & D_n g_m(a) \end{pmatrix} \begin{pmatrix} \xi_1 \\ \ldots \\ \xi_n \end{pmatrix}$.
Nach 7.21 (ii) gilt $DF_a(x) = \begin{pmatrix} D_1 F(a) & \ldots & D_n F(a) \end{pmatrix} \begin{pmatrix} \xi_1 \\ \ldots \\ \xi_n \end{pmatrix}$ und somit die Behauptung. □

7.25 Anwendungen :

(i) **Umrechnen** der partiellen Ableitung **in andere Koordinaten**, z.B. in Polarkoordinaten: $f(\xi, \eta) = f(\rho \cos\theta, \rho \sin\theta) =: F(\rho, \theta)$.

$$\frac{\partial F}{\partial \rho} = \frac{\partial f}{\partial \xi}\frac{\partial \xi}{\partial \rho} + \frac{\partial f}{\partial \eta}\frac{\partial \eta}{\partial \rho} = \cos\theta \frac{\partial f}{\partial \xi} + \sin\theta \frac{\partial f}{\partial \eta}$$
$$\frac{\partial F}{\partial \theta} = -\rho \sin\theta \frac{\partial f}{\partial \xi} + \rho \cos\theta \frac{\partial f}{\partial \eta}$$

(ii) **Implizites Differenzieren**
Gesucht seien die partiellen Ableitungen von $\zeta = f(\xi, \eta)$.
Sei $g(\xi, \eta, \zeta) = 0$ die implizite Form von $\zeta = f(\xi, \eta)$, also $g(\xi, \eta, f(\xi, \eta)) = F(\xi, \eta) = 0$. Dann gilt

$$0 = \frac{\partial F}{\partial \xi} = \frac{\partial g}{\partial \xi}\frac{\partial \xi}{\partial \xi} + \frac{\partial g}{\partial \eta}\frac{\partial \eta}{\partial \xi} + \frac{\partial g}{\partial \zeta}\frac{\partial \zeta}{\partial \xi} = D_1 g + D_3 g \frac{\partial f}{\partial \xi} \quad ,$$

also $\frac{\partial f}{\partial \xi} = -\frac{D_1 g}{D_3 g}$ (sofern $D_3 g \neq 0$ ist). Entsprechend $\frac{\partial f}{\partial \eta} = -\frac{D_2 g}{D_3 g}$.
Ebenso: Wenn $\eta = f(\xi)$ explizit und $g(\xi, \eta) = 0$ die implizite Form ist, so ist $\eta' = -\frac{D?}{D_2 ?}$.
Z.B. $\eta = \sqrt{\frac{c}{\sin\xi}}$, $g(\xi, \eta) = \eta^2 \sin\xi - c = 0$, $\eta' = \frac{-\eta^2 \cos\xi}{2\eta \sin\xi} = \frac{-\eta}{2\tan\xi}$, $\eta' = \frac{-\sqrt{c}}{2\sqrt{\sin\xi}\tan\xi}$.

(iii) **Fehlerabschätzung**
$D \Subset \mathbb{R}^n$, $f : D \to \mathbb{R}$,
$x = (\xi_1, \ldots, \xi_n) \in D$ der wahre Wert,
$x_0 = (\xi_{01}, \ldots, \xi_{0n}) \in D$ der (gemessene) Näherungswert,

$h = x - x_0$ der Meßfehler,
$|\Delta\xi_i| = |\xi_i - \xi_{0i}| \le \sigma_i$ die Schranken für den Meßfehler.
Es ist $|\Delta f| = |f(x) - f(x_0)| = |Df_{x_0}(x - x_0) + \mathrm{o}(||x - x_0||)| \le$ $|\sum D_i f(x_0)(\xi_i - \xi_{0i})| + |\,\mathrm{o}(||x - x_0||)|$, also ist $S \approx \sum |D_i f(x_0)|\sigma_i$ eine ungefähre Fehlerschranke, wenn $|\Delta\xi_i| \le \sigma_i$ ist.

7.26 Definitionen : Es seien $x, y \in \mathbb{R}^n$.

(i) $\langle x, y\rangle := \sum_{i=1}^{n} \xi_i \eta_i$ heißt **Standardbilinearform** auf $\mathbb{R}^n$ (manchmal auch **inneres Produkt**).

(ii) x heißt **orthogonal** zu y, geschrieben $x \perp y$, wenn $\langle x, y\rangle = 0$ ist.

(iii) $\sphericalangle(x, y) := \arccos \frac{\langle x,y\rangle}{||x||_2||y||_2}$ heißt der **Winkel** zwischen x und y.

(iv) Der Operator **Nabla**, geschrieben $\bigtriangledown$, wird definiert als $\bigtriangledown := (D_1, \ldots, D_n)^t$.

(v) $\bigtriangledown f := \begin{pmatrix} D_1 f \\ \vdots \\ D_n f \end{pmatrix} =: \operatorname{grad} f$ heißt der **Gradient** von $f : D \to \mathbb{R}$ $(D \Subset \mathbb{R}^n)$.

(vi) f heißt eine **Potentialfunktion** des Vektorfeldes $\bigtriangledown f : D \to \mathbb{R}^n$.

(vii) Sei $g : D \to \mathbb{R}^n$. $\langle \bigtriangledown, g\rangle = \sum_{i=1}^{n} D_i g^i =: \operatorname{div} g$ heißt **Divergenz** von g.

(viii) $\bigtriangledown \times g =: \operatorname{rot} g$ heißt **Rotation** von g. Aber das „Kreuzprodukt" $\times$ behandeln wir erst in Mathematik III !

