Bergische Universität – Gesamthochschule Wuppertal – Fachbereich Mathematik

Analysis 2

Kapitel 2 Differentialrechnung im $\mathbb{R}^n$

Vorlesungsausarbeitung zum SS 2001

von Prof. Dr. Klaus Fritzsche

Inhaltsverzeichnis

§ 1 Totale Differenzierbarkeit

Erinnern wir uns an die Differenzierbarkeit in einer Veränderlichen:

Es sei $I \subset \mathbb{R}$ ein Intervall. Eine Funktion $f : I \to \mathbb{R}^k$ heißt differenzierbar in $t_0 \in I$, falls eine der folgenden äquivalenten Aussagen erfüllt ist:

1. Es existiert der Grenzwert $\displaystyle\lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}$.

2. Es gibt einen Vektor $\mathbf{a} \in \mathbb{R}^k$ und eine Funktion $\varphi : I \to \mathbb{R}^k$, mit

$$f(t) = f(t_0) + (t - t_0) \cdot \mathbf{a} + \varphi(t) \quad \text{auf } I$$

und

$$\lim_{t \to t_0} \frac{\varphi(t)}{t - t_0} = 0.$$

3. Es gibt eine in t_0 stetige Abbildung $\Delta : I \to \mathbb{R}^k$ mit

$$f(t) = f(t_0) + (t - t_0) \cdot \Delta(t).$$

Unter der Ableitung von f in t_0 versteht man den Vektor

$$f'(t_0) := \mathbf{a} = \Delta(t_0) = \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}.$$

Was bedeutet die Differenzierbarkeit von f in t_0 anschaulich? Im Falle $k > 1$ kann man sich f als Parametrisierung einer Kurve C vorstellen, die im Punkte $f(t_0)$ so „glatt“ ist, daß sie dort eine eindeutig bestimmte Tangente besitzt. Was ist aber nun eine Tangente? Das ist eine Gerade, die C in $f(t_0)$ „berührt“. Also bleibt die Frage: Was bedeutet es, daß sich zwei Kurven in einem Punkt berühren?

Damit sich zwei Kurven $\alpha, \beta : I \to \mathbb{R}^n$ in einem Punkt $\mathbf{x}_0$ berühren, muß es sicherlich ein $t_0 \in I$ geben, so daß $\alpha(t_0) = \beta(t_0) = \mathbf{x}_0$ ist. Aber das reicht nicht! Wir erwarten, daß sich die beiden Kurven bei $\mathbf{x}_0$ regelrecht aneinander anschmiegen. Der Ausdruck $\|\alpha(t) - \beta(t)\|$ muß für $t \to t_0$ hinreichend schnell gegen Null streben. Im Falle $k = 1$ wissen wir, daß der Graph der Funktion $t \mapsto |t - t_0|$ bei t_0 einen Knick und zwei verschiedene Tangenten hat. Ist aber φ eine Funktion mit $\varphi(t) \to 0$ für $t \to t_0$, so streben alle Differenzenquotienten von $(t - t_0)\varphi(t)$ für $t \to t_0$ gegen Null, und die x-Achse ist die eindeutig bestimmte Tangente an den Graphen. Deshalb sagt man, daß sich α und β in $\mathbf{x}_0$ berühren, falls eine Funktion φ mit Werten im $\mathbb{R}^k$ existiert, so daß $\alpha(t) - \beta(t) = (t - t_0) \cdot \varphi(t)$ ist und $\varphi(t)$ für $t \to t_0$ gegen Null konvergiert.

Nun brauchen wir noch die Parametrisierung der Tangente an die durch f parametrisierte Kurve. Dies geschieht durch

$$T_f(t) := f(t_0) + (t - t_0) \cdot \mathbf{a},$$

wobei $\mathbf{a}$ der Richtungsvektor und $f(t_0)$ der „Stützvektor" ist. Die Parametrisierung ist gerade so gemacht, daß $T_f(t_0) = f(t_0)$ ist. Die obige Bedingung (2) entspricht also genau der anschaulichen Vorstellung, daß f in t_0 differenzierbar ist, wenn die durch f parametrisierte Kurve in $f(t_0)$ von einer (eindeutig bestimmten) Geraden berührt wird. Im Falle $k = 1$ können wir $f : I \to \mathbb{R}$ durch die Parametrisierung $F(t) := (t, f(t))$ des Graphen von f ersetzen. Dann parametrisiert F eine Kurve im $\mathbb{R}^2$, wir behalten die Vorstellung von der Tangente an eine Kurve und f ist natürlich genau dann differenzierbar, wenn F es ist.

Die Parametrisierung T_f der Tangente (im Falle $k > 1$) ist eine affin-lineare Abbildung der Gestalt

$$T_f(t_0 + h) = f(t_0) + L(h),$$

mit einer linearen Abbildung $L : \mathbb{R} \to \mathbb{R}^k$, die durch $L(h) = h \cdot \mathbf{a}$ gegeben ist. So ist (2) äquivalent zu folgender Aussage:

1. $f(t) = f(t_0) + L(t - t_0) + \varphi(t)$ für $t \in I$.
2. $\lim\limits_{t \to t_0} \dfrac{\varphi(t)}{|t - t_0|} = 0.$

Ist nun $f = f(x_1, \ldots, x_n)$ eine Funktion von mehreren Variablen mit Werten in $\mathbb{R}^k$ und $k > n$, so parametrisiert f ein n-dimensionales Gebilde (wir werden das am Ende des Kapitels präzisieren). Wir werden dann f in $\mathbf{y}_0 = f(\mathbf{x}_0)$ differenzierbar nennen, wenn das parametrisierte Gebilde in $\mathbf{y}_0$ so glatt ist, daß es dort eine eindeutig bestimmte Tangentialebene der Dimension n besitzt. Ist $k \le n$ (z.B. im Falle einer skalaren Funktion f), so behelfen wir uns wieder mit dem Graphen, der durch $F(\mathbf{x}) = (\mathbf{x}, f(\mathbf{x}))$ parametrisiert wird.

Auch im Falle einer Funktion $f : \mathbb{R}^n \to \mathbb{R}^k$ (mit $n < k$) wird die Tangentialebene durch eine affin-lineare Abbildung parametrisiert:

$$T_f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}),$$

diesmal mit einer linearen Abbildung $L : \mathbb{R}^n \to \mathbb{R}^k$.

Deshalb beschäftigen wir uns zunächst einmal mit linearen Abbildungen.

Ist $L : \mathbb{R}^n \to \mathbb{R}^k$ linear, so gibt es eine Matrix $A \in M_{k,n}(\mathbb{R})$, so daß gilt:

$$L(\mathbf{v}) = \mathbf{v} \cdot A^t.$$

Man beachte, daß wir hier immer mit Zeilen-Vektoren arbeiten! Schreiben wir $L = L_A$, so ist

$$L_{A \cdot B}(\mathbf{v}) = \mathbf{v} \cdot (A \cdot B)^t = \mathbf{v} \cdot (B^t \cdot A^t) = L_B(\mathbf{v}) \cdot A^t = L_A(L_B(\mathbf{v})),$$

also

$$L_{A \cdot B} = L_A \circ L_B.$$

Sind E, F zwei (endlich-dimensionale) $\mathbb{R}$-Vektorräume, so bezeichnen wir mit $L(E, F)$ den Vektorraum aller linearen Abbildungen von E nach F. Die Zuordnung $A \mapsto L_A$ liefert einen Vektorraum-Isomorphismus

$$M_{k,n}(\mathbb{R}) \to L(\mathbb{R}^n, \mathbb{R}^k)\,.$$

Wir betrachten noch zwei Spezialfälle:

a) Ist F beliebig, so wird durch $L \mapsto L(1)$ eine lineare Abbildung $L(\mathbb{R}, F) \to F$ definiert. Tatsächlich ist dies ein Isomorphismus, die Umkehrabbildung ordnet einem Vektor $a \in F$ die lineare Abbildung $t \mapsto t \cdot a$ zu. Man spricht hier auch vom *kanonischen Isomorphismus* $L(\mathbb{R}, F) \cong F$.

b) Die Elemente von $L(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$ sind die *Linearformen* auf dem $\mathbb{R}^n$. Ist $\mathbf{a} \in \mathbb{R}^n$, so wird eine Linearform $\lambda_{\mathbf{a}}$ auf dem $\mathbb{R}^n$ definiert durch

$$\lambda_{\mathbf{a}}(\mathbf{x}) := \mathbf{x} \bullet \mathbf{a} = \mathbf{x} \cdot \mathbf{a}^{\,t}\,.$$

Offensichtlich ist dann $\mathbf{a} \in M_{1,n}(\mathbb{R})$ auch schon die Matrix, die $\lambda_{\mathbf{a}}$ beschreibt.

