12. Differentialrechnung im $\mathbb{R}^n$

Partielle und totale Differenzierbarkeit. Im Folgenden sei $U \subseteq \mathbb{R}^n$ offen, X ein Banachraum und $f : U \to X$ eine Funktion.

12.1. Motivation. Wir betrachten die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,y) = \sin(2x + 3y).$$

Wir können f nach x ableiten und erhalten $2\cos(2x+3y)$. Ebenso wir können f nach y ableiten und erhalten $3\cos(2x+3y)$. Um zu kennzeichnen, nach welcher Variable man ableitet, schreibt man

$$\frac{\partial f}{\partial x}(x,y) = 2\cos(2x+3y) \text{ bzw. } \frac{\partial f}{\partial y}(x,y) = 3\cos(2x+3y)$$

(mit der von Leibniz eingeführten geschwungenen Variante ∂ des Buchstabens d) statt $\frac{d}{dx}$ bzw. $\frac{d}{dy}$. Allgemein sieht das so aus:

12.2. Partielle Ableitung. Es sei $x \in U$. Wir nennen f in x nach x_j partiell differenzierbar, falls der Grenzwert

$$\frac{\partial f}{\partial x_j}(x) = \lim_{h\to 0} \frac{f(x_1,\dots,x_{j-1},x_j+h,x_{j+1},\dots,x_n) - f(x_1,\dots,x_n)}{h}$$

existiert. Man nennt ihn die partielle Ableitung von f nach x_j in x und schreibt $\frac{\partial f}{\partial x_j}$ oder auch $\partial_{x_j} f(x)$, manchmal sogar nur $f_{x_j}(x)$.

Ist f in jedem Punkt partiell nach x_j differenzierbar, so heißt f auf U partiell nach x_j differenzierbar.

Ist f nach allen Variablen x_j, $j = 1,\dots,n$, partiell differenzierbar in x bzw. auf U, so heißt f partiell differenzierbar in x bzw. auf U. Sind die partiellen Ableitungen stetig, so heißt f stetig partiell differenzierbar. Ebenso definiert man höhere partielle Ableitungen, z.B. $\frac{\partial^2 f}{\partial x_2 \partial x_1}(x)$.

12.3. Bemerkungen.

(a) Bezeichnet man mit e_j den j-ten Einheitsvektor, so schreibt sich die partielle Ableitung

$$\frac{\partial f}{\partial x_j}(x) = \lim_{h\to 0} \frac{f(x+he_j) - f(x)}{h}.$$

(b) Wählt man die $x_k, k \neq j$ fest und betrachtet die Funktion

$$g(t) = f(x_1,\dots,x_{j-1},t,x_{j+1},\dots,x_n)$$

die für t nahe bei x_j definiert ist, so gilt – falls der Grenzwert existiert –

$$\frac{\partial f}{\partial x_j}(x) = g'(x_j).$$

Die partielle Ableitung ist also eine gewöhnliche Ableitung bei festgehaltenen anderen Variablen.

12.4. Richtungsableitung. Statt der Vektoren $e_1,\dots,e_n$ kann man auch beliebige andere Vektoren einsetzen: Ist v ein beliebiger Vektor in $\mathbb{R}^n$, so ist die partielle Ableitung von f nach v in x definiert durch

$$\frac{\partial f}{\partial v}(x) = \lim_{h\to 0} \frac{f(x+hv) - f(x)}{h},$$

sofern der Limes existiert.

12.5. Beispiel.

(a) $f : \mathbb{R}^2 \to \mathbb{C}^2$ sei definiert durch
$$f(x_1, x_2) = (e^{3x_1+ix_2}, 2x_1^2x_2^3).$$
Dann ist
$$\frac{\partial f}{\partial x_1} = (3\,e^{3x_1+ix_2}, 4x_1x_2^3), \qquad \frac{\partial f}{\partial x_2} = (i\,e^{3x_1+ix_2}, 6x_1^2x_2^2),$$
f ist daher auf ganz $\mathbb{R}^2$ stetig partiell differenzierbar; sogar – wie man leicht sieht – unendlich oft.

(b) $r : \mathbb{R}^n \to \mathbb{R}$ sei definiert durch $r(x) = \|x\| = (\sum_{j=1}^n x_j^2)^{1/2}$. Dann ist r auf $\mathbb{R}^n \setminus \{0\}$ stetig partiell differenzierbar, mit
$$\frac{\partial r}{\partial x_j} = \frac{1}{2}(\sum x_j^2)^{-1/2} \cdot 2x_j = \frac{x_j}{r}.$$

Aus der Beobachtung in 12.3(b) folgt sofort:

12.6. Satz.

(a) *$f : U \to \mathbb{R}^m$ ist in x partiell nach x_j differenzierbar, genau dann, wenn alle Komponentenfunktionen f_k, $k = 1, \dots, m$, in x nach x_j partiell differenzierbar sind. In diesem Fall ist die k-te Komponente von $\partial f/\partial x_j$ gleich $\partial f_k/\partial x_j$.*

(b) *$f : U \to \mathbb{C}^m$ ist in x partiell nach x_j differenzierbar, genau dann, wenn alle Real- und Imaginärteile $\mathrm{Re}\, f_k$ und $\mathrm{Im}\, f_k$ der Komponentenfunktionen in x nach x_j partiell differenzierbar sind. Dann ist $\mathrm{Re}\,(\partial f_k/\partial x_j) = \partial \mathrm{Re}\, f_k/\partial x_j$, analog für den Imaginärteil.*

(c) *Es gelten Summen-, Produkt- und Quotientenregel.*

12.7. Schwäche der partiellen Ableitung. Auf $D \subseteq \mathbb{R}$ differenzierbare Funktionen sind stetig (Folgerung 7.4). Partiell differenzierbare Funktionen sind jedoch i. Allg. nicht stetig. Wir betrachten z. B. für $n \geq 2$ die Funktion
$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) = \begin{cases} \dfrac{x_1 \cdot \ldots \cdot x_n}{\|x\|^{2n}}, & x \neq 0, \\ f(0) = 0 \end{cases}.$$
Dann ist f in 0 nicht stetig, da die Folge $(1/k, \dots, 1/k)$ gegen Null konvergiert, aber
$$f(\frac{1}{k}, \dots, \frac{1}{k}) = \frac{k^{-n}}{(nk^{-2})^n} = \left(\frac{k}{n}\right)^n \to \infty.$$
Andererseits ist f in 0 nach $x_1, \dots, x_n$ partiell differenzierbar, denn
$$\frac{f(0 + he_j) - f(0)}{h} = \frac{0 - 0}{h} = 0.$$
Beachte: Auf $\mathbb{R}^n \setminus \{0\}$ ist f als Komposition beliebig oft partiell differenzierbarer Funktionen beliebig oft partiell differenzierbar.

Beispielsweise erhalten wir mit $r = \|x\|$.
$$\frac{\partial f}{\partial x_j} = \frac{x_1 \ldots x_{j-1}x_{j+1} \ldots x_n}{r^{2n}} - 2n\frac{x_1 \ldots x_n}{r^{2n+1}} \cdot \frac{x_j}{r}.$$

12.8. Gradient, Divergenz, Rotation.

(a) Es sei $f : U \to \mathbb{R}$ (!) partiell differenzierbar. Dann heißt

$$\operatorname{grad} f(x) = \left(\frac{\partial f}{\partial x_1}(x) \dots \frac{\partial f}{\partial x_n}(x)\right) \in \mathbb{R}^n$$

der *Gradient* von f in x. Manchmal schreibt man auch $\nabla f(x)$ „Nabla f“. Die Funktion $\operatorname{grad} f : U \to \mathbb{R}$ ordnet jedem Punkt $x \in U \subseteq \mathbb{R}^n$ den Vektor $\operatorname{grad} f(x) \in \mathbb{R}^n$ zu.

Ganz allgemein nennt man eine Funktion $g : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ ein Vektorfeld.

(b) Ist $g : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ ein partiell differenzierbares Vektorfeld, so heißt $\operatorname{div} g = \sum_{j=1}^n \frac{\partial g_j}{\partial x_j}$ die Divergenz von g in x.

Formal kann man schreiben

$$\operatorname{div} g = \langle \nabla, g \rangle,$$

wobei ∇ als der „Vektor“ $(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n})$ interpretiert wird.

(c) Ist $g : U \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ ein partiell differenzierbares Vektorfeld, so heißt

$$\operatorname{rot} g = \left(\frac{\partial g_3}{\partial x_2} - \frac{\partial g_2}{\partial x_3}, \frac{\partial g_1}{\partial x_3} - \frac{\partial g_3}{\partial x_1}, \frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2}\right)$$

die Rotation von g in x (Englisch: $\operatorname{curl} g$). Formal ist

$$\operatorname{rot} g = \nabla \times g.$$

12.9. Satz. (Satz von Schwarz). *Ist $f : U \subseteq \mathbb{R}^n \to X$ zweimal stetig partiell differenzierbar, so ist*

$$\partial_{x_j}\partial_{x_k} f(x) = \partial_{x_k}\partial_{x_j} f(x), \quad x \in U;$$

man kann also die Reihenfolge der partiellen Ableitungen vertauschen.

Mehrfache Anwendung des Satzes zeigt, dass man für eine ℓ-mal stetig partiell differenzierbare Funktionen die Reihenfolge für ℓ partielle Ableitungen vertauschen kann.

Beweis. 1. Vereinfachung: Es genügt, ihn für $n = 2$ zu zeigen, da partielle Ableitungen gewöhnliche Ableitungen bei festgehaltenen anderen Variablen sind und die restlichen Variablen keine Rolle spielen.