7.27 Bemerkungen :

(i) Wegen 5.5 liegt $\frac{\langle x,y\rangle}{||x||_2||y||_2}$ im Definitionsbereich des arccos. Die Relation $\langle x, y\rangle = ||x||_2||y||_2 \cos \sphericalangle(x, y)$ wird oft benutzt.

(ii) Wenn $f \in C^{(1)}$, so ist $Df_a(x) = \langle \bigtriangledown f(a), x\rangle$ (**Zuwuchsformel**).

(iii) Es sei $\alpha(t) : \mathbb{R} \to D \Subset \mathbb{R}^n$ differenzierbar in $t_0 \in \mathbb{R}$ und es sei $f : D \to \mathbb{R}$ differenzierbar. Dann ist $(f \circ \alpha)'(t_0) = \langle \bigtriangledown f(\alpha_0), \alpha'(t_0)\rangle$, wobei $\alpha_0 := \alpha(t_0)$ und $\alpha'(t_0) := (\alpha^{1'}(t_0), \ldots, \alpha^{n'}(t_0))^t$ ist.
Denn:
$(f \circ \alpha)'(t_0) \cdot h = D(f \circ \alpha)_{t_0}(h) = Df_{\alpha(t_0)}(D\alpha_{t_0}(h)) = \langle \bigtriangledown f(\alpha_0), D\alpha_{t_0}(h)\rangle$.

Nun ist $D\alpha_{t_0}(h) = \begin{pmatrix} D\alpha^1_{t_0}(h) \\ \dots \\ D\alpha^n_{t_0}(h) \end{pmatrix} = \begin{pmatrix} \alpha^{1\prime}(t_0)h \\ \dots \\ \alpha^{n\prime}(t_0)h \end{pmatrix}$, also folgt die Behauptung.

(iv) $\alpha^{'}(t_0)$ von (iii) ist doch der Jacobi-Vektor von α. Daher heißt $\alpha^{'}(t_0)$ der **Tangentenvektor an die Kurve** $\alpha(t)$ in t_0.

Denn: $\alpha^{'}(t_0) = \frac{D\alpha_{t_1}(h)}{h} = \begin{pmatrix} \lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{\alpha^1(t_0+h)-\alpha^1(t_0)}{h} \\ \dots \\ \lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{\alpha^n(t_0+h)-\alpha^{(n)}(t_0)}{h} \end{pmatrix} = \lim\limits_{\substack{h\to 0\\ h\neq 0}} \frac{\alpha(t_0+h)-\alpha(t_0)}{h}$.

(v) Für $D \Subset \mathbb{R}^n$ und differenzierbare Funktionen $f : D \to \mathbb{R}$, $\alpha : \mathbb{R} \to D$ gilt $(f \circ \alpha)'(t)h = D(f \circ \alpha)_t(h) = Df_{\alpha(t)}(D\alpha_t(h)) = Df_{\alpha(t)}(\alpha'(t))h$, also $Df_{\alpha(t)}(\alpha'(t)) = (f \circ \alpha)'(t)$. Letzterer Ausdruck gibt doch die Änderung von f in Richtung der durch $\alpha(t)$ beschriebenen Kurve an (, also auch in Richtung von $\alpha'(t)$). Das liefert uns eine „anschauliche "Interpretation der totalen Ableitung:
$Df_a(h)$ gibt die Änderung von der Funktion f im Punkte a in Richtung h an.
Wenn $||h|| = 1$ ist, so heißt $Df_a(h)$ auch manchmal die **Richtungsableitung** von f im Punkte a in Richtung h. Diese wird (in Anlehnung an die Schreibweise für partielle Ableitungen) auch geschrieben als $D_h f(a)$.

(vi) Die Änderung $Df_a(h) = \langle \bigtriangledown f(a), h\rangle = || \bigtriangledown f(a)||\,||h|| cos \sphericalangle(\bigtriangledown f(a), h)$ ist am größten, wenn $\sphericalangle(\bigtriangledown f(a), h) = 0$ ist, d. h., wenn h in Richtung von $\bigtriangledown f(a)$ zeigt. Anders gewendet:

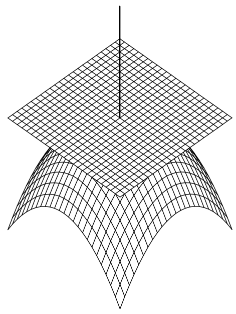
> **Slogan:**$\bigtriangledown f = gradf$ zeigt stets in Richtung des größten Anstiegs von f, und $- \bigtriangledown f$ zeigt in Richtung der stärksten Abnahme von f.

(vii) Nun ergibt sich nach 7.27(v) $\langle \bigtriangledown f(a), \alpha'(t)\rangle = Df_{\alpha(t)}(\alpha'(t)) = (f\circ\alpha)'(t)$, also ist $\bigtriangledown f(\alpha(t) \perp \alpha(t)$ äquivalent mit $f \circ \alpha = const.$

> **Slogan:** $\bigtriangledown f(a)$ ist ein **Normalenvektor** auf der **Niveauhyperfläche** $f = const$ im Punkte a

(viii) Ein Beispiel aus der Physik: Die elektrische Feldstärke $E(x)$ hat in einem elektrostatischen Spannungsfeld $U(x)$ die Richtung des stärksten Spannungsabfalls, da $E(x) = -grad\, U(x)$ gilt.

(ix) Gleichung der Tangentialhyperebene:



Wir erhalten also eine Gleichung für die Tangentialhyperebene im Punkte a an die Niveauhyperfläche $f = const$ durch $\langle \bigtriangledown f(a), x - a \rangle = 0$.