Die Zuordnung $\mathbf{a} \mapsto \lambda_{\mathbf{a}}$ liefert einen Isomorphismus von $\mathbb{R}^n$ auf $L(\mathbb{R}^n, \mathbb{R})$. Die Umkehrung ist gegeben durch

$$\lambda \mapsto (\lambda(\mathbf{e}_1), \ldots, \lambda(\mathbf{e}_n)).$$

In diesem Fall ist der Isomorphismus $\mathbb{R}^n \cong L(\mathbb{R}^n, \mathbb{R})$ bestimmt durch das euklidische Skalarprodukt und die dazu passende Orthonormalbasis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Man kann für jeden anderen endlich-dimensionalen Vektorraum E (mit einem Skalarprodukt $<\ldots, \ldots>$) einen Isomorphismus $E \to L(E, \mathbb{R}) = E^*$ definieren durch

$$a \mapsto \lambda_a : (x \mapsto <x\,,\, a>)\,.$$

Sei nun F ein beliebiger endlich-dimensionaler $\mathbb{R}$-Vektorraum mit Basis $\{z_1, \ldots, z_m\}$. Ist $B \subset \mathbb{R}^n$ offen und $f : B \to F$ eine Abbildung, so kann man f in der Form

$$f(\mathbf{x}) = \sum_{\nu=1}^{m} f_\nu(\mathbf{x}) \cdot z_\nu$$

schreiben, wobei die Koeffizienten $f_\nu : B \to \mathbb{R}$ skalare Funktionen sind. Ist f Einschränkung einer linearen Abbildung $f : \mathbb{R}^n \to F$, so sind die f_ν Linearformen auf $\mathbb{R}^n$.

Die Abbildung f ist genau dann in $\mathbf{x}_0 \in B$ stetig, wenn alle Koeffizienten f_ν in $\mathbf{x}_0$ stetig sind.

Beispiele.

1. Sei $F = L(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$. Zu den Einheitsvektoren $\mathbf{e}_i \in \mathbb{R}^n$ gehören die Linearformen $\lambda_{\mathbf{e}_i} = \varepsilon^i \in (\mathbb{R}^n)^*$ mit

$$\varepsilon^i(\mathbf{x}) = \mathbf{x} \cdot \mathbf{e}_i^t = x_i, \text{ also } \varepsilon^i(\mathbf{e}_j) = \delta_{ij}.$$

Man bezeichnet $\{\varepsilon^1, \dots, \varepsilon^n\}$ auch als *duale Basis* (zur Standard-Basis) von $(\mathbb{R}^n)^*$. Jede Linearform

$$\lambda : (x_1, \dots, x_n) \mapsto a_1 x_1 + \cdots + a_n x_n$$

kann in der Form $\lambda = a_1\varepsilon^1 + \cdots + a_n\varepsilon^n$ geschrieben werden.

Eine Abbildung $f = f_1\varepsilon^1 + \cdots + f_n\varepsilon^n : B \to L(\mathbb{R}^n, \mathbb{R})$ ist genau dann stetig, wenn alle f_i stetig sind.[1] Dabei ist $f_i(\mathbf{x}) = f(\mathbf{x})(\mathbf{e}_i)$. Deshalb ist f auch genau dann stetig, wenn für jeden Vektor $\mathbf{v}$ die Abbildung $\mathbf{x} \mapsto f(\mathbf{x})(\mathbf{v})$ stetig ist.

Ist – etwas allgemeiner – $F = L(\mathbb{R}^n, \mathbb{R}^k)$, so gilt das gleiche Kriterium: $f : B \to F$ ist genau dann stetig, wenn $f(\mathbf{x})(\mathbf{v})$ für jeden Vektor $\mathbf{v}$ stetig in $\mathbf{x}$ ist. Das ist jeweils eine Abbildung von B nach $\mathbb{R}^k$.

2. Sei $F = M_{k,n}(\mathbb{R})$. Dann bilden die Matrizen $E_{\nu\mu}$, die aus einer 1 an der Stelle (ν, μ) und lauter Nullen sonst bestehen, eine Basis von F. Eine Abbildung $f : B \to M_{k,n}(\mathbb{R})$ hat dann die Gestalt

$$f(\mathbf{x}) = \sum_{\nu,\mu} f_{\nu\mu}(\mathbf{x}) \cdot E_{\nu\mu} = \begin{pmatrix} f_{11}(\mathbf{x}) & \cdots & f_{1n}(\mathbf{x}) \\ \vdots & & \vdots \\ f_{k1}(\mathbf{x}) & \cdots & f_{kn}(\mathbf{x}) \end{pmatrix}.$$

Analog zu der Situation bei den linearen Abbildungen gilt hier: f ist genau dann stetig, wenn die Abbildung $\mathbf{x} \mapsto \mathbf{v} \cdot f(\mathbf{x})^t$ für jeden Vektor $\mathbf{v}$ stetig ist.

1.1 Satz. *Sei $B \subset \mathbb{R}^n$ offen, $\mathbf{x}_0 \in B$ und $f : B \to \mathbb{R}^m$ eine Abbildung. Dann sind die folgenden Aussagen äquivalent:*

1. *Es gibt eine lineare Abbildung $L : \mathbb{R}^n \to \mathbb{R}^k$, so daß gilt:*

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0\,.$$

2. *Es gibt eine lineare Abbildung $L : \mathbb{R}^n \to \mathbb{R}^k$ und eine Abbildung $\varphi : B \to \mathbb{R}^k$, so daß gilt:*

   (a) $f(\mathbf{x}) = f(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0) + \varphi(\mathbf{x})$.

[1] Wir kennen solche Abbildungen f schon in der Gestalt Pfaffscher Formen $f = f_1\,dx_1 + \cdots + f_n\,dx_n$.

(b) $\lim_{\mathbf{x} \to \mathbf{x}_0} \dfrac{\varphi(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$

3. Es gibt eine in $\mathbf{x}_0$ *stetige Abbildung* $\delta : B \to L(\mathbb{R}^n, \mathbb{R}^k)$ *mit*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \delta(\mathbf{x})(\mathbf{x} - \mathbf{x}_0)\,.$$

4. Es gibt eine in $\mathbf{x}_0$ *stetige Abbildung* $\Delta : B \to M_{k,n}(\mathbb{R})$ *mit*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^{\,t}\,.$$

Beweis:

(1) $\Longrightarrow$ (2): Setze $\varphi(\mathbf{x}) := f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)$. Nach Voraussetzung folgt:

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\varphi(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}_0\|} = 0\,.$$

(2) $\Longrightarrow$ (3): Für $\mathbf{x} \in B$ sei $\delta(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^k$ definiert durch

$$\delta(\mathbf{x})(\mathbf{v}) := L(\mathbf{v}) + \frac{(\mathbf{x} - \mathbf{x}_0) \bullet \mathbf{v}}{\|\mathbf{x} - \mathbf{x}_0\|^2} \cdot \varphi(\mathbf{x}), .$$

Offensichtlich ist $\delta(\mathbf{x})$ linear und

$$\delta(\mathbf{x})(\mathbf{x} - \mathbf{x}_0) = L(\mathbf{x} - \mathbf{x}_0) + \varphi(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}_0)\,.$$

Sei $F(\mathbf{h}, \mathbf{v}) := \dfrac{\mathbf{h} \bullet \mathbf{v}}{\|\mathbf{h}\|^2} \cdot \varphi(\mathbf{x}_0 + \mathbf{h})$. Mit Cauchy-Schwarz folgt, daß $\|F(\mathbf{h}, \mathbf{v})\|$ bei festem $\mathbf{v}$ für $\mathbf{h} \to \mathbf{0}$ gegen Null konvergiert, also $\delta(\mathbf{x})(\mathbf{v})$ für $\mathbf{x} \to \mathbf{x}_0$ gegen $L(\mathbf{v})$. Das bedeutet, daß δ stetig in $\mathbf{x}_0$ ist.

(3) $\Longrightarrow$ (4): Sei $\Delta(\mathbf{x}) \in M_{k,n}(\mathbb{R})$ die Matrix mit mit $\delta(\mathbf{x})(\mathbf{v}) = \mathbf{v} \cdot \Delta(\mathbf{x})^{\,t}$. Dann ist

$$(\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^{\,t} = L(\mathbf{x} - \mathbf{x}_0) + \varphi(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}_0)\,,$$

und Δ ist stetig in $\mathbf{x}_0$.

(4) $\Longrightarrow$ (1): Ist Δ gegeben, so setzen wir $L(\mathbf{v}) := \mathbf{v} \cdot \Delta(\mathbf{x}_0)^{\,t}$. Dann ist

$$f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0) = (\mathbf{x} - \mathbf{x}_0) \cdot (\Delta(\mathbf{x})^{\,t} - \Delta(\mathbf{x}_0)^{\,t})\,.$$

Also strebt

$$\frac{f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \frac{\mathbf{x} - \mathbf{x}_0}{\|\mathbf{x} - \mathbf{x}_0\|} \cdot (\Delta(\mathbf{x})^{\,t} - \Delta(\mathbf{x}_0)^{\,t})$$

gegen Null, für $\mathbf{x} \to \mathbf{x}_0$. ∎

Definition. Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}^k$ eine Abbildung.

f heißt in $\mathbf{x}_0 \in B$ *(total) differenzierbar*, falls eine der äquivalenten Bedingungen des Satzes erfüllt ist.