2. Vereinfachung: Es genügt, ihn für $U = \{(y_1, y_2) : |y_1|, |y_2| < c\}$ für ein $c > 0$ und $(x_1, x_2) = (0, 0)$ zu zeigen. Durch evtl. Verkleinern erreichen wir dir obige Form von U. Ist ferner $(a, b) \in U$, so betrachte $g(x_1, x_2) = f(x_1 + a, x_2 + b)$. Die Ableitungen von f in (a, b) sind die von g in $(0, 0)$. Wir schreiben $\partial_1 f$ statt $\partial_{x_1} f$, $\partial_{12} f$ statt $\partial_{x_1}\partial_{x_2} f$ usw..

Nun zum Beweis. Wir zeigen:

$$\partial_{21} f(0, 0) = \lim_{\substack{(h,k)\to(0,0)\\ h,k\neq 0}} \frac{f(h, k) - f(0, k) - f(h, 0) + f(0, 0)}{hk} = \partial_{12} f(0, 0).$$

Zunächst die erste Identität: Wir bezeichnen den obigen Bruch mit $D(h, k)$ und schreiben (Hauptsatz der Differential- und Integralrechnung, angewendet auf die Funktion $t \mapsto f(t, k)$):

$$\frac{1}{h}(f(h, k) - f(0, k)) = \frac{1}{h} \int_0^h \partial_1 f(t, k)\, dt \overset{t=sh}{=} \int_0^1 \partial_1 f(sh, k) ds.$$

Ebenso ist $\frac{1}{h}(f(h, 0) - f(0, 0)) = \int_0^1 \partial_1 f(sh, 0)\, ds$, also

$$D(h, k) = \int_0^1 \frac{\partial_1 f(sh, k) - \partial_1 f(sh, 0)}{k}\, ds. \tag{1}$$

Nun wenden wir - analog wie eben - den Hauptsatz auf die Funktion $u \mapsto \partial_1 f(sh, u)$ an und erhalten :

(2) $$D(h,k) = \int_0^1 \Big(\int_0^1 \partial_{21} f(sh, tk)\, dt \Big) ds.$$

Ist $\varepsilon > 0$ vorgelegt, so ist wegen der Stetigkeit von $\partial_{21} f$

$$\|\partial_{21} f(sh, tk) - \partial_{21} f(0,0)\| < \varepsilon$$

falls $|h|, |k|$ hinreichend klein sind (beachte: $|sh| \le |h|$ und $|tk| \le |k|$). Beachten wir, dass $\int_0^1 \Big(\int_0^1 \partial_{21} f(0,0) dt \Big) ds = \partial_{21} f(0,0)$, so erhalten wir

$$\|D(h,k) - \partial_{21} f(0,0)\| = \left\| \int_0^1 \Big(\int_0^1 \partial_{21} f(sh, tk)\, dt \Big) ds - \partial_{21} f(0,0) \right\|$$

$$= \left\| \int_0^1 \Big(\int_0^1 \partial_{21} f(sh, tk) - \partial_{21} f(0,0)\, dt \Big) ds \right\| \le \int_0^1 \Big(\int_0^1 \|\partial_{21} f(sh, tk) - \partial_{21} f(0,0)\|\ dt \Big) ds < \varepsilon.$$

Damit haben wir die erste Identität bewiesen. Für die zweite schreibe

$$D(h,k) = \frac{f(h,k) - f(0,k) - f(h,0) + f(0,0)}{hk}$$

und argumentiere wie oben mit vertauschten Rollen. $\triangleleft$

12.10. Lemma. *Es seien $f : U \subseteq \mathbb{R}^3 \to \mathbb{R}$ und $g : U \subseteq \mathbb{R}^3 \to \mathbb{R}^3$ zweimal stetig partiell differenzierbar. Dann ist*

$$\operatorname{rot} \operatorname{grad} f = 0 \qquad \text{und} \qquad \operatorname{div} \operatorname{rot} g = 0.$$

Beweis. Nachrechnen. Beispiel: Erste Komponente von $\operatorname{rot}\operatorname{grad} f$ ist $\partial_{x_2}\partial_{x_3} f - \partial_{x_3}\partial_{x_2} f = 0$ nach Schwarz.

12.11. Operatoren der Mathematischen Physik. Es sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{C}$ zweimal stetig partiell differenzierbar. Man setzt

$$\Delta f = \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j^2}$$

und nennt $\Delta = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}$ den Laplace-Operator. Wichtigster Operator der Math. Physik. Beachte: $\Delta = \operatorname{div}\operatorname{grad}$.

Die Gleichung $\Delta f = 0$ heißt Laplacegleichung; ihre Lösungen *harmonische Funktionen*. Die inhomogene Gleichung $\Delta f = g$ (bei gegebenem g und gesuchtem f) heißt meist Poissongleichung.

Beispiel: $\Delta U = 4\pi\rho$ (U elektro-magnetisches Potential, ρ Ladungsdichte) ist die Gleichung für das elektro-magnetische Potential bei gegebener Ladungsverteilung.

Weiterhin sei $I \subseteq \mathbb{R}$ ein Intervall. Für Funktionen $f : U \times I \to \mathbb{R}$ heißt

$$\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \Delta_x f = 0$$

die Wellengleichung; hier bedeutet $\Delta_x = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}$; c ist die Ausbreitungsgeschwindigkeit.

Ferner ist

$$\frac{1}{k}\frac{\partial f}{\partial t} - \Delta_x f = 0$$

die Wärmeleitungsgleichung. Hier ist $k \ge 0$ die Leitfähigkeit.

Meist skaliert man die Konstanten c und k auf 1. Der Wärmeleitungsoperator ist dann $\partial_t - \Delta$, der Wellenoperator $\partial_t^2 - \Delta$.

Achtung: Sind die Variablen x und t, und kommt ein Δ vor, so versteht man die Gleichung meist so, dass Δ nur bzgl. der x-Variablen wirkt.

12.12. Beispiel. Es sei $f : \mathbb{R}_+ \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Wir betrachten die Funktion

$$N : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$$

definiert durch $N(x) = f(r(x))$; wie vorher ist $r(x) = \|x\|$. Dann gilt nach 12.5(b)

$$\operatorname{grad} N(x) = f'(r(x))\frac{x}{r(x)},$$

also

$$\begin{aligned}
\Delta N(x) &= \operatorname{div}\operatorname{grad} f(r(x)) = \operatorname{div} f'(r(x))\frac{x}{r(x)}\\
&= \sum_{j=1}^n f''(r(x))\frac{x_j^2}{r^2(x)} + \sum_{j=1}^n f'(r(x))\left(-\frac{x_j}{r^2(x)}\frac{x_j}{r(x)} + \frac{1}{r(x)}\right)\\
&= f''(r(x)) + (n-1)\frac{f'(r(x))}{r(x)},
\end{aligned}$$

da $\sum x_j^2 = \|x\|^2 = r^2$. Speziell für $f(r) = \dfrac{1}{r^{n-2}}$ ist

$$\Delta N = (2-n)(1-n)r^{-n} + (n-1)(2-n)r^{-n} = 0,$$

sowie für $f(r) = \ln r$ im Fall $n = 2$ ebenfalls

$$\Delta N(x) = -\frac{1}{r^2(x)} + (2-1)\frac{1/r(x)}{r(x)} = 0.$$

In $x = 0$ ist N nicht differenzierbar. Man kann jedoch – wie man z.B. in einem Kurs 'Partielle Differentialgleichungen' lernt – die Ableitung als Distribution definieren. Es gilt $\Delta N = c_n\delta$. Dabei ist $c_2 = 2\pi$ und $c_n = (n-2)\omega_n$ für $n \geq 3$; ω_n ist das Volumen der Einheitssphäre in $\mathbb{R}^n$ und δ die Delta-'Funktion' (besser Delta-Distribution). Man nennt N/c_n *Newtonpotential*.

12.13. Definition/Erinnerung.

(i) Eine *Umgebung* einer Menge M ist eine offene Menge U, die M enthält.

(ii) Ist $U \subseteq \mathbb{R}^n$ offen, und $x \in U$, so existiert ein $\varepsilon > 0$ mit $B(x,\varepsilon) \subseteq U$. Insbesondere ist $x + h \in U$ für alle $h \in \mathbb{R}^n$ mit $\|h\| < \varepsilon$.

(iii) In Satz 7.3 hatten wir gesehen, dass für eine Funktion $f : D \subseteq \mathbb{R} \to X$ äquivalent ist:

  (a) f ist in t_0 differenzierbar.

  (b) Es gibt ein $c \in X$ und eine Funktion $\varphi : D \to X$ mit $\lim_{t\to t_0} \frac{\varphi(t)}{t-t_0} = 0$ so, dass

$$f(t) = f(t_0) + c(t-t_0) + \varphi(t).$$

  In diesem Fall ist $c = f'(t_0)$.

Wir *definieren* nun die Differenzierbarkeit für Funktionen von mehreren Variablen mit Hilfe dieser Eigenschaft. Man erhält also einen neuen Differenzierbarkeitsbegriff, den man im Unterschied zur partiellen Differenzierbarkeit als *totale Differenzierbarkeit* bezeichnet. Anders als im Fall einer Variable sind die Eigenschaften nicht äquivalent.

12.14. Totale Differenzierbarkeit. Eine Funktion $f : U \to X$ heißt in $x \in U$ total differenzierbar, falls eine (von x abhängige) lineare Abbildung

$$A : \mathbb{R}^n \to X$$

existiert, so dass für alle h mit $\|h\| < \varepsilon$ (ε wie in 12.13) gilt:

$$f(x+h) = f(x) + Ah + \varphi(h),$$

wobei $\varphi : \{h : \|h\| < \varepsilon\} \to X$ eine Funktion ist mit

$$\lim_{\|h\|\to 0} \frac{\varphi(h)}{\|h\|} = 0. \tag{1}$$

Die lineare Abbildung A ist die Ableitung von f in x. Schreibe $A = f'(x)$ (auch $A = \partial f(x)$).