Die lineare Abbildung $Df(\mathbf{x}_0) := L = \delta(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^k$ nennt man die *Ableitung* von f in $\mathbf{x}_0$, die Matrix $J_f(\mathbf{x}_0) := \Delta(\mathbf{x}_0) \in M_{k,n}(\mathbb{R})$ die *Funktionalmatrix* oder *Jacobi-Matrix* von f in $\mathbf{x}_0$

1.2 Satz. *Ist f in $\mathbf{x}_0$ differenzierbar, so ist die Ableitung in $\mathbf{x}_0$ eindeutig bestimmt, und $J_f(\mathbf{x}_0)$ ist die Matrix, die zu $Df(\mathbf{x}_0)$ gehört.*

BEWEIS: Gibt es zwei Darstellungen

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \delta_1(\mathbf{x})(\mathbf{x} - \mathbf{x}_0) = f(\mathbf{x}_0) + \delta_2(\mathbf{x})(\mathbf{x} - \mathbf{x}_0)$$

mit in $\mathbf{x}_0$ stetigen Abbildungen $\delta_1, \delta_2 : B \to L(\mathbb{R}^n, \mathbb{R}^k)$, so folgt:

$$(\delta_1 - \delta_2)(\mathbf{x}_0 + t\mathbf{v})(\mathbf{v}) = \mathbf{0} \text{ für } t \neq 0 \text{ und jedes } \mathbf{v}.$$

Wegen der Stetigkeit von δ_1 und δ_2 in $\mathbf{x}_0$ erhält man dann $\delta_1(\mathbf{x}_0) = \delta_2(\mathbf{x}_0)$. Die zweite Behauptung ist trivial. ∎

1.3 Satz. *Ist f in $\mathbf{x}_0$ differenzierbar, so ist f dort auch stetig.*

BEWEIS: Mit Eigenschaft (3) folgt die Behauptung sofort. ∎

1.4 Satz. *Ist $f : B \to \mathbb{R}$ in $\mathbf{x}_0$ differenzierbar, so ist f dort in Richtung jedes Vektors $\mathbf{v}$ differenzierbar, und es ist*

$$Df(\mathbf{x}_0)(\mathbf{v}) = D_{\mathbf{v}} f(\mathbf{x}_0).$$

Ist $f = (f_1, \ldots, f_k) : B \to \mathbb{R}^k$ in $\mathbf{x}_0$ differenzierbar, so ist

$$Df(\mathbf{x}_0) = (Df_1(\mathbf{x}_0), \ldots, Df_k(\mathbf{x}_0)).$$

BEWEIS: Im ersten Fall sei $L := Df(\mathbf{x}_0) \in \operatorname{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R})$. Wir benutzen die Darstellung $f(\mathbf{x}) = f(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0) + \varphi(\mathbf{x})$. Für $t \neq 0$ ist dann

$$\frac{1}{t}(f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)) = L(\mathbf{v}) + \frac{\varphi(\mathbf{x}_0 + t\mathbf{v})}{\|t\mathbf{v}\|} \cdot \|\mathbf{v}\|.$$

Die rechte Seite konvergiert für $t \to 0$ gegen $L(\mathbf{v})$.

Ist $f = (f_1, \ldots, f_k)$ in $\mathbf{x}_0$ differenzierbar, so gibt es Abbildungen

$$\delta_1, \ldots, \delta_k : B \to L(\mathbb{R}^n, \mathbb{R}),$$

die in $\mathbf{x}_0$ stetig sind, so daß gilt:

$$(f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})) = (f_1(\mathbf{x}_0), \ldots, f_k(\mathbf{x}_0)) + (\delta_1(\mathbf{x})(\mathbf{x} - \mathbf{x}_0), \ldots, \delta_1(\mathbf{x})(\mathbf{x} - \mathbf{x}_0)).$$

Daraus ergibt sich unmittelbar die Behauptung. ∎

1.5 Folgerung. *Ist $f = (f_1, \ldots, f_k) : B \to \mathbb{R}^k$ in $\mathbf{x}_0$ differenzierbar, so ist*

$$J_f(\mathbf{x}_0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \dfrac{\partial f_k}{\partial x_1}(\mathbf{x}_0) & \cdots & \dfrac{\partial f_k}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}.$$

BEWEIS: Die Matrix der linearen Abbildung $Df(\mathbf{x}_0)$ besitzt die Spalten

$$J_f(\mathbf{x}_0) \cdot \mathbf{e}_\nu^t = Df(\mathbf{x}_0)(\mathbf{e}_\nu)^t = D_{\mathbf{e}_\nu} f(\mathbf{x}_0)^t = ((f_1)_{x_\nu}(\mathbf{x}_0), \ldots, (f_k)_{x_\nu}(\mathbf{x}_0))^t.$$

∎

Wie man sieht, ist

$$J_f(\mathbf{x}_0) = \begin{pmatrix} \nabla f_1(\mathbf{x}_0) \\ \vdots \\ \nabla f_k(\mathbf{x}_0) \end{pmatrix}.$$

1.6 Satz. *Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}^k$ in der Nähe von $\mathbf{x}_0 \in B$ partiell differenzierbar. Sind alle partiellen Ableitungen von f in $\mathbf{x}_0$ stetig, so ist f in $\mathbf{x}_0$ (total) differenzierbar.*

BEWEIS: Wir wählen $\varepsilon > 0$ so, daß f auf $U_\varepsilon(\mathbf{x}_0) \subset B$ partiell differenzierbar ist. Zu jedem $\mathbf{x} \in U_\varepsilon(\mathbf{x}_0)$ konstruieren wir die Punktekette $\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n$ mit $\mathbf{z}_0 = \mathbf{x}_0$ und $\mathbf{z}_i = \mathbf{z}_{i-1} + (x_i - x_i^{(0)}) \cdot \mathbf{e}_i$, wie in Kapitel I, §2. Dann gibt es Punkte $\mathbf{c}_i = \mathbf{c}_i(\mathbf{x})$ auf der Verbindungsstrecke von $\mathbf{z}_{i-1}$ nach $\mathbf{z}_i$, so daß gilt:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}_i) \cdot (x_i - x_i^{(0)}).$$

Wir setzen dann

$$\Delta(\mathbf{x}) := (\frac{\partial f}{\partial x_1}(\mathbf{c}_1), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{c}_n)).$$

Offensichtlich ist $f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^t$, und wegen der Stetigkeit der partiellen Ableitungen konvergiert $\Delta(\mathbf{x})$ für $\mathbf{x} \to \mathbf{x}_0$ gegen $\nabla f(\mathbf{x}_0)$. Also ist f in $\mathbf{x}_0$ total differenzierbar. ∎

1.7 Satz. *Sind $f, g : B \to \mathbb{R}^k$ in $\mathbf{x}_0$ differenzierbar und $r, s \in \mathbb{R}$, so ist auch $r \cdot f + s \cdot g$ in $\mathbf{x}_0$ differenzierbar, und es gilt:*

$$D(r \cdot f + s \cdot g)(\mathbf{x}_0) = r \cdot Df(\mathbf{x}_0) + s \cdot Dg(\mathbf{x}_0).$$

Beweis: Trivial! ∎

Beispiele.

1. Sei $f(\mathbf{x}) \equiv \mathbf{c}$ konstant. Wir können $\delta = 0$ setzen und erhalten: f ist überall differenzierbar, und $Df(\mathbf{x}) = 0$ für alle $\mathbf{x}$.

2. Sei $f : \mathbb{R}^n \to \mathbb{R}^k$ linear. Dann ist $f(\mathbf{x}) - f(\mathbf{x}_0) = f(\mathbf{x} - \mathbf{x}_0)$, und wir können $\delta(\mathbf{x}) \equiv f$ setzen. So erhalten wir $Df(\mathbf{x}_0) = f$.

3. Sei $A \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix und $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch $f(\mathbf{x}) := \mathbf{x} \cdot A \cdot \mathbf{x}^t$. Dann ist $\mathbf{x} \cdot A \cdot \mathbf{y}^t = \mathbf{y} \cdot A \cdot \mathbf{x}^t$, und es gilt:

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}_0) &= \mathbf{x} \cdot A \cdot \mathbf{x}^t - \mathbf{x}_0 \cdot A \cdot \mathbf{x}_0^t \\ &= \mathbf{x}_0 \cdot A \cdot (\mathbf{x} - \mathbf{x}_0)^t + (\mathbf{x} - \mathbf{x}_0) \cdot A \cdot \mathbf{x}_0^t \\ &= \ + (\mathbf{x} - \mathbf{x}_0) \cdot A \cdot (\mathbf{x} - \mathbf{x}_0)^t \\ &= 2 \cdot (\mathbf{x} - \mathbf{x}_0) \cdot A \cdot \mathbf{x}_0^t + (\mathbf{x} - \mathbf{x}_0) \cdot A \cdot (\mathbf{x} - \mathbf{x}_0)^t \\ &= (\mathbf{x} - \mathbf{x}_0) \cdot (2 \cdot \mathbf{x}_0 \cdot A + (\mathbf{x} - \mathbf{x}_0) \cdot A)^t . \end{aligned}$$

Also setzen wir $\Delta(\mathbf{x}) := 2 \cdot \mathbf{x}_0 \cdot A + (\mathbf{x} - \mathbf{x}_0) \cdot A = (\mathbf{x} + \mathbf{x}_0) \cdot A$. Dann ist

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^t$$

und

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \Delta(\mathbf{x}) = 2 \cdot \mathbf{x}_0 \cdot A .$$

Also ist f in jedem Punkt $\mathbf{x}_0$ differenzierbar, und $Df(\mathbf{x}_0)(\mathbf{v}) = 2 \cdot \mathbf{v} \cdot A \cdot \mathbf{x}_0^t$.

In Koordinaten sieht die Situation folgendermaßen aus:

$$f(x_1, \ldots, x_n) = \sum_{i,j} a_{ij} x_i x_j \quad \text{und} \quad \frac{\partial f}{\partial x_\nu}(\mathbf{x}_0) = 2 \cdot \sum_{j=1}^{n} a_{\nu j} x_j^{(0)} .$$

1.8 Satz (Kettenregel). *Sei $B \subset \mathbb{R}^n$ offen, $f : B \to \mathbb{R}^k$ in $\mathbf{x}_0 \in B$ differenzierbar, $G \subset \mathbb{R}^k$ offen, $f(B) \subset G$ und $g : G \to \mathbb{R}^m$ in $f(\mathbf{x}_0)$ differenzierbar.*

Dann ist auch $g \circ f : B \to \mathbb{R}^m$ in $\mathbf{x}_0$ differenzierbar, und

$$D(g \circ f)(\mathbf{x}_0) = Dg(f(\mathbf{x}_0)) \circ Df(\mathbf{x}_0) .$$

Beweis: Sei $\mathbf{y}_0 := f(\mathbf{x}_0)$. Wir haben Darstellungen

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^t \\ \text{und} \quad g(\mathbf{y}) &= g(\mathbf{y}_0) + (\mathbf{y} - \mathbf{y}_0) \cdot \Delta^*(\mathbf{y}), \end{aligned}$$

wobei Δ in $\mathbf{x}_0$ und Δ^* in $\mathbf{y}_0$ stetig ist. Dann folgt:

$$\begin{aligned} g \circ f(\mathbf{x}) - g \circ f(\mathbf{x}_0) &= (f(\mathbf{x}) - f(\mathbf{x}_0)) \cdot \Delta^*(f(\mathbf{x}))^t \\ &= (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^t \cdot \Delta^*(f(\mathbf{x}))^t \\ &= (\mathbf{x} - \mathbf{x}_0) \cdot (\Delta^*(f(\mathbf{x})) \cdot \Delta(\mathbf{x}))^t \,. \end{aligned}$$

Also ist $g \circ f$ in $\mathbf{x}_0$ differenzierbar, und

$$\begin{aligned} D(g \circ f)(\mathbf{x}_0)(\mathbf{v}) &= \mathbf{v} \cdot (\Delta^*(f(\mathbf{x}_0)) \cdot \Delta(\mathbf{x}_0))^t \\ &= (\mathbf{v} \cdot \Delta(\mathbf{x}_0))^t \cdot \Delta^*(f(\mathbf{x}_0))^t \\ &= Dg(f(\mathbf{x}_0))(Dg(\mathbf{x}_0)(\mathbf{v})). \end{aligned}$$

∎

Bemerkung. Ist $\alpha : [a, b] \to \mathbb{R}^n$ ein differenzierbarer Weg, so ist

$$D\alpha(t_0)(v) = v \cdot \alpha'(t_0)\,.$$

Ist $f : B \to \mathbb{R}$ eine differenzierbare Funktion, so ist

$$Df(\mathbf{x}_0)(\mathbf{v}) = \mathbf{v} \cdot \nabla f(\mathbf{x}_0)^t\,.$$

Liegt nun die Spur von α in B, so ist $f \circ \alpha : [a, b] \to \mathbb{R}$ differenzierbar, und es gilt:

$$D(f \circ \alpha)(t_0)(v) = Df(\alpha(t_0))(D\alpha(t_0)(v)) = v \cdot \alpha'(t_0) \cdot \nabla f(\alpha(t_0))^t\,.$$

Daraus folgt: $(f \circ \alpha)'(t_0) = \nabla f(\alpha(t_0)) \bullet \alpha'(t_0)$. Ist f sogar stetig differenzierbar, so ist das genau die spezielle Kettenregel.

Beispiel.

Sei $\Phi : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^m$ eine bilineare Abbildung, also linear in beiden Argumenten. Sei $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^k \times \mathbb{R}^k$. Dann ist

$$\Phi(\mathbf{x}, \mathbf{y}) - \Phi(\mathbf{x}_0, \mathbf{y}_0) = \Phi(\mathbf{x} - \mathbf{x}_0, \mathbf{y}) + \Phi(\mathbf{x}_0, \mathbf{y} - \mathbf{y}_0).$$

Die Abbildung $\delta(\mathbf{x}, \mathbf{y}) : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^m$ mit

$$\delta(\mathbf{x}, \mathbf{y})(\mathbf{v}, \mathbf{w}) := \Phi(\mathbf{v}, \mathbf{y}) + \Phi(\mathbf{x}_0, \mathbf{w})$$

ist linear, und sie ist sicherlich stetig in $(\mathbf{x}_0, \mathbf{y}_0)$. Daher ist Φ in $(\mathbf{x}_0, \mathbf{y}_0)$ differenzierbar, und

$$D\Phi(\mathbf{x}_0, \mathbf{y}_0)(\mathbf{v}, \mathbf{w}) = \Phi(\mathbf{v}, \mathbf{y}_0) + \Phi(\mathbf{x}_0, \mathbf{w}) .$$

Ist nun $B \subset \mathbb{R}^n$ offen und sind $f, g : B \to \mathbb{R}^k$ zwei differenzierbare Abbildungen, so ist auch $\Phi \circ (f, g) : B \to \mathbb{R}$ differenzierbar, und es gilt:

$$\begin{aligned} D(\Phi \circ (f, g))(\mathbf{x}_0)(\mathbf{v}) &= D\Phi(f(\mathbf{x}_0), g(\mathbf{x}_0)) \circ D(f, g)(\mathbf{x}_0)(\mathbf{v}) \\ &= \Phi(Df(\mathbf{x}_0)(\mathbf{v}), g(\mathbf{x}_0)) + \Phi(f(\mathbf{x}_0), Dg(\mathbf{x}_0)(\mathbf{v})). \end{aligned}$$

Das ist eine Verallgemeinerung der Produktregel. Haben f und g Werte in $\mathbb{R}$, so ist

$$D(f \cdot g)(\mathbf{x}_0) = f(\mathbf{x}_0) \cdot Dg(\mathbf{x}_0) + g(\mathbf{x}_0) \cdot Df(\mathbf{x}_0) .$$

Ist E ein endlich-dimensionaler reeller Vektorraum, so kann E (etwa mit Hilfe einer Basis) mit einer Norm versehen werden. Damit wird E zu einem metrischen Raum, und die so gewonnene Topologie auf E ist unabhängig von der Norm. Ist F ein weiterer endlich-dimensionaler Vektorraum, so ist jede lineare Abbildung von $f : E \to F$ stetig. Der Beweis funktioniert genauso wie bei linearen Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^k$. Da die Kugel $B = \{x \in E : \|x\| \le 1\}$ kompakt ist, nimmt f auf B sein Maximum an.

Definition. Sei $f : E \to F$ eine lineare Abbildung zwischen endlich-dimensionalen Vektorräumen. Dann nennt man

$$\|f\| := \sup_{\|\mathbf{x}\| \le 1} \|f(\mathbf{x})\|$$

die *Norm* von f.

Tatsächlich erfüllt $\|f\|$ alle Eigenschaften einer Norm und macht $L(E, F)$ zu einem normierten (endlich-dimensionalen) Vektorraum.

Bemerkung. Ist $f : E \to F$ linear und $\mathbf{x} \in E$, so ist $\|f(\mathbf{x})\| \le \|f\| \cdot \|\mathbf{x}\|$. Im Falle $\mathbf{x} = \mathbf{0}$ ist das klar, und im Falle $\mathbf{x} \ne \mathbf{0}$ ist

$$\frac{\|f(\mathbf{x})\|}{\|\mathbf{x}\|} = \|f(\frac{\mathbf{x}}{\|\mathbf{x}\|})\| \le \|f\| .$$

1.9 Satz. *Ist $f : \mathbb{R}^n \to \mathbb{R}$ die durch $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{a}^t$ gegebene Linearform, so ist $\|f\| = \|\mathbf{a}\|$.*

BEWEIS: Für beliebiges $\mathbf{x} \in \mathbb{R}^n$ ist

$$|f(\mathbf{x})| = |\mathbf{a} \bullet \mathbf{x}| \le \|\mathbf{a}\| \cdot \|\mathbf{x}\| .$$

Daraus folgt: $\|f\| \le \|\mathbf{a}\|$. Aber für $\mathbf{x} := \mathbf{a}/\|\mathbf{a}\|$ gilt: $\|\mathbf{x}\| = 1$ und $f(\mathbf{x}) = \|\mathbf{a}\|^2/\|\mathbf{a}\| = \|\mathbf{a}\|$. Daher ist $\|f\| = \|\mathbf{a}\|$. ∎

Ist $A \in M_{k,n}(\mathbb{R})$ eine Matrix und L_A die zugehörige lineare Abbildung (mit $L_A(\mathbf{x}) := \mathbf{x} \cdot A^t$), so setzt man $\|A\| := \|L_A\|$.