Man nennt f (total) differenzierbar auf U, falls f in jedem Punkt von U (total) differenzierbar ist.

12.15. Bemerkung. Ist $X = \mathbb{C}^m$, so ist A eine lineare Abbildung von $\mathbb{R}^n$ nach $\mathbb{C}^m$, also durch eine Matrix in $\mathrm{Mat}_{mn}(\mathbb{C})$ die sog. *Jacobi-Matrix* gegeben

12.16. Beispiel.

(a) Es sei $f(x) = Tx$ eine lineare Abbildung von $\mathbb{R}^n$ nach X. Dann ist $f(x+h) = T(x+h) = Tx + Th = f(x) + Th$, somit ist f total differenzierbar und $f'(x) = T$ ist konstant. Hier ist $\varphi \equiv 0$.

(b) Es sei $f : \mathbb{R}^n \to \mathbb{R}$ gegeben durch $f(x) = \langle x, x\rangle = \|x\|^2$. Dann ist

$$f(x+h) = \langle x+h, x+h\rangle = f(x) + 2\langle x, h\rangle + \|h\|^2.$$

Dann ist $f'(x)$ die durch $f'(x)h = 2\langle x, h\rangle$ definierte lineare Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}$. Die Jacobi-Matrix ist der Zeilenvektor x, und $\varphi(h) = \|h\|^2$.

12.17. Landausche Symbole. Ist φ eine auf einer Umgebung von $0 \in \mathbb{R}^n$ definierte Funktion mit $\lim \varphi(h)/\|h\| \to 0$ für $h \to 0$, so schreibt man $\varphi(h) = o(\|h\|)$. Relation 12.14(1) heißt also

$$f(x+h) - f(x) - Ah = o(\|h\|) \quad \text{für} \quad h \to 0.$$

Allgemeiner: Sind φ, ψ zwei solche Funktionen und ist $\psi(h) \neq 0$ für alle $h \neq 0$, so schreibt man

$$\begin{aligned} \varphi(h) &= o(\psi(h)), \quad \text{falls} \quad \varphi(h)/\psi(h) \to 0 \quad \text{für} \quad h \to 0 \\ \varphi(h) &= O(\psi(h)), \quad \text{falls} \quad \varphi(h)/\psi(h) \quad \text{beschränkt für} \quad h \to 0. \end{aligned}$$

12.18. Satz. *Für $f : U \to \mathbb{C}^m$ und $x \in U$ ist äquivalent*

(a) *f total differenzierbar in x;*
(b) *alle Komponentenfunktionen f_j, $j = 1, \ldots, m$, sind total differenzierbar*
(c) *Real- und Imaginärteil jeder Komponente ist total differenzierbar*

Beweis. Alle Aussagen folgen sofort, wenn man sich überlegt, wie sich der Beitrag von Ah zusammensetzt:

(b) Hat f die Ableitung $A = (a_{ij})$, so hat die i-te Komponente die i-te Zeile von A als Ableitung.

(c) Die Ableitung des Realteils der i-ten Komponente ist der Realteil der Ableitung. Ist also die Ableitung von f_i die (Zeilen-)Matrix $(a_{i1}, \ldots, a_{in})$, so ist die von $\mathrm{Re}\, f_i$ die Matrix $(\mathrm{Re}\, a_{i1}, \ldots, \mathrm{Re}\, a_{in})$. $\triangleleft$

Zum weiteren Verständnis benötigen wir:

12.19. Definition und Lemma. Es sei $T : \mathbb{R}^n \to X$ eine lineare Abbildung und $e_1, \ldots, e_n$ die Einheitsvektoren in $\mathbb{R}^n$. Dann ist T stetig: Für $x, h \in \mathbb{R}^n$ gilt

$$\|T(x+h) - Tx\| = \|Th\| \le \sum \|h_j Te_j\| \le \|h\|_1 \max\{\|Te_j\| : j = 1, \ldots, n\} \to 0,$$

falls $h \to 0$.

Der Einheitskreis in $\mathbb{R}^n$ ist eine kompakte Menge (da abgeschlossen und beschränkt). Die stetige Abbildung $v \mapsto \|Tv\|$ nimmt dort also ihr Maximum an. Wir nennen

$$\|T\| = \max_{\|v\|=1} \|Tv\|$$

die Norm der linearen Abbildung T. Es gilt dann sogar

$$\|T\| = \sup_{v\neq 0} \frac{\|Tv\|}{\|v\|},$$

denn

$$\sup_{v\neq 0} \frac{\|Tv\|}{\|v\|} = \sup_{v\neq 0} \frac{\|T(v/\|v\|)\|}{\|v/\|v\|\|} \leq \max_{\|w\|=1} \|Tw\| \leq \sup_{v\neq 0} \frac{\|Tv\|}{\|v\|}.$$

Insbesondere ist

$$\|Tv\| \leq \|T\|\|v\| \text{ für alle } v.$$

Man bezeichnet mit $\mathcal{L}(\mathbb{R}^n, X)$ den Raum der (stetigen) linearen Abbildungen von $\mathbb{R}^n$ nach X. Mit der soeben definierten Norm ist er ein normierter Raum. Man kann zeigen, dass er sogar ein Banachraum ist.

Ist $X = \mathbb{C}^m$ und T durch eine Matrix (t_{ij}) gegeben, so hat man eine einfache Abschätzung für die Norm:

$$\|T\| \leq \Big(\sum_{ij} |t_{ij}|^2\Big)^{1/2}.$$

Das sieht man so:

$$\begin{aligned}
\|Tv\|^2 &= \sum_{i=1}^m |\sum_{j=1}^n t_{ij}v_j|^2 \\
&\leq \sum_{i=1}^m \left(\sum_{j=1}^n |t_{ij}|^2\right)\left(\sum_{j=1}^n |v_j|^2\right) \qquad \text{(Cauchy-Schwarz)} \\
&= \sum_{ij} |t_{ij}|^2 \|v\|^2.
\end{aligned}$$

12.20. Satz. *Es sei $f : U \subseteq \mathbb{R}^n \to X$ in $x \in U$ total differenzierbar mit $f'(x) = A$. Dann gilt:*

(a) *f ist in x stetig.*

(b) *f ist in x in alle Koordinatenrichtungen partiell differenzierbar.*

(c) *Ist $X = \mathbb{C}^m$ und A durch die Matrix (a_{ij}) gegeben, so gilt*

$$\frac{\partial f_i}{\partial x_j}(x) = a_{ij}.$$

Mit anderen Worten: $f'(x) = \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}.$

Beweis.

(a) Nach Definition der totalen Differenzierbarkeit ist

(1) $$f(x+h) = f(x) + Ah + \varphi(h).$$

Da $Ah \to 0$ für $h \to 0$ und $\varphi(h) \to 0$ für $h \to 0$, folgt die Stetigkeit.

(b) Aus (1) folgt

$$\frac{f(x+te_j)-f(x)}{t}=Ae_j+\frac{\varphi(te_j)}{t}.$$

Da $\varphi(te_j)/|t|=\varphi(te_j)/\|te_j\|\to 0$, folgt $\frac{\partial f}{\partial x_j}(x)=Ae_j$.

(c) Ist $X=\mathbb{C}^m$ so ist nach 12.6 die Ableitung der i-ten Komponentenfunktion nach x_j die i-te Komponente der Ableitung. Also ist $\partial f_i/\partial x_j = i$-te Komponente von $Ae_j=a_{ij}$. $\triangleleft$

12.21. Satz. *Es sei $U\subseteq\mathbb{R}^n$ offen und $f:U\to X$ partiell differenzierbar. In $x\in U$ seien alle partiellen Ableitungen stetig. Dann ist f in x total differenzierbar.*

Beweis. Schreibe $h\in\mathbb{R}^n$ in der Form $h=\sum h_je_j$ und setze $x^{(\nu)}=x+\sum_{j=1}^{\nu}h_je_j$. Dann gilt $x^{(0)}=x, x^{(n)}=x+h$. Die Werte $x^{(\nu)}$ und $x^{(\nu+1)}$ unterscheiden sich nur in einer Komponente. Also können wir auf f die (eindimensionale) Charakterisierung der Differenzierbarkeit aus Satz 7.3 anwenden

$$f(x^{(\nu)})=f(x^{(\nu-1)})+\frac{\partial f}{\partial x_\nu}(x^{(\nu-1)})h_\nu+\varphi_\nu(h_\nu),$$

wobei

(1) $$g_\nu(h_\nu)=o(h_\nu),\ \text{also erst recht}\ \ \varphi_\nu(h_\nu)=o(\|h\|).$$

Es folgt

$$f(x+h)-f(x)=\sum_{\nu=1}^{n}[f(x^{(\nu)})-f(x^{(\nu-1)})]=\sum_{\nu=1}^{n}\frac{\partial f}{\partial x_\nu}(x^{(\nu-1)})h_\nu+\sum_{\nu=1}^{n}\varphi_\nu(h_\nu).$$

Wir definieren A durch $Ae_\nu=(\partial f/\partial x_\nu)(x)$. Dann folgt

$$f(x+h)-f(x)-Ah=\sum_{\nu=1}^{n}\left(\frac{\partial f}{\partial x_\nu}(x^{(\nu-1)})-\frac{\partial f}{\partial x_\nu}(x)\right)h_\nu+\sum_{\nu=1}^{n}\varphi_\nu(h_\nu)=o(\|h\|),$$

denn da $(\partial f/\partial x_\nu)(x^{(\nu-1)})-(\partial f/\partial x_\nu)(x)\to 0$ für $h\to 0$ (auf Grund der Stetigkeit von $\partial f/\partial x_\nu$ in x) geht die erste Summe auch bei Division durch $\|h\|$ für $h\to 0$ gegen Null, und wegen (1) ist dies auch für die zweite der Fall. $\triangleleft$

12.22. Definition. Wir nennen eine Funktion $f:U\to X$ auf U stetig total differenzierbar, falls f total differenzierbar ist und die Abbildung $f':U\to\mathcal{L}(\mathbb{R}^n,X)$, $x\mapsto f'(x)$ stetig ist.

Wir erhalten also aus den Sätzen 12.20 und 12.21 folgendes wichtige Resultat:

12.23. Satz. *$f:U\to X$ ist auf U genau dann stetig partiell differenzierbar wenn f auf U stetig total differenzierbar ist. Man spricht daher einfach von stetig differenzierbaren Funktionen.*

Anwendungen.

12.24. Die Tangentialfläche an den Funktionsgraphen. Es sei $f:U\to\mathbb{R}$ total differenzierbar in $x\in U$. Der Graph von f ist die Menge

$$G_f=\{(y,f(y)):y\in U\}\subseteq\mathbb{R}^{n+1}.$$

Die Differenzierbarkeit von f besagt, dass f bei x_0 gut durch die affin-lineare Funktion $l(x_0+h)=f(x_0)+f'(x_0)h$ genähert werden kann. Die Tangentialfläche an den Graphen in dem Punkt $(x_0,f(x_0))\in G_f$ ist der Graph dieser Funktion, d.h. der affin-lineare Raum

$$\{(x_0+h,f(x_0)+f'(x_0)h):h\in\mathbb{R}^n\}=(x_0,f(x_0))+\{(h,f'(x_0)h):h\in\mathbb{R}^n\}\subseteq\mathbb{R}^{n+1}.$$

12.25. Satz (Kettenregel). Es seien $U \subseteq \mathbb{R}^n$ und $V \subseteq \mathbb{R}^k$ offen, $f : U \to X, g : V \to U$ Abbildungen. Ist g differenzierbar in $x \in V$ und f differenzierbar in $g(x) \in U$, so ist $f \circ g : V \to X$ differenzierbar in x, und es gilt

$$(f \circ g)'(x) = f'(g(x)) \circ g'(x),$$

wobei $\circ$ auf der rechten Seite die Komposition von linearen Abbildungen bedeutet. Insbesondere ergibt sich die Formel

$$\frac{\partial (f \circ g)}{\partial x_j}(x) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(g(x)) \frac{\partial g_k}{\partial x_j}(x).$$

Beweis. Für $y = g(x)$ schreibe $f(y+h) = f(y) + Ah + \varphi(h)$ und $g(x+h) = g(x) + Bh + \psi(h)$ mit $A = f'(y)$, $B = g'(x)$. Dann ist

$$\begin{aligned} (f \circ g)(x+h) &= f(g(x+h)) = f(g(x) + Bh + \psi(h)) \\ &= f \circ g(x) + ABh + A\psi(h) + \varphi(Bh + \psi(h)). \end{aligned}$$

Es bleibt also nur zu zeigen, dass $A\psi(h) + \varphi(Bh + \psi(h)) = o(\|h\|)$. Dazu: Zunächst ist

$$\frac{\|A\psi(h)\|}{\|h\|} = \left\| A \left(\frac{\psi(h)}{\|h\|} \right) \right\| \overset{12.19}{\leq} \|A\| \left\| \frac{\psi(h)}{\|h\|} \right\| \to 0.$$

Nun zu $\varphi(Bh + \psi(h))$. Für alle $h \neq 0$ mit $Bh + \psi(h) = 0$ ist nichts zu zeigen. Stets ist $\|Bh + \psi(h)\| \leq \|B\| \|h\| + \frac{\psi(h)}{\|h\|} \|h\| \to 0$. Es folgt:

$$\frac{\|\varphi(Bh + \psi(h))\|}{\|h\|} = \frac{\|\varphi(Bh + \psi(h))\|}{\|Bh + \psi(h)\|} \frac{\|Bh + \psi(h)\|}{\|h\|} \leq \frac{\|\varphi(Bh + \psi(h))\|}{\|Bh + \psi(h)\|} \left(\|B\| + \frac{\psi(h)}{\|h\|} \right) \longrightarrow 0.$$

◁

12.26. Folgerung. Es sei $U \subseteq \mathbb{R}^n$ offen, $x \in U, v \in \mathbb{R}^n$ und $f : U \to \mathbb{R}(!)$ differenzierbar in x. Dann gilt

$$\frac{\partial f}{\partial v}(x) = \langle \operatorname{grad} f(x), v \rangle.$$

Nach der Cauchy-Schwarzschen Ungleichung ist stets $|\langle \operatorname{grad} f(x), v \rangle| \leq \|\operatorname{grad} f(x)\| \, \|v\|$, wobei Gleichheit nur bei linearer Abhängigkeit von $\operatorname{grad} f(x)$ und v gilt.

Es folgt: Unter allen v mit $\|v\| = 1$ wird $\frac{\partial f}{\partial v}(x)$ maximal für das (eindeutig bestimmte) v, das in Richtung von $\operatorname{grad} f(x)$ zeigt. Der Gradient gibt daher die Richtung des steilsten Anstiegs von f an.

Beweis. Definiere die Funktion $g : \mathbb{R} \to \mathbb{R}^n$ durch $g(t) = x + tv$. Sie ist differenzierbar in 0 mit $g'(0) = v$. Für hinreichend kleines $\varepsilon > 0$ ist $x + tv \in U$ für alle $|t| < \varepsilon$. Nach 12.25 ist $f \circ g$ differenzierbar in 0, und es gilt

$$\frac{\partial f}{\partial v}(x) = \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t} = (f \circ g)'(0) = f'(g(0))g'(0) = f'(x) \circ v = \langle \operatorname{grad} f(x), v \rangle.$$

12.27. Satz (Hauptsatz der Differential- und Integralrechnung). Sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to X$ stetig differenzierbar. Sei $x \in U$ und $h \in \mathbb{R}^n$ so klein, dass $x + th \in U$ für alle $t \in [0,1]$. Dann gilt

$$f(x+h) - f(x) = \int_0^1 f'(x+th) \, dt \cdot h. \tag{1}$$

Beweis. Wir definieren die Funktion $g : [0,1] \to X$ durch $g(t) = f(x+th)$. Dann gilt nach der Kettenregel $g'(t) = f'(x+th)h$. Folglich

$$f(x+h) - f(x) = g(1) - g(0) = \int_0^1 g'(t)\,dt = \int_0^1 f'(x+th)\,dt \cdot h.$$

$\lhd$

12.28. Folgerung (Schrankensatz). In 12.27 sei zusätzlich $\|\partial_{x_j} f(x+th)\| \le L_j$ für alle $0 \le t \le 1, j = 1, \ldots, n$. Dann folgt aus 12.27(1), dass $\|f(x+h) - f(x)\| \le \sum_{j=1}^{n} L_j |h_j|$.

12.29. Multi-Indizes. Es sei $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$. Wir setzen $|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n$: Länge des Multi-Index, $\alpha! = \alpha_1! \cdot \ldots \cdot \alpha_n!$.

Wir schreiben $\partial_x^\alpha f = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_n}^{\alpha_n} f$, falls f $|\alpha|$-mal stetig differenzierbar ist. Für $x \in \mathbb{R}^n$ setzen wir $x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$.

12.30. Hilfssatz. Es sei $f : U \to X$ N-mal stetig differenzierbar, $x \in U$, $h \in \mathbb{R}^n$ mit $x + th \in U$ für $0 \le t \le 1$. Dann gilt für die durch $g(t) = f(x+th)$ definierte Funktion: g ist N-mal stetig differenzierbar auf $[0,1]$ und

$$\frac{d^N}{dt^N} g(t) = \sum_{|\alpha|=N} \frac{N!}{\alpha!} \partial^\alpha f(x+th) h^\alpha.$$

Beweis. Aus der Kettenregel ergibt sich $g'(t) = \sum_{j=1}^n \partial_{x_j} f(x+th) h_j$ und so:

$$\frac{d^N}{dt^N} g(t) = \sum_{j_N=1}^{n} \cdots \sum_{j_1=1}^{n} \partial_{x_{j_N}} \ldots \partial_{x_{j_1}} f(x+th) h_{j_1} \cdots h_{j_N}.$$

Nach dem Satz von Schwarz kann man die Reihenfolge der Ableitungen vertauschen und die Terme für gleiches α zusammenfassen. Das sind dann jeweils $N!/\alpha!$ Stück. (Dazu fragt man sich, wieviele verschiedene N-Tupel $(j_1, \ldots, j_N)$ man aus einem Multi-Index α der Länge N erhält. Zunächst kann man die N Ableitungen aus α in beliebiger Reihenfolge aufführen. Das liefert $N!$ Möglichkeiten. Davon sind jedoch $\alpha!$ Stück gleich.) $\lhd$

12.31. Satz (Taylorformel mit Restglied). *Es seien U, x, h wie in 12.30 und $f : U \to X$ N-mal stetig differenzierbar. Dann gilt*

$$f(x+h) = \sum_{|\alpha|=0}^{N-1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=N} \frac{N}{\alpha!} \int_0^1 (1-t)^{N-1} \partial^\alpha f(x+th)\,dt \cdot h^\alpha.$$

Beweis. Definiere $g : [0,1] \to X$ durch $g(t) = f(x+th)$. Dann gilt nach der Formel von Taylor in $\mathbb{R}$

$$g(1) = \sum_{k=0}^{N-1} \frac{1}{k!} \frac{d^k}{dt^k} g(0) + \frac{1}{(N-1)!} \int_0^1 (1-t)^{N-1} \frac{d^N}{dt^N} g(t)\,dt.$$

Mit 12.30 folgt sofort die Behauptung. $\lhd$

12.32. Definition. Es sei $f : U \to \mathbb{R}$ zweimal stetig differenzierbar. Dann setzt man

$$(\text{Hess}\, f)(x) = (\partial_{x_j} \partial_{x_k} f(x))_{j,k=1,\ldots,n}$$

„Hessesche Matrix von f in x“: Symmetrische $n \times n$-Matrix nach dem Satz von Schwarz.