1.10 Satz.

1. *Es ist* $\|A \cdot B\| \le \|A\| \cdot \|B\|$.
2. *Ist* $A = (a_{ij})$, *so ist* $\|A\| \le (\sum_{i,j} a_{ij}^2)^{1/2}$.

Beweis: 1) Es ist

$$\|L_{A\cdot B}(\mathbf{x})\| = \|L_A(L_B(\mathbf{x}))\| \le \|A\| \cdot \|L_B(\mathbf{x})\| \le \|A\| \cdot \|B\| \cdot \|\mathbf{x}\|.$$

2) Bezeichnen wir mit $z_i(A) = (a_{i1}, \ldots, a_{in})$ die i-te Zeile von A, so ist $z_i(A)^t$ die i-te Spalte von A^t, also $z_i(A) = \mathbf{e}_i \cdot A$. Dann gilt:

$$\begin{aligned} \|\mathbf{x} \cdot A^t\|^2 &= \|\sum_{i=1}^n (\mathbf{x} \cdot z_i(A)^t)\mathbf{e}_i\|^2 \\ &= \sum_{i=1}^n (\mathbf{x} \cdot z_i(A)^t)^2 \\ &\le \sum_{i=1}^n \|\mathbf{x}\|^2 \cdot \|z_i(A)\|^2 \\ &= \|\mathbf{x}\|^2 \cdot \sum_{i=1}^n \sum_{j=1}^n (a_{ij})^2. \end{aligned}$$

■

1.11 Mittelwertsatz. *Sei* $B \subset \mathbb{R}^n$ *offen und* $f : B \to \mathbb{R}$ *differenzierbar. Wenn mit den Punkten* $\mathbf{a}, \mathbf{b} \in B$ *auch deren Verbindungsstrecke zu* B *gehört, so gibt es einen Punkt* $\mathbf{c}$ *auf dieser Verbindungsstrecke mit*

$$f(\mathbf{b}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{b} - \mathbf{a}).$$

Beweis: Sei $\varphi(t) := f((1-t)\mathbf{a} + t\mathbf{b})$ für $0 \le t \le 1$. Dann ist $\varphi(0) = f(\mathbf{a})$, $\varphi(1) = f(\mathbf{b})$ und

$$\varphi'(t) = Df((1-t)\mathbf{a} + t\mathbf{b})(\mathbf{b} - \mathbf{a}).$$

Nach dem Mittelwertsatz in einer Veränderlichen gibt es ein $\xi \in (0,1)$ mit $\varphi(1) - \varphi(0) = \varphi'(\xi)$. Mit $\mathbf{c} := (1-\xi)\mathbf{a} + \xi\mathbf{b}$ folgt: $f(\mathbf{b}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{b} - \mathbf{a})$. ■

1.12 Folgerung (Schrankensatz). *Sei* $B \subset \mathbb{R}^n$ *offen und konvex,* $f : B \to \mathbb{R}$ *differenzierbar und* $\|Df(\mathbf{x})\| \le C$ *für alle* $\mathbf{x} \in B$. *Dann ist*

$$|f(\mathbf{b}) - f(\mathbf{a})| \le C \cdot \|\mathbf{b} - \mathbf{a}\| \text{ für alle } \mathbf{a}, \mathbf{b} \in B.$$

BEWEIS: Es interessiert nur der Fall, daß $\mathbf{a} \neq \mathbf{b}$ ist. Wegen der Konvexität von B liegt die Verbindungsstrecke von $\mathbf{a}$ und $\mathbf{b}$ ganz in B. Dann gibt es ein $\mathbf{c}$ auf der Verbindungsstrecke mit $f(\mathbf{b}) - f(\mathbf{a}) = Df(\mathbf{c})(\mathbf{b} - \mathbf{a})$, und es ist

$$|f(\mathbf{b}) - f(\mathbf{a})| = |Df(\mathbf{c})(\mathbf{b} - \mathbf{a})| \leq \|Df(\mathbf{c})\| \cdot \|\mathbf{b} - \mathbf{a}\| \leq C \cdot \|\mathbf{b} - \mathbf{a}\|.$$

∎

1.13 Satz. *Sei $B \subset \mathbb{R}^n$ offen und zusammenhängend, $f : B \to \mathbb{R}$ differenzierbar. Dann gilt:*

$$Df(\mathbf{x}) \equiv 0 \iff f \text{ konstant auf } B.$$

BEWEIS: Die eine Richtung („$\Leftarrow$“) ist trivial.

Für die andere Richtung wählen wir einen festen Punkt $\mathbf{a} \in B$. Ist $\mathbf{x} \neq \mathbf{a}$ ein beliebiger Punkt von B, so kann man $\mathbf{x}$ innerhalb von B durch einen Streckenzug mit $\mathbf{a}$ verbinden. Es gibt also Punkte $\mathbf{z}_0 = \mathbf{a}, \mathbf{z}_1, \ldots, \mathbf{z}_N = \mathbf{x}$ in B, so daß die Verbindungsstrecke von $\mathbf{z}_{i-1}$ und $\mathbf{z}_i$ ganz in B liegt. Auf dieser Verbindungsstrecke gibt es ein $\mathbf{c}_i$ mit

$$f(\mathbf{z}_i) - f(\mathbf{z}_{i-1}) = Df(\mathbf{c}_i)(\mathbf{z}_i - \mathbf{z}_{i-1}) = 0.$$

Dann ist auch $f(\mathbf{x}) - f(\mathbf{a}) = 0$. Das bedeutet, daß f konstant ist. ∎

Ist E ein normierter $\mathbb{R}$-Vektorraum (mit Norm N) und (f_n) eine Folge von Abbildungen von einer offenen Menge $B \subset \mathbb{R}^n$ nach E, so kann man wie üblich die Begriffe *(gleichmäßige) Konvergenz* und *Cauchyfolge* erklären.

1. (f_n) konvergiert gegen $f : B \to E$, falls gilt:

   $\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N}$, so daß für $n \geq n_0$ gilt: $N(f_n(\mathbf{x}) - f(\mathbf{x})) < \varepsilon$ für alle $\mathbf{x} \in B$.

2. (f_n) ist eine Cauchyfolge, falls gilt:

   $\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N}$, so daß für $n, m \geq n_0$ gilt: $N(f_n(\mathbf{x}) - f_m(\mathbf{x})) < \varepsilon$ für $\mathbf{x} \in B$.

1.14 Theorem. *Sei $B \subset \mathbb{R}^n$ offen und (f_n) eine Folge von differenzierbaren Abbildungen $f_n : B \to \mathbb{R}^k$, die punktweise gegen eine Abbildung $f : B \to \mathbb{R}^k$ konvergiert. Wenn die Folge der Ableitungen (Df_n) auf B gleichmäßig gegen eine Abbildung $g : B \to L(\mathbb{R}^n, \mathbb{R}^k)$ konvergiert, dann ist f differenzierbar und $Df = g$.*

BEWEIS: Wir gehen so ähnlich wie bei dem Beweis des entsprechenden Satzes in einer Veränderlichen vor. Sei $\mathbf{x}_0$ ein fester Punkt von B und $U = U_r(\mathbf{x}_0)$ eine Kugelumgebung, deren Abschluß noch in B liegt. Setzen wir $f_{nm} := f_n - f_m$, so gilt für $\mathbf{x} \in U$:

$$\|f_{nm}(\mathbf{x}) - f_{nm}(\mathbf{x}_0)\| \le \|\mathbf{x} - \mathbf{x}_0\| \cdot \sup_U \|Df_{nm}\|. \qquad (*)$$

Nun sei $\varepsilon > 0$. Wir wählen ein $n_0 \in \mathbb{N}$, so daß

$$\sup_U \|Df_{nm}\| < \varepsilon \text{ und } \sup_U \|Df_n - g\| < \varepsilon$$

für $n, m \ge n_0$ ist. Wegen der gleichmäßigen Konvergenz von (f_n) gegen g auf B ist das möglich.

Lassen wir in $(*)$ m gegen Unendlich gehen, so erhalten wir:

$$\|(f_n(\mathbf{x}) - f(\mathbf{x})) - (f_n(\mathbf{x}_0) - f(\mathbf{x}_0))\| \le \|\mathbf{x} - \mathbf{x}_0\| \cdot \varepsilon \text{ für } n \ge n_0.$$

Jetzt halten wir ein solches n fest. Da f_n differenzierbar ist, gibt es ein $\delta > 0$, so daß für $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ gilt:

$$\|f_n(\mathbf{x}) - f_n(\mathbf{x}_0) - Df_n(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\| \le \|\mathbf{x} - \mathbf{x}_0\| \cdot \varepsilon.$$

Dann folgt:

$$\begin{aligned}
\|f(\mathbf{x}) - f(\mathbf{x}_0) - g(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\| &\le \|(f_n(\mathbf{x}) - f(\mathbf{x})) - (f_n(\mathbf{x}_0) - f(\mathbf{x}_0))\| \\
&\quad + \|f_n(\mathbf{x}) - f_n(\mathbf{x}_0) - Df_n(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\| \\
&\quad + \|Df_n(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) - g(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\| \\
&\le 3\varepsilon \cdot \|\mathbf{x} - \mathbf{x}_0\|.
\end{aligned}$$

Das bedeutet, daß f in $\mathbf{x}_0$ differenzierbar und $Df(\mathbf{x}_0) = g(\mathbf{x}_0)$ ist. ∎

Bemerkung. Zum Schluß dieses Paragraphen kommen wir zurück zu dem Problem, die Tangentialebene an den Graphen einer differenzierbaren Funktion zu finden.

Ist $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ in $\mathbf{x}_0$ differenzierbar, so soll $T_{(\mathbf{x}_0, f(\mathbf{x}_0))}(G_f)$ diejenige affine Hyperebene im $\mathbb{R}^{n+1}$ sein, die den Graphen G_f über $\mathbf{x}_0$ berührt. Nach unseren Überlegungen zu Anfang des Paragraphen ist das der Graph der affin-linearen Abbildung $T_f : \mathbb{R}^n \to \mathbb{R}$, die gegeben ist durch

$$T_f(\mathbf{x}) := f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0).$$

Dann ist

$$\begin{aligned}
T_{(\mathbf{x}_0, f(\mathbf{x}_0))}(G_f) &= \{(\mathbf{x}, T_f(\mathbf{x})) \,:\, \mathbf{x} \in \mathbb{R}^n\} \\
&= \{(\mathbf{x}_0 + \mathbf{h}, f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{h})) \,:\, \mathbf{h} \in \mathbb{R}^n\} \\
&= \{(\mathbf{x}, z) \in \mathbb{R}^{n+1} \,:\, z - f(\mathbf{x}_0) = Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\}.
\end{aligned}$$

§ 2 Höhere Ableitungen und Taylorformel

Definition. $E_1, E_2, \ldots, E_q$ und F seien reelle Vektorräume. Eine Abbildung

$$\varphi : E_1 \times \ldots \times E_q \to F$$

heißt *(q-fach) multilinear*, wenn sie in jedem Argument linear ist, d.h. wenn gilt:

1. $\varphi(v_1, \ldots, v_i + v_i', \ldots, v_q) = \varphi(v_1, \ldots, v_i, \ldots, v_q) + \varphi(v_1, \ldots, v_i', \ldots, v_q)$.
2. $\varphi(v_1, \ldots, c \cdot v_i, \ldots, v_q) = c \cdot \varphi(v_1, \ldots, v_i, \ldots, v_q)$ für $c \in \mathbb{R}$.

Den Vektorraum aller q-fach multilinearen Abbildungen von $E_1 \times \ldots \times E_q$ nach F bezeichnen wir mit $L_q(E_1, \ldots, E_q; F)$. Ist $E_1 = \ldots = E_q =: E$, so schreiben wir auch kurz $L_q(E; F)$. Ist $F = \mathbb{R}$, so sprechen wir von *Multilinearformen*.

Wir betrachten hier nur endlich-dimensionale (normierte) Vektorräume.

Beispiele.

1. Die „Evaluationsabbildung“ $\mathrm{ev} : L(E, F) \times E \to F$ mit $\mathrm{ev}(L, v) := L(v)$ ist bilinear.
2. Die Elemente von $L_2(E; \mathbb{R})$ nennt man auch *Bilinearformen* auf E. Ist auf E sogar ein Skalarprodukt $< \ldots, \ldots >$ gegeben, so definiert jede lineare Abbildung $\lambda : E \to E$ eine Bilinearform φ auf E durch

   $$\varphi(x, y) :=< \lambda(x), y > \ .$$

   Auf dem $\mathbb{R}^n$ kann jede Bilinearform mit Hilfe einer Matrix beschrieben werden:

   $$\varphi(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot B \cdot \mathbf{w}^t \, .$$

   Die Einträge b_{ij} in der Matrix B sind dann gegeben durch $b_{ij} = \mathbf{e}_i \cdot B \cdot \mathbf{e}_j^t$.

   Nun kommt eine begrifflich etwas schwierige, aber wichtige Betrachtung: Ist $\mathbf{v} \in \mathbb{R}^n$ fest, so wird durch $\mathbf{w} \mapsto \mathbf{v} \cdot B \cdot \mathbf{w}^t$ eine Linearform $\lambda_{\mathbf{v}}$ auf dem $\mathbb{R}^n$ definiert. Die Zuordnung $\lambda : \mathbf{v} \mapsto \lambda_{\mathbf{v}}$ ist eine lineare Abbildung

   $$\lambda : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}),$$

   also ein Element aus $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}))$.

   Ist umgekehrt ein $\lambda \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}))$ gegeben, so erhält man eine Bilinearform $\varphi_\lambda : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, indem man definiert:

   $$\varphi_\lambda(\mathbf{v}, \mathbf{w}) := \lambda(v)(w).$$

   Auch wenn E und F beliebige Vektorräume sind, wird durch diese Formel ein Isomorphismus $L(E, L(E, F)) \to L_2(E; F)$ definiert.

3. Was mit bilinearen Abbildungen funktioniert, das geht z.B. auch mit allgemeinen Multilinearformen. Ist E ein endlich-dimensionaler Vektorraum, so gibt es für alle $q \in \mathbb{N}$ einen Isomorphismus

$$L(E, L(E, \dots, L(E, \mathbb{R}) \dots)) \to L_q(E; \mathbb{R}),$$

mit $\lambda \mapsto \varphi_\lambda$ und $\varphi_\lambda(v_1, \dots, v_q) := \lambda(v_1)(v_2) \dots (v_q)$.

Bemerkung. Jede multilineare Abbildung (zwischen endlich-dimensionalen Vektorräumen) ist stetig.

BEWEIS: Wir betrachten nur den Fall einer bilinearen Abbildung $\varphi : E_1 \times E_2 \to F$. Zunächst wählen wir Basen $\{a_1, \dots, a_n\}$ von E_1 und $\{b_1, \dots, b_m\}$ von E_2. Dann gilt für $v = \sum_i v_i a_i$ und $w = \sum_j w_j b_j$

$$\|\varphi(v, w)\| = \|\sum_{i,j} v_i w_j \varphi(a_i, b_j)\| \leq C \cdot \|(v_1, \dots, v_n)\| \cdot \|(w_1, \dots, w_m)\| \cdot \sum_{i,j} \|\varphi(a_i, b_j)\|,$$

mit einer geeigneten Konstanten $C > 0$. Daraus folgt die Stetigkeit im Nullpunkt.

Wegen $\varphi(v, w) - \varphi(v_0, w_0) = \varphi(v - v_0, w) + \varphi(v_0, w - w_0)$ ergibt sich daraus die Stetigkeit in allen anderen Punkten. ■

Definition. E und F seien endlich-dimensionale Vektorräume, und $B \subset E$ sei eine offene Teilmenge. Eine Abbildung $f : B \to F$ heißt in einem Punkt $x_0 \in B$ *differenzierbar*, wenn es eine Abbildung $\delta : B \to L(E, F)$ gibt, so daß gilt:

1. $f(x) = f(x_0) + \delta(x)(x - x_0)$ für $x \in B$.

2. δ ist stetig in x_0.

Die lineare Abbildung $Df(x_0) := \delta(x_0) \in L(E, F)$ heißt dann die *Ableitung* von f in x_0.

Wie im $\mathbb{R}^n$ folgt aus der Differenzierbarkeit die Stetigkeit, es gilt die Linearität, die Produktregel und die Kettenregel. Eine Funktionalmatrix können wir allerdings nur in Abhängigkeit von Basen definieren.

Sei nun $B \subset E$ offen, $f : B \to F$ eine Abbildung und $\{z_1, \dots, z_m\}$ eine Basis von F, so daß man f in der Form

$$f(x) = \sum_{\nu=1}^{m} f_\nu(x) \cdot z_\nu$$

schreiben kann, mit skalaren Funktionen $f_\nu : B \to \mathbb{R}$. Die Abbildung f ist genau dann in x_0 differenzierbar, wenn alle f_ν in x_0 differenzierbar sind, und es gilt: $Df(x_0) = \sum_\nu Df_\nu(x_0) \cdot z_\nu$.

Wir versuchen jetzt herauszufinden, was unter der zweiten Ableitung einer Funktion zu verstehen ist.