12.33. Folgerung. Es sei $f : U \to \mathbb{R}$ zweimal stetig differenzierbar und $x \in U$. Dann gilt

$$f(x+h) = f(x) + \langle \operatorname{grad} f(x), h\rangle + \frac{1}{2}\langle \operatorname{Hess} f(x)h, h\rangle + R_3(x,h), \tag{1}$$

wobei $R_3(x,h)/\|h\|^2 \to 0$ für $h \to 0$.

Beweis. Definiere R_3 durch (1). Vergleich mit dem Restglied $R_2(x,h) = \sum_{|\alpha|=2} \frac{2}{\alpha!} \int_0^1 \dots$ aus der Taylorentwicklung für $N = 2$ liefert:

$$\begin{aligned} R_3(x,h) &= R_2(x,h) - \frac{1}{2}\langle \operatorname{Hess} f(x)h, h\rangle \\ &= \sum_{|\alpha|=2} \frac{h^\alpha}{\alpha!}\left(2\int_0^1 (1-t)\frac{\partial^\alpha}{\partial x^\alpha} f(x+th)\, dt - \frac{\partial^\alpha f}{\partial x^\alpha}(x)\right). \end{aligned}$$

Wegen $\int_0^1 (1-t)\, dt = \frac{1}{2}$ ist der letzte Ausdruck

$$= \sum_{|\alpha|=2} \frac{h^\alpha}{\alpha!}\left(2\int_0^1 (1-t)\left(\frac{\partial^\alpha}{\partial x^\alpha} f(x+th) - \frac{\partial^\alpha}{\partial x^\alpha} f(x)\right) dt\right). \tag{2}$$

Wegen der Stetigkeit der zweiten Ableitungen existiert zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit

$$\left\| \frac{\partial^\alpha}{\partial x^\alpha} f(x+y) - \frac{\partial^\alpha}{\partial x^\alpha} f(x)\right\| < \varepsilon \quad \text{falls } \|y\| < \delta.$$

Die Norm des Integrals in (2) ist also $< \varepsilon$ falls $\|h\| < \delta$. Da für $|\alpha| = 2$ gilt $|h^\alpha| \le \|h\|^2$, folgt $\|R_3(x,h)\| \le \varepsilon\|h\|^2$ und somit die Behauptung. ◁

Lokale Extrema. In diesem Abschnitt werden wir sehen, dass sich mit den richtigen Begriffsbildungen die Suche nach Extremstellen weitgehend analog zum eindimensionalen Fall gestaltet.

12.34. Definition. Es sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine Funktion. Man nennt $x \in U$ ein lokales Maximum für f, falls

(1) $f(x) \ge f(y)$ für alle y in einer Umgebung von x

Analog heißt x lokales Minimum, falls

(2) $f(x) \le f(y)$ für alle y in einer Umgebung von x.

Extremum ist der Oberbegriff für Maximum oder Minimum.

Man spricht von einem isolierten Maximum/Minimum/Extremum, falls Gleichheit in (1) bzw. (2) nur für $x = y$ gilt.

12.35. Satz. *Es sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ partiell differenzierbar und x lokales Extremum für f. Dann ist $f'(x) = 0$, d.h. $\partial_{x_j} f(x) = 0$ für $j = 1, \dots, n$.*

Beweis. Für $j = 1, \dots, n$ betrachte $g(t) = f(x+te_j)$. Da U offen ist, ist g für $-\varepsilon < t < \varepsilon$ definiert. Die Funktion g ist differenzierbar nach der Kettenregel und hat in 0 ein lokales Extremum. Es folgt: $\partial_{x_j} f(x) = g'(0) = 0$. ◁

12.36. Definition. Es sei A selbstadjungierte Matrix. Dann ist $\langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$, somit $\langle Ax, x\rangle \in \mathbb{R}$. Man nennt A

- positiv definit, falls $\langle Ax, x\rangle > 0$ für alle $x \ne 0$.
- negativ definit, falls $\langle Ax, x\rangle < 0$ für alle $x \ne 0$.
- indefinit, falls es sowohl ein x mit $\langle Ax, x\rangle > 0$ als auch ein y mit $\langle Ay, y\rangle < 0$ gibt.

Beispiel:

- $A = \mathrm{Id}$ positiv definit: $\langle Ax, x\rangle = \langle x, x\rangle = \|x\|^2 > 0$ für alle $x \neq 0$.
- $A = -\mathrm{Id}$ negativ definit: $\langle Ax, x\rangle = \langle -x, x\rangle = -\|x\|^2 < 0$ für alle $x \neq 0$.

12.37. Definitheit und Eigenwerte. Eine selbstadjungierte Matrix ist diagonalisierbar mit reellen Eigenwerten nach einem Satz der Linearen Algebra. Dann gilt:

- A positiv definit $\Leftrightarrow$ alle Eigenwerte sind positiv.
- A negativ definit $\Leftrightarrow$ alle Eigenwerte sind negativ.
- A indefinit $\Leftrightarrow$ es gibt sowohl positive als auch negative Eigenwerte.

Beweis. Es sei $T^{-1}AT = D$ eine Diagonalmatrix und T orthogonal. Dann ist

$$0 < \langle Ax, x\rangle = \langle TDT^{-1}x, x\rangle = \langle DT^{-1}x, T^*x\rangle \overset{\text{T orth}}{=} \langle DT^{-1}x, T^{-1}x\rangle$$

für alle $x \neq 0$ genau dann, wenn $0 < \langle Dy, y\rangle = \sum_j \lambda_j y_j^2$ für alle $y \neq 0$ mit den Eigenwerten λ_j von A. Dies wiederum ist genau dann der Fall, wenn alle λ_j positiv sind.

Analog argumentiert man bei negativer Definitheit oder Indefinitheit. ◁

12.38. Hurwitz-Kriterium. Es sei $A = (a_{ij})_{i,j=1,\dots,n}$ eine reelle, symmetrische $n \times n$-Matrix. Dann gilt:

$$A \text{ positiv definit} \Leftrightarrow \det (a_{ij})_{1\le i,j\le k} > 0, \quad k = 1, \dots, n.$$

12.39. Satz. *Es sei $U \subseteq \mathbb{R}^n$ offen, $f : U \to \mathbb{R}$ zweimal stetig differenzierbar und $x \in U$ mit* $\operatorname{grad} f(x) = 0$.

(a) *Ist* $(\operatorname{Hess} f)(x)$ *positiv/negativ definit, so hat f in x ein isoliertes Minimum/Maximum.*

(b) *Ist* $(\operatorname{Hess} f)(x)$ *indefinit, so hat f in x kein lokales Extremum.*

Beweis. Nach 12.33 ist

$$f(x + h) = f(x) + \langle \operatorname{grad} f(x), h\rangle + \frac{1}{2}\langle \operatorname{Hess} f(x)h, h\rangle + R_3(x, h),$$

wobei

(1) $$\lim_{h\to 0} R_3(x, h)/\|h\|^2 = 0.$$

Hier ist $\operatorname{grad} f(x) = 0$. Ist $\operatorname{Hess} f(x)$ positiv definiert und $\lambda > 0$ der kleinste Eigenwert, so ist

$$\langle \operatorname{Hess} f(x)h, h\rangle \ge \lambda \|h\|^2.$$

Andererseits existiert wegen (1) ein $\delta > 0$ mit

$$|R_3(x, h)| < \frac{\lambda}{4}\|h\|^2,$$

falls $\|h\| < \delta$. Es folgt:

$$f(x + h) > f(x) + \frac{\lambda}{4}\|h\|^2 > f(x), \quad \|h\| < \delta.$$

Analog schließen wir für negativ definites $\operatorname{Hess} f(x)$.

(b) Wähle eine Eigenvektor h_+ zu einem positiven Eigenwert λ_+, und einen Eigenvektor h_- zu einem negativen Eigenwert λ_-, von $\operatorname{Hess} f(x)$ mit $\|h_\pm\| = 1$.

Dann ist $\langle \operatorname{Hess} f(x)h_+, h_+\rangle = \lambda_+ \|h_+\|^2 = \lambda_+$ und $\langle \operatorname{Hess} f(x)h_-, h_-\rangle = \lambda_-$.

Für $t \to 0$ betrachte $f(x+th_+)$ und $f(x+th_-)$. Es gilt

$$f(x+th_\pm) - f(x) = \frac{1}{2}\langle \mathrm{Hess} f(x) th_\pm, th_\pm\rangle + R_3(x, th_\pm) = \frac{1}{2}\lambda_\pm t^2 + o(t^2).$$

Für kleines t ist dieser Ausdruck positiv bei "+" und negativ bei "–". ◁

Inverse und implizite Funktionen.

12.40. Definition. Eine Abbildung $f : U \to V$ zwischen zwei offenen Teilmengen des $U, V \subset \mathbb{R}^n$ heißt (C^1-)Diffeomorphismus, falls f bijektiv ist und sowohl f als auch die Umkehrabbildung f^{-1} stetig differenzierbar sind.