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ überall differenzierbar. Dann ist $Df : B \to L(\mathbb{R}^n, \mathbb{R})$ definiert. Wir können Df in der Form $Df = \sum_\nu g_\nu \varepsilon^\nu$ schreiben, mit gewissen Funktionen $g_\nu : B \to \mathbb{R}$. Nun gilt:

$$D_j f(\mathbf{x}) = Df(\mathbf{x})(\mathbf{e}_j) = \big(\sum_{\nu=1}^n g_\nu(\mathbf{x})\varepsilon^\nu\big)(\mathbf{e}_j) = \sum_{\nu=1}^n g_\nu(\mathbf{x})\varepsilon^\nu(\mathbf{e}_j) = g_j(\mathbf{x}).$$

Daher können wir schreiben:

$$Df = \sum_{j=1}^n (D_j f) \cdot \varepsilon^j \,.$$ [2]

Nun nehmen wir an, daß Df in $x_0 \in B$ ein weiteres Mal differenzierbar ist. Dann ist

$$D(Df)(\mathbf{x}_0) = \sum_{j=1}^n D(D_j f)(\mathbf{x}_0) \cdot \varepsilon^j \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R})).$$

Daraus folgt

$$D(Df)(\mathbf{x}_0)(\mathbf{v}) = \sum_{j=1}^n D(D_j f)(\mathbf{x}_0)(\mathbf{v}) \cdot \varepsilon^j \,,$$

mit

$$D(D_j f)(\mathbf{x}_0)(\mathbf{v}) = \big(\sum_i D_i(D_j f)(\mathbf{x}_0) \cdot \varepsilon^i\big)(\mathbf{v}) = \sum_i D_i(D_j f)(\mathbf{x}_0) \cdot v_i \,.$$

So erhalten wir schließlich

$$D^2 f(\mathbf{x}_0)(\mathbf{v}, \mathbf{w}) := D(Df)(\mathbf{x}_0)(\mathbf{v})(\mathbf{w}) = \sum_{j=1}^n \sum_{i=1}^n D_i(D_j f)(\mathbf{x}_0) v_i w_j \,.$$

Die Bilinearform $D^2 f(\mathbf{x}_0) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ nennt man die *zweite Ableitung* von f in $\mathbf{x}_0$. Voraussetzung für ihre Existenz ist die Differenzierbarkeit von f in der Nähe von $\mathbf{x}_0$ und die Differenzierbarkeit von Df im Punkt $\mathbf{x}_0$.

Existiert $D^2 f$ in einer ganzen Umgebung U von $\mathbf{x}_0$ und ist $D^2 f : U \to L_2(\mathbb{R}^n; \mathbb{R})$ wiederum in $\mathbf{x}_0$ differenzierbar, so nennt man $D^3 f(\mathbf{x}_0) \in L_3(\mathbb{R}^n; \mathbb{R})$ mit

$$D^3 f(\mathbf{x}_0)(\mathbf{u}, \mathbf{v}, \mathbf{w}) := D(D^2 f)(\mathbf{x}_0)(\mathbf{u})(\mathbf{v}, \mathbf{w})$$

die dritte Ableitung von f in $\mathbf{x}_0$, und so geht es weiter. Wir werden hier aber ab der Ordnung 3 nur noch mit partiellen Ableitungen arbeiten.

[2] Im Grunde ist das nur eine andere Schreibweise für die Gleichung $df = \sum_{j=1}^n f_{x_j}\, dx_j$.

Definition. Sei f in $\mathbf{x}_0 \in \mathbb{R}^n$ zweimal differenzierbar. Dann heißt die Matrix

$$H_f(\mathbf{x}_0) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) \;\middle|\; i, j = 1, \ldots, n \right)$$

die *Hesse–Matrix* von f in $\mathbf{x}_0$.

2.1 Satz. *Ist f in $\mathbf{x}_0 \in \mathbb{R}^n$ zweimal differenzierbar, so ist die Hesse–Matrix $H_f(\mathbf{x}_0)$ symmetrisch.*

BEWEIS: Sind die zweiten partiellen Ableitungen von f in $\mathbf{x}_0$ sogar stetig, so folgt die Symmetrie der Hesse–Matrix aus dem Satz von Schwarz. Wird die Stetigkeit der 2. Ableitungen nicht vorausgesetzt, so ist ein etwas subtilerer Beweis notwendig, auf den wir hier jedoch verzichten wollen. ∎

Bemerkung. Im Falle $n = 2$ ist

$$H_f(x, y) = \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{pmatrix}.$$

Die zweite Ableitung läßt sich nun auch folgendermaßen beschreiben:

$$D^2 f(\mathbf{x}_0)(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot H_f(\mathbf{x}_0) \cdot \mathbf{w}^{\,t}.$$

2.2 Folgerung. *Die zweite Ableitung ist eine symmetrische Bilinearform:*

$$D^2 f(\mathbf{x}_0)(\mathbf{v}, \mathbf{w}) = D^2 f(\mathbf{x}_0)(\mathbf{w}, \mathbf{v}).$$

Als nächstes wollen wir eine Taylorformel für Funktionen von mehreren Veränderlichen herleiten. Wie versprochen, wollen wir dabei aber nur mit partiellen Ableitungen arbeiten.

Sei f in der Nähe von $\mathbf{x}_0 \in \mathbb{R}^n$ genügend oft differenzierbar. Wir betrachten den Weg $\alpha(t) := \mathbf{x}_0 + t\mathbf{h}$, mit $\mathbf{h} := \mathbf{x} - \mathbf{x}_0$, und untersuchen die Funktion

$$g(t) := f \circ \alpha(t) = f(\mathbf{x}_0 + t\mathbf{h}).$$

Auf jeden Fall ist

$$g'(t) = Df(\mathbf{x}_0 + t\mathbf{h})(\mathbf{h}) = \mathbf{h} \bullet \nabla f(\mathbf{x}_0 + t\mathbf{h}).$$

Wir wollen die höheren Ableitungen von g berechnen.

Sei P der Differentialoperator

$$P = \mathbf{h} \bullet \nabla = h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\,.$$

Dann ist $(Pf) \circ \alpha = (f \circ \alpha)'$, und per Induktion folgt:

$$(P^k f) \circ \alpha = (f \circ \alpha)^{(k)}\,.$$

Der Induktionsschritt sieht dabei folgendermaßen aus:

$$\begin{aligned} (P^{k+1} f) \circ \alpha &= P(P^k f) \circ \alpha = ((P^k f) \circ \alpha)' \\ &= ((f \circ \alpha)^{(k)})' = (f \circ \alpha)^{(k+1)}. \end{aligned}$$

Um $g^{(k)}(t) = (\mathbf{h} \bullet \nabla)^k f(\mathbf{x}_0 + t\mathbf{h})$ zu berechnen, brauchen wir einen Satz über Polynome.

2.3 Satz.

$$(x_1 + \cdots + x_n)^k = \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{k!}{\alpha_1! \cdots \alpha_n!} x_1^{\alpha_1} \cdots x_n^{\alpha_n}\,.$$

Beweis: (Induktion nach n)

Der Induktionsanfang ist trivial. Zum Induktionsschluß:

$$\begin{aligned} (x_1 + \cdots + x_{n+1})^k &= ((x_1 + \cdots + x_n) + x_{n+1})^k \\ &= \sum_{m + \alpha_{n+1} = k} \frac{k!}{m!\alpha_{n+1}!} (x_1 + \cdots + x_n)^m x_{n+1}^{\alpha_{n+1}} \\ &= \sum_{m + \alpha_{n+1} = k} \frac{k!}{m!\alpha_{n+1}!} \sum_{\alpha_1 + \cdots + \alpha_n = m} \frac{m!}{\alpha_1! \cdots \alpha_n!} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \cdot x_{n+1}^{\alpha_{n+1}} \\ &= \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{k!}{\alpha_1! \cdots \alpha_n!} x_1^{\alpha_1} \cdots x_n^{\alpha_n}. \end{aligned}$$

■

Da die h_i Konstanten und die partiellen Ableitungen vertauschbar sind, kann man $(\mathbf{h} \bullet \nabla)^k$ genauso wie den Ausdruck $(x_1 + \cdots + x_n)^k$ berechnen. Es folgt:

$$g^{(k)}(t) = (\mathbf{h} \bullet \nabla)^k f(\mathbf{x}_0 + t\mathbf{h}) = k! \sum_{|\alpha| = k} \frac{1}{\alpha!} D^\alpha f(\mathbf{x}_0 + t\mathbf{h}) \cdot \mathbf{h}^\alpha\,.$$

Dabei ist $\alpha! := \alpha_1! \cdots \alpha_n!$, $|\alpha| := \alpha_1 + \cdots + \alpha_n$ und $D^\alpha f := D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f$, sowie $\mathbf{h}^\alpha := h_1^{\alpha_1} \cdots h_n^{\alpha_n}$ für einen Vektor $\mathbf{h} = (h_1, \ldots, h_n)$.

Ist f k-mal differenzierbar, so nennt man

$$T_k f(\mathbf{x}; \mathbf{x}_0) := \sum_{|\alpha| \le k} \frac{1}{\alpha!} D^\alpha f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^\alpha$$

das k-te *Taylorpolynom* von f in $\mathbf{x}_0$.