12.41. Lemma. *Ist $f : U \to V$ ein Diffeomorphismus, so ist $f'(x)$ invertierbar für jedes $x \in U$ und*

$$(f^{-1})'(f(x)) = (f'(x))^{-1}.$$

◁

Beweis. Die Identität $f^{-1}(f(x)) = x$ liefert beim Ableiten: $(f^{-1})'(f(x)) \circ f'(x) = \mathrm{Id}$.

12.42. Frage. Wäre es sinnvoll, den Begriff des Diffeomorphismus $f : U \to V$ für $U \subseteq \mathbb{R}^n$ und $V \subseteq \mathbb{R}^k$ einzuführen?

Antwort: Nein, denn wie oben wäre $f'(x)$ invertierbar. Dies geht nur für $k = n$.

12.43. Lemma. *Es seien $U, V \subseteq \mathbb{R}^n$ offen, $f : U \to V$ ein Diffeomorphismus und $W \subseteq U$ offen. Dann ist $f(W)$ offen. Mit anderen Worten: Diffeomorphismen bilden offene Mengen auf offene Mengen ab.*

Beweis. Wegen der Bijektivität ist $W = f^{-1}(f(W))$. Also ist $f(W) = (f^{-1})^{-1}(W)$ als das Urbild der offenen Menge W unter der stetigen Abbildung f^{-1} offen, vgl. Satz 6.14. ◁

12.44. Satz von der lokalen Invertierbarkeit. *Es sei $U \subseteq \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ stetig differenzierbar. Ferner sei für ein $a \in U$ die Ableitung $f'(a)$ invertierbar. Dann gibt es offene Umgebungen $V (\subseteq U)$ von a und V' von $f(a)$ in $\mathbb{R}^n$ derart, dass*

$$f : V \to V'$$

ein Diffeomorphismus ist.

Für den Beweis benötigen wir:

12.45. Bezeichnungen. Für $x, y \in \mathbb{R}^n$ sei $[x, y] = \{x + t(y - x) : 0 \le t \le 1\}$: Strecke zwischen x und y.

Für $\varepsilon > 0$ sei $\overline{B(x,\varepsilon)} = \{y \in \mathbb{R}^n : \|z - x\| \le \varepsilon\}$. Man kann dies als Definition nehmen; es ist jedoch auch leicht zu sehen, dass dies der *Abschluss* der Menge $B(x, \varepsilon)$ ist, d.h. die kleinste abgeschlossene Menge, die $B(x, \varepsilon)$ enthält.

Beweis von Satz 12.44. 1. Vereinfachung: Es genügt, den Satz für $a = f(a) = 0$ zu beweisen. Wir können nämlich statt f die durch $F(x) = f(x + a) - f(a)$ definierte Funktion F betrachten. Sie erfüllt $F(0) = 0$, und $F'(0)$ ist invertierbar. Man sieht leicht, dass die Invertierbarkeit von F zwischen zwei Umgebungen von 0 äquivalent ist zur Invertierbarkeit von f wie oben.

2. Vereinfachung: Wir können annehmen, dass $f'(0) = \mathrm{Id}$ ist. Sonst betrachten wir $F(x) = f'(0)^{-1}f(x)$. Ist $F : V_1 \to V_2$ invertierbar für Umgebungen V_1, V_2 von 0, so ist $f : V_1 \to V_2'$ invertierbar, wobei $V_2' = \{f'(0)y : y \in V_2\}$. Diese Menge ist offen nach 12.43, denn Multiplikation mit $f'(0)$ ist ein Diffeomorphismus.

3. Bijektivität: Für festes y in einer Umgebung von 0 müssen wir ein x finden mit $f(x) = y$. Wir definieren die Abbildung $F_y : U \to \mathbb{R}^n$ durch $F_y(x) = y + x - f(x)$. Dann ist $y = f(x) \Leftrightarrow F_y(x) = x$, d.h. wenn x ein Fixpunkt von F_y ist. Um dessen Existenz zu sichern, wenden wir den Banachschen Fixpunktsatz an.

[Erinnerung BFS: M abgeschl. Teilmenge eines Banachraums, $F : M \to M$ mit $\|F(x_1) - F(x_2)\| \le c\|x_1 - x_2\|$ für ein $c < 1$ und alle x_1, x_2. Dann existiert genau ein $x \in M$ mit $F(x) = x$.]

Nach dem Schrankensatz 12.28 gilt:

$$\|F_y(x_1) - F_y(x_2)\| \le \max_{u \in [x_1, x_2]} \|F_y'(u)\| \|x_1 - x_2\| \le \max_{u \in [x_1, x_2]} \|\mathrm{Id} - f'(u)\| \|x_1 - x_2\|.$$

Wir wählen daher nun $r > 0$ so klein, dass $\overline{B(0, 2r)} \subseteq U$ und $\|\mathrm{Id} - f'(u)\| < 1/2$ für alle $u \in \overline{B(0, 2r)}$ – das geht, weil $f'(0) = \mathrm{Id}$ und f' stetig ist.

Weil mit x_1 und x_2 auch die Strecke $[x_1, x_2]$ in $\overline{B(0, 2r)}$ liegt, folgt:

$$\|F_y(x_1) - F_y(x_2)\| \le \frac{1}{2}\|x_1 - x_2\|, \quad x_1, x_2 \in \overline{B(0, 2r)}.$$

Für $\|y\| < r$ und $x \in \overline{B(0, 2r)}$ ist dann wegen $y = F_y(0)$:

$$\|F_y(x)\| = \|y + (F_y(x) - F_y(0))\| \le \|y\| + \frac{1}{2}\|x\| < 2r. \tag{1}$$

Somit erfüllt $F_y : \overline{B(0, 2r)} \to \overline{B(0, 2r)}$ für jedes y mit $\|y\| < r$ die Voraussetzungen des Banachschen Fixpunktsatzes, hat also genau einen Fixpunkt in $\overline{B(0, 2r)}$. Mit anderen Worten: Für jedes $y \in B(0, r)$ existiert genau ein $x \in B(0, 2r)$ mit $f(x) = y$. Setzen wir also $V' = B(0, r)$ und $V = f^{-1}(V') \cap B(0, 2r)$, so sind dies offene Umgebungen von Null und

$$f : V \to V' \text{ ist bijektiv.}$$

Wir definieren $g : V' \to V$ als Umkehrabbildung.

4. Differenzierbarkeit von g in 0: Wegen $f(0) = 0$ und der Differenzierbarkeit von f in 0 mit Ableitung Id ist

$$f(x) = x + \varphi(x) \text{ mit } \varphi(x) = o(\|x\|).$$

Schreiben wir $g(y) = x$ so erhalten wir:

$$g(y) = y + \psi(y) \text{ mit } \psi(y) = -\varphi(f^{-1}(y)).$$

Nun zeigt (1) im obigen Beweis, dass aus $\|y\| \le r$ folgt, dass $\|x\| \le 2r$ ist. Da wir r beliebig verkleinern können, folgt, dass $\|f^{-1}(y)\| \le 2\|y\|$, falls $\|y\| \le r$, und somit

$$\frac{\|\psi(y)\|}{\|y\|} = \frac{\|\varphi(f^{-1}(y))\|}{\|f^{-1}(y)\|} \frac{\|f^{-1}(y)\|}{\|y\|} \to 0 \text{ für } \|y\| \to 0.$$

Daher ist g differenzierbar in 0 mit Ableitung $g'(0) = \mathrm{Id}$.

5. Differenzierbarkeit auf einer Umgebung. Nach Voraussetzung ist $\det f'(a) \neq 0$. Da f' stetig und die Determinante eine stetige Funktion ist, gibt es eine Umgebung von a, in der $\det f'(x) \neq 0$, somit $f'(x)$ invertierbar ist. Wir verkleinern ggf. V und V' so, dass f' auf ganz V invertierbar ist.

Für jedes x können wir den bisher bewiesenen Teil des Satzes anwenden. Wir schließen, dass f in einer Umgebung von x invertierbar ist und die Umkehrfunktion $g = f^{-1}$ (die wegen der Eindeutigkeit mit der obigen übereinstimmt) in $y = f(x)$ differenzierbar ist. Damit ist g auf ganz V' differenzierbar, insbesondere auch stetig. Die Identität $f^{-1}(f(x)) = x$ liefert $(f^{-1})'(f(x)) = f'(x)^{-1}$ bzw. $g'(y) = f'(g(y))$. Somit ist g' stetig und f ein Diffeomorphismus. $\triangleleft$

12.46. Beispiel. Polarkoordinaten in $\mathbb{R}^2$.

$$\begin{array}{rcl} f:\mathbb{R}_+\times\mathbb{R} & \to & \mathbb{R}^2\\ f(r,\varphi) & = & (r\cos\varphi, r\sin\varphi)\\ f'(r,\varphi) & = & \begin{pmatrix} \cos\varphi & -r\sin\varphi\\ \sin\varphi & r\cos\varphi\end{pmatrix}\\ \det f'(r,\varphi) & = & r\neq 0 \text{ für alle } r\in\mathbb{R}_+.\end{array}$$

Also ist f für jeden Wert von $(r,\varphi)\in\mathbb{R}_+\times\mathbb{R}$ lokal invertierbar; es ist

(1) $$(f^{-1})'(f(r,\varphi)) = (f'(r,\varphi))^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi\\ -\frac{\sin\varphi}{r} & \frac{\cos\varphi}{r}\end{pmatrix}.$$

Setzt man $x=f_1(r,\varphi)=r\cos\varphi$, $y=f_2(r,\varphi)=r\sin\varphi$, so ist $f^{-1}(r,\varphi)=(x,y)$, $r=\sqrt{x^2+y^2}$, $\cos\varphi=\frac{x}{r}$, $\sin\varphi=\frac{y}{r}$ und (1) schreibt sich

$$(f^{-1})'(x,y) = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}}\\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2}\end{pmatrix}.$$

Beachte: Die Abbildung f ist nicht global invertierbar: Für beliebiges $r>0$ ist

$$f(r,0)=(r,0)=f(r,2\pi).$$

12.47. Beispiel: Implizite Funktionen. Gegeben ist eine Funktion $F:U\to\mathbb{R}^m$ mit $U\subseteq\mathbb{R}^{m+k}$. Gesucht sind (möglichst alle) $(x,y)\in U$ mit $F(x,y)=0$. Man nennt dies eine 'implizite' Gleichung für x,y. Beispiel $F(x,y)=x^2+y^2-1$; man sucht alle $x,y\in\mathbb{R}$ mit $F(x,y)=0$. Wissen: Die Lösungsmenge ist die Einheitskreislinie.

Man kann $F(x,y)=0$ als eine Gleichung für y auffassen, die von einem Parameter x abhängt. Die intuitive Idee ist, dass man für jedes x nach y auflösen, d.h. dieses Gleichungssystem auch in der 'expliziten' Form $y=g(x)$ schreiben kann, mit einer geeigneten Funktion g, so dass $F(x,y)=0\Leftrightarrow y=g(x)$.

Das stimmt natürlich nicht immer: Ist etwa $F(x,y)=x^2+y^2-1$, so gibt es für $|x|>1$ keine Lösung von $F(x,y)=0$ in $\mathbb{R}$, für $|x|<1$ gibt es zwei Lösungen, nämlich $y(x)=\pm\sqrt{1-x^2}$, nur für $x=\pm1$ ist $y=0$ eindeutig.

Aber: Diese Überlegung zeigt auch, dass für $|x|<1$ die Gleichung $F(x,y)=0$ in $\mathbb{R}_+$ (d.h. für $y>0$) die eindeutige Lösung $y=\sqrt{1-x^2}$ in der gesuchten Form hat, diese hängt sogar stetig differenzierbar von x ab. Durch Einschränkung des Definitionsbereichs von F erhalten wir also Eindeutigkeit der Lösung. Dies macht der folgende Satz präzise:

12.48. Satz (Satz über die implizite Funktion). *Es sei $U\subseteq\mathbb{R}^k\times\mathbb{R}^m$ offen und $F:U\to\mathbb{R}^m$ stetig differenzierbar.*

Für einen Punkt $(a,b)\in U$ sei $F(a,b)=0$ und die $m\times m$-Matrix $\frac{\partial F}{\partial y}(a,b)$ invertierbar. Dann gibt es offene Umgebungen V_1 von a, V_2 von b und eine stetig differenzierbare Abbildung $g:V_1\to V_2$ mit $F(x,g(x))=0$.

Ferner gilt: Ist $(x,y)\in V_1\times V_2$ mit $F(x,y)=0$, so ist $y=g(x)$ (d. h. die Gleichung ist dort eindeutig nach y auflösbar) und

(1) $$\frac{\partial g}{\partial x}(x) = -\left(\frac{\partial F}{\partial y}(x,g(x))\right)^{-1}\frac{\partial F}{\partial x}(x,g(x)).$$

Beweis. Wir betrachten die Abbildung $f : U \to \mathbb{R}^k \times \mathbb{R}^m$ mit $(x, y) \mapsto (x, F(x, y))$. Es ist

$$f'(x,y) = \begin{pmatrix} \mathrm{Id} & 0 \\ \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial y}(x,y) \end{pmatrix}.$$

Nach dem Kästchensatz ist also $f'(a, b)$ invertierbar. Der Satz über die lokale Invertierbarkeit liefert Umgebungen W von (a, b) und W' von $(a, F(a, b)) = (a, 0)$, so dass $f : W \to W'$ ein Diffeomorphismus ist. OBdA: $W = V_0 \times V_2$ für offene Umgebungen V_0 von a und V_2 von b. Die Umkehrabbildung $f^{-1} : W' \to W$ hat die Form $f^{-1}(u, v) = (u, H(u, v))$ mit einer geeigneten Abbildung $H : W' \to V_2$. Mit f^{-1} ist auch H stetig differenzierbar.

Da W' offene Umgebung von $(a, 0)$ ist, finden wir eine Umgebung V_1 von a mit $V_1 \subseteq V_0$ und $V_1 \times \{0\} \subseteq W'$. Es sei $(x, y) \in V_1 \times V_2 \subseteq W$. Dann ist $x \in V_1$, also $(x, 0) \subset W'$. Somit ist $f^{-1}(x, 0)$ definiert, also auch $H(x, 0)$. Ferner

$$F(x, y) = 0 \Leftrightarrow f(x, y) = (x, 0) \Leftrightarrow (x, y) = f^{-1}(x, 0) \Leftrightarrow y = H(x, 0).$$

Wir können daher $g : V_1 \to V_2$ durch $g(x) = H(x, 0)$ definieren. Für $(x, y) \in V_1 \times V_2$ ist dann $F(x, y) = 0 \Leftrightarrow g(x) = y$. Ferner ist g stetig differenzierbar. Aus der Gleichung

$$F(x, g(x)) = 0$$

folgt (1) durch Differenzieren mit der Kettenregel:

$$\frac{\partial F}{\partial x}(x, g(x)) + \frac{\partial F}{\partial y}(x, g(x))\frac{\partial g}{\partial x}(x) = 0.$$

$\triangleleft$

12.49. Beispiel.

(a) Wir betrachten wieder $F(x, y) = x^2 + y^2 - 1$ $(k = m = 1)$. Es ist $\frac{\partial F}{\partial x} = 2x$, $\frac{\partial F}{\partial y} = 2y$. Also ist $\frac{\partial F}{\partial x}(x, y) \neq 0$ für alle (x, y) mit $x \neq 0$ und $\frac{\partial F}{\partial y} \neq 0$ für alle (x, y) mit $y \neq 0$.

Also: Sind $a, b \neq 0$ und ist $F(a, b) = a^2 + b^2 - 1$, so ist die Gleichung $F(x, y) = 0$ in einer Umgebung von (a, b) sowohl nach x als auch nach y auflösbar.

Ist $a = 0,\ b \neq 0$, mit $F(a, b) = 0$, d. h. $b = \pm 1$, so ist die Gleichung nahe (a, b) nach y auflösbar, über Auflösbarkeit nach x sagt der Satz nichts (geht auch nicht).

Analog ist für $a \neq 0,\ b = 0$ mit $F(a, b) = 0$ die Gleichung nahe (a, b) nach x auflösbar, keine Aussage über Auflösbarkeit nach y (geht nicht).

(b) Wir betrachten die Funktion $F(x, y) = x^y + y^x - 2$. Für $(a, b) = (1, 1)$ ist $F(1, 1) = 0$. Hier ist $\partial_y F(x, y) = x^y \ln x + xy^{x-1}$, also $\partial_y F(1, 1) = 1$. Somit wissen wir, dass die Menge $F(x, y) = 0$ in der Form $y = g(x)$ darstellbar ist, ohne dass wir die Funktion g explizit hinschreiben könnten (zumindest wüsste ich nicht, wie).

Extrema unter Nebenbedingungen.

12.50. Erinnerung an Schulaufgaben. Sie haben 400 m Weidezaun und sollen damit eine möglichst große rechteckige Weide einzäunen.Wie seiht die optimale Lösung aus?

Wir bezeichnen mit x und y die Seitenlängen des Rechtecks. Die Fläche F der Weide ist dann

$$F = xy.$$

Die Tatsache, dass 400 m Zaun zur Verfügung stehen (der auch genutzt werden sollte), bedeutet, dass für den Umfang U des Rechtecks gilt:

$$U = 400 = 2x + 2y.$$

Aus der Nebenbedingung schließen wir, dass $y = 200 - x$. Einsetzen in die Zielfunktion liefert

$$F = x(200 - x).$$

Zur Suche nach dem Extremum setzt man die Ableitung Null. Das liefert

$$200 - 2x = 0, \text{ bzw. } x = 100.$$

Aus der Nebenbedingung folgt dann $y = 100$. Die Weide ist also quadratisch mit der Seitenlänge 100 m.

Wie sehen wir das mit heutigen Augen? Wir haben die Funktion

$$F : [0, \infty[\times[0, \infty[\to \mathbb{R}, \qquad F(x, y) = xy.$$

die kein (lokales) Maximum annimmt. Der Gradient ist $(1, 1)$, also nirgends Null.

Die Nebenbedingung schränkt den Definitionsbereich auf

$$M = \{(x, y) \in [0, \infty[\times[0, \infty[: 2x + 2y - 400 = 0\}$$

ein. Diese Menge ist abgeschlossen und beschränkt, also nach Heine-Borel kompakt. Also hat F dort ein Maximum. Allerdings ist es nicht so einfach zu finden; das Kriterium mit dem Gradienten nützt uns nichts.

Was hilft, ist, dass wir die Nebenbedingung $g(x, y) = 2x + 2y - 400 = 0$, eine implizite Funktion, nach y auflösen können:

$$2x + 2y - 400 = 0 \Leftrightarrow y = h(x) = 200 - x.$$

Wir können nun also die Funktion

$$\tilde{F} : [0, 400] \to \mathbb{R}, \qquad \tilde{F}(x) = F(x, h(x)) = xh(x) = x(200 - x)$$

betrachten, die auf einem Intervall definiert ist. Das Maximum wird im Innern angenommen. Also gilt dort $\tilde{F}'(x) = 0$, somit nach der Kettenregel:

$$0 = \frac{d}{dx}\tilde{F}(x) = \frac{\partial F}{\partial x}(x, h(x)) + \frac{\partial F}{\partial y}(x, h(x))h'(x).$$

Im speziellen Fall liefert Auflösen dieser Gleichung nach x unmittelbar den x-Wert der Maximalstelle. Aus der Nebenbedingung erhalten wir den y-Wert durch $y = h(x)$.

Im Allgemeinen wird das etwas, aber nicht viel komplizierter, wie wir nun sehen werden.

12.51. Satz (Extrema unter Nebenbedingungen). *Es sei $U \subseteq \mathbb{R}^n$ offen, $g : U \to \mathbb{R}^k$ stetig differenzierbar,*

$$M = \{x \in U : g(x) = 0\}$$

und $a = (a_1, \ldots, a_n) \in M$ ein Punkt mit $\mathrm{Rang}(g'(a)) = k$ (also maximal). Ferner sei $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion, die in a ein lokales Extremum unter der Nebenbedingung $g(x) = 0$ annimmt (das heißt, es gibt eine Umgebung U von a derart, dass entweder $f(a) \leq f(x)$ für alle $x \in M \cap U$ oder $f(a) \geq f(x)$ für alle $x \in M \cap U$).

Dann existiert $\lambda = (\lambda_1, \ldots, \lambda_k)$ (λ_k: "Lagrangesche Multiplikatoren") so, dass

$$\mathrm{grad}\, f(a) = \lambda\, g'(a).$$

Wir erhalten zusammen mit der Nebenbedingungs-Gleichung $g(x) = 0$ ein System von $n + k$ Gleichungen für die $n + k$ Variablen $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k$, das zu lösen ist.

Beweis. Es gibt nach Annahme k Variablen $x_{l_1}, \ldots, x_{l_k}$ derart, dass die Matrix

$$\left(\frac{\partial g_i}{\partial x_{l_j}}\right)_{\substack{i=1,\ldots,k \\ j=1,\ldots,k}}$$

invertierbar ist. Nach Umnummerieren können wir annehmen, dass dies $x_1. \ldots, x_k$ sind.

Nach dem Satz von der impliziten Funktion 12.48 können wir nach $x_1, \dots, x_k$ auflösen, d. h. es existiert

- eine Umgebung V'' von $(a_{k+1}, \dots, a_n) \subseteq \mathbb{R}^{n-k}$,
- eine Umgebung V' von $(a_1, \dots, a_k) \subseteq \mathbb{R}^k$, und
- eine stetig differenzierbare Funktion $h : V'' \to V'$ derart, dass
$$g(x) = 0 \Leftrightarrow (x_1, \dots, x_k) = h(x_{k+1}, \dots, x_n)$$
für alle $x' = (x_1, \dots, x_k) \in V', x'' = (x_{k+1}, \dots, x_n) \in V''$.

Wir definieren nun $F : V'' \to \mathbb{R}$ durch

$$F(x'') = f(h(x''), x''). \tag{1}$$

Da $(h(x''), x'')$ in M liegt, hat F in a'' ein lokales Extremum. Also ist

$$\operatorname{grad} F(a'') = 0.$$

Dies liefert nach (1)

$$\frac{\partial f}{\partial x'}(h(a''), a'')\frac{\partial h}{\partial x''}(a'') + \frac{\partial f}{\partial x''}(h(a''), a'') = 0. \tag{2}$$

Andererseits ist $g(h(x''), x'') = 0$, also durch Ableiten nach x'':

$$\frac{\partial g}{\partial x'}(h(x''), x'')\frac{\partial h}{\partial x''}(x'') + \frac{\partial g}{\partial x''}(h(x''), x'') = 0. \tag{3}$$

Gleichungen (2) und (3) liefern

$$-\frac{\partial f}{\partial x'}(a)\left(\frac{\partial g}{\partial x'}(a)\right)^{-1}\frac{\partial g}{\partial x''}(a) + \frac{\partial f}{\partial x''}(a) = 0. \tag{4}$$

Setzen wir also $\lambda = \frac{\partial f}{\partial x'}(a)\left(\frac{\partial g}{\partial x'}(a)\right)^{-1}$, dann gilt

$$\begin{aligned}
\lambda\frac{\partial g}{\partial x}(a) &= \lambda\left(\frac{\partial g}{\partial x'}(a), \frac{\partial g}{\partial x''}(a)\right) \\
&= \frac{\partial f}{\partial x'}(a)\left(\frac{\partial g}{\partial x'}(a)\right)^{-1}\left(\frac{\partial g}{\partial x'}(a), \frac{\partial g}{\partial x''}(a)\right) \\
&\overset{(4)}{=} \left(\frac{\partial f}{\partial x'}(a), \frac{\partial f}{\partial x''}(a)\right) = \operatorname{grad} f(a).
\end{aligned}$$

◁

12.52. Beispiel. $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x, y, z) = x + y + z$, $g : \mathbb{R}^3 \to \mathbb{R}^2$, $g(x, y, z) = (x^2 + y^2 - 1, z)$.
Klar: f, g sind stetig differenzierbar

$$g'(x, y, z) = \begin{pmatrix} 2x & 2y & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Menge $M = \{(x, y, z) : g(x, y, z) = 0\}$ ist kompakt, also hat f auf M ein Maximum und ein Minimum. Mit a bezeichnen wir eines dieser Extrema.

Auf M hat g' den Rang 2, da wegen $x^2 + y^2 = 1$ stets $x \neq 0$ oder $y \neq 0$. Nach 12.51 existiert $\lambda = (\lambda_1, \lambda_2)$ mit

$$\operatorname{grad} f(a) = \lambda\, g'(a)$$

also

$$(1, 1, 1) = (\lambda_1 \cdot 2x, \lambda_1 \cdot 2y, \lambda_2).$$

Es folgt

$$
\begin{array}{llcl}
(1) & 2x\lambda_1 & = & 1 \\
(2) & 2y\lambda_1 & = & 1 \\
(3) & \lambda_2 & = & 1.
\end{array}
$$

Ferner ist wegen der Nebenbedingung

$$
\begin{array}{llcl}
(4) & x^2+y^2 & = & 1 \\
(5) & z & = & 0.
\end{array}
$$

Wegen (1) ist $\lambda_1 \neq 0$, (1) und (2) liefern dann $x = y$, (4) liefert dann $x = y = \pm\frac{1}{2}\sqrt{2}$, (5) liefert $z = 0$. Wir haben also zwei Kandidaten: $a_1 = (\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}, 0)$ und $a_2 = (-\frac{1}{2}\sqrt{2}, -\frac{1}{2}\sqrt{2}, 0)$. Einsetzen liefert $f(a_1) = \sqrt{2}$, $f(a_2) = -\sqrt{2}$. Somit ist a_1 das Maximum und a_2 das Minmum von f unter der Nebenbedingung g.

12.53. Definition. Für eine Funktion $g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ heißen die Mengen

$$N_c(g) = \{x \in U : g(x) = c\}$$

die Niveaumengen zum Niveau c von g.

12.54. Lemma. *Ist $g : U \to \mathbb{R}^n$ differenzierbar, $g(x) = c$, so steht* $\operatorname{grad} g(x)$ *senkrecht auf $N_c(g)$ in x und zwar in folgendem Sinne:*

Für jede differenzierbare Kurve $\alpha :]-\varepsilon, \varepsilon[\to \mathbb{R}^n$ mit $\alpha(]-\varepsilon, \varepsilon[) \subseteq N_c(g)$ und $\alpha(0) = x$ ist $\langle \operatorname{grad} g(x), \alpha'(0)\rangle = 0$.

[Diese Bedingung heißt eigentlich, dass der Gradient in x senkrecht auf dem Tangentialraum an $N_c(f)$ steht. Dazu mehr in Analysis 3.]

Beweis. $g \circ \alpha :]-\varepsilon, \varepsilon[\to \mathbb{R}$ ist differenzierbar mit $g \circ \alpha(t) = c$ für alle $t \in]-\varepsilon, \varepsilon[$. Ableiten in 0 liefert:

$$0 = g'(\alpha(0)) \circ \alpha'(0) = \langle \operatorname{grad} g(x), \alpha'(0)\rangle.$$

$\lhd$

12.55. Geometrischer 'Beweis' des Satzes 12.51 für k=1. Die Menge M ist eine Niveaumenge für g (nämlich zum Niveau 0). Der Gradient von g in a steht also senkrecht auf M im Sinne von 12.54. Nach Annahme ist er nicht 0.

Nun weiß man (das werden wir noch sehen), dass die Tangentialvektoren an M in x im Sinne von 12.54 (d.h. die Ableitungsvektoren $\alpha'(0)$ aller solcher Kurven) einen $(n-1)$-dimensionalen Vektorraum aufspannen.

Aus Dimensionsgründen ist also die Existenz einer Zahl λ mit

$$(1) \qquad \operatorname{grad} f(a) = \lambda \operatorname{grad} g(a)$$

äquivalent dazu, dass $\operatorname{grad} f(a)$ in a auf M senkrecht steht.

Übrigens: Stünde $\operatorname{grad} f(a)$ nicht senkrecht auf M, so gäbe es eine differenzierbare Kurve $\alpha :]-\varepsilon, \varepsilon[\to \mathbb{R}^n$ mit $\alpha(t) \in M$ für alle t, $\alpha(0) = a$ und

$$(f \circ \alpha)'(0) = \langle \operatorname{grad} f(a), \alpha'(0)\rangle \neq 0.$$

Damit wäre jedoch $f \circ \alpha$ entweder streng monoton wachsend oder streng monoton fallend und daher a kein Extremum unter der Nebenbedingung M.