2.4 Satz (Taylorentwicklung). *Sei $B \subset \mathbb{R}^n$ eine offene Menge, die sternförmig bezüglich $\mathbf{x}_0 \in B$ ist, und $f : B \to \mathbb{R}$ eine k-mal stetig differenzierbare Funktion. Dann gibt es eine Darstellung $f = T_k f + R_k$, wobei gilt:*

1. $\displaystyle\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{R_k(\mathbf{x})}{\|\mathbf{x}-\mathbf{x}_0\|^k} = 0.$
2. *Ist f sogar $(k+1)$-mal differenzierbar, so gibt es zu jedem $\mathbf{x} \in B$ ein $\xi \in [0,1]$, so daß gilt:*

$$R_k(\mathbf{x}) = \sum_{|\alpha|=k+1} \frac{1}{\alpha!} D^\alpha(\mathbf{x}_0 + \xi(\mathbf{x}-\mathbf{x}_0))(\mathbf{x}-\mathbf{x}_0)^\alpha.$$

Beweis: Wir betrachten zunächst den Fall, daß f sogar $(k+1)$-mal differenzierbar ist. Sei $\alpha(t) := \mathbf{x}_0 + t(\mathbf{x}-\mathbf{x}_0)$. Dann ist auch $g(t) := f \circ \alpha(t)$ $(k+1)$-mal differenzierbar. Die Taylorformel in einer Veränderlichen liefert zu jedem t ein $\xi = \xi(t)$ zwischen 0 und t, so daß gilt:

$$g(t) = \sum_{i=0}^{k} \frac{g^{(i)}(0)}{i!} t^i + \frac{1}{(k+1)!} g^{(k+1)}(\xi) t^{k+1}\,.$$

Setzen wir $t = 1$, so erhalten wir

$$f(\mathbf{x}) = \sum_{|\alpha|\le k} \frac{1}{\alpha!} D^\alpha f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)^\alpha + \sum_{|\alpha|=k+1} \frac{1}{\alpha!} D^\alpha f(\mathbf{x}_0 + \xi(\mathbf{x}-\mathbf{x}_0))(\mathbf{x}-\mathbf{x}_0)^\alpha.$$

Ist f nur k-mal stetig differenzierbar, so setzen wir $\mathbf{h} := \mathbf{x} - \mathbf{x}_0$ und erhalten

$$\begin{aligned} f(\mathbf{x}) &= T_{k-1} f(\mathbf{x};\mathbf{x}_0) + \sum_{|\alpha|=k} \frac{1}{\alpha!} D^\alpha f(\mathbf{x}_0 + \xi\mathbf{h})\mathbf{h}^\alpha \\ &= T_k f(\mathbf{x};\mathbf{x}_0) + \sum_{|\alpha|=k} \frac{1}{\alpha!} \big(D^\alpha f(\mathbf{x}_0 + \xi\mathbf{h}) - D^\alpha f(\mathbf{x}_0)\big)\mathbf{h}^\alpha. \end{aligned}$$

Setzen wir $\varphi_\alpha(\mathbf{h}) := \dfrac{1}{\alpha!}\big(D^\alpha f(\mathbf{x}_0 + \xi\mathbf{h}) - D^\alpha f(\mathbf{x}_0)\big)$, so erhalten wir

$$f(\mathbf{x}) = T_k f(\mathbf{x};\mathbf{x}_0) + \sum_{|\alpha|=k} \varphi_\alpha(\mathbf{h})\mathbf{h}^\alpha.$$

Für $|\alpha| = k$ ist

$$\frac{|\mathbf{h}^\alpha|}{\|\mathbf{h}\|^k} = \frac{|h_1|^{\alpha_1}\cdots|h_n|^{\alpha_n}}{\|\mathbf{h}\|^{\alpha_1}\cdots\|\mathbf{h}\|^{\alpha_n}} \le 1\,.$$

Daraus folgt:

$$|\sum_{|\alpha|=k} \varphi_\alpha(\mathbf{h})\mathbf{h}^\alpha|/\|\mathbf{h}\|^k \le \sum_{|\alpha|=k} |\varphi_\alpha(\mathbf{h})| \to 0 \text{ für } \mathbf{h} \to \mathbf{0},$$

wegen der Stetigkeit von $D^\alpha f$ in $\mathbf{x}_0$. ■

Wie in der Theorie von einer Veränderlichen sei $\mathcal{C}^k(B)$ die Menge aller k-mal stetig differenzierbaren Funktionen auf B. Dabei ist auch $k = \infty$ zugelassen.

Definition. Sei $M \subset \mathbb{R}^n$ eine Teilmenge, $f : M \to \mathbb{R}$ stetig, $\mathbf{a} \in M$ ein Punkt.

f hat in $\mathbf{a}$ auf M ein *relatives (oder lokales) Maximum* (bzw. ein *relatives (oder lokales) Minimum*), wenn es eine offene Umgebung $U(\mathbf{a}) \subset \mathbb{R}^n$ gibt, so daß

$$f(\mathbf{x}) \le f(\mathbf{a}) \quad (\text{bzw.} \quad f(\mathbf{x}) \ge f(\mathbf{a})\ )$$

für alle $\mathbf{x} \in U \cap M$ ist. In beiden Fällen spricht man von einem *relativen (oder lokalen) Extremum.*

Gilt die Ungleichung sogar für alle $\mathbf{x} \in M$, so spricht man von einem *absoluten (oder globalen)* Maximum oder Minimum.

2.5 Satz. *Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ in $\mathbf{a} \in B$ differenzierbar.*

Besitzt f in $\mathbf{a}$ ein relatives Extremum, so ist $\nabla f(\mathbf{a}) = \mathbf{0}$.

Beweis: Für $i = 1, \dots, n$ besitzt auch $g_i(t) := f(\mathbf{a} + t\mathbf{e}_i)$ in $t = 0$ ein lokales Extremum. Nach dem notwendigen Kriterium aus der Differentialrechnung einer Veränderlichen muß dann $(g_i)'(0) = 0$ sein. Es ist aber

$$(g_i)'(0) = \frac{\partial f}{\partial x_i}(\mathbf{a}), \text{ für } i = 1, \dots, n.$$

Daraus folgt die Behauptung. ■

Definition. Ist f in $\mathbf{a}$ differenzierbar und $\nabla f(\mathbf{a}) = \mathbf{0}$, so heißt $\mathbf{a}$ ein *stationärer (oder kritischer) Punkt* von f.

Ein stationärer Punkt $\mathbf{a}$ von f heißt *Sattelpunkt* von f, falls es in jeder Umgebung $U(\mathbf{a})$ Punkte $\mathbf{b}$ und $\mathbf{c}$ gibt, so daß $f(\mathbf{b}) < f(\mathbf{a}) < f(\mathbf{c})$ ist.

Wir formulieren die Taylorsche Formel in einem speziellen Fall:

2.6 Satz (Taylorformel 2.Ordnung). *Sei $\mathbf{a} \in \mathbb{R}^n$, $B = B_r(\mathbf{a})$ eine offene Kugel um $\mathbf{a}$, $f : B \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gibt es eine auf $B_r(\mathbf{0})$ definierte Funktion R, so daß gilt:*

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h}^t + \frac{1}{2}\mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^t + R(\mathbf{h})$$

und

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|^2} = 0\,.$$

Beweis: Für $|\alpha| = 2$ gibt es die Möglichkeiten

$$\begin{aligned}
(\alpha_1,\ldots,\alpha_n) &= (2,0,0,\ldots,0),\\
(\alpha_1,\ldots,\alpha_n) &= (1,1,0,\ldots,0),\\
&\vdots\\
(\alpha_1,\ldots,\alpha_n) &= (1,0,\ldots,0,1),\\
(\alpha_1,\ldots,\alpha_n) &= (0,2,0,\ldots,0),\\
&\vdots\\
(\alpha_1,\ldots,\alpha_n) &= (0,0,\ldots,0,2).
\end{aligned}$$

Daher ist

$$\sum_{|\alpha|=2} \frac{1}{\alpha!} D^\alpha f(\mathbf{a}) \cdot \mathbf{h}^\alpha = \frac{1}{2}\sum_{i=1}^n f_{x_i x_i}(\mathbf{a}) h_i^2 + \sum_{i<j} f_{x_i x_j}(\mathbf{a}) h_i h_j = \frac{1}{2}\sum_{i,j} f_{x_i x_j}(\mathbf{a}) h_i h_j\,.$$

■

Ist nun f in $\mathbf{a}$ stationär, also

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^t + R(\mathbf{h}),$$

so hängt das Verhalten von f in der Nähe von $\mathbf{a}$ im Wesentlichen von der Hesse-Matrix ab, denn $R(\mathbf{h})$ verschwindet ja in $\mathbf{a}$ von höherer Ordnung. Das führt uns zu einem ähnlichen hinreichenden Kriterium für Extremwerte, wie wir es aus der eindimensionalen Theorie kennen. Allerdings ist die Lage hier doch etwas komplizierter.

Ist $A \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix, so nennt man die Funktion

$$q(\mathbf{h}) := \mathbf{h} \cdot A \cdot \mathbf{h}^t$$

eine *quadratische Form*. Es ist $q(t\mathbf{h}) = t^2 \cdot q(\mathbf{h})$ für $t \in \mathbb{R}$ und $\mathbf{h} \in \mathbb{R}^n$. Insbesondere ist natürlich $q(\mathbf{0}) = 0$.

Definition. Eine quadratische Form $q(\mathbf{h})$ heißt

$$\begin{aligned}
\text{positiv semidefinit} &:\iff && q(\mathbf{h}) \ge 0 \text{ für alle } \mathbf{h},\\
\text{positiv definit} &:\iff && q(\mathbf{h}) > 0 \text{ für alle } \mathbf{h} \ne \mathbf{0},\\
\text{negativ semidefinit} &:\iff && q(\mathbf{h}) \le 0 \text{ für alle } \mathbf{h},\\
\text{negativ definit} &:\iff && q(\mathbf{h}) < 0 \text{ für alle } \mathbf{h} \ne \mathbf{0},\\
\text{indefinit} &:\iff && \exists\, \mathbf{h}_1, \mathbf{h}_2 \text{ mit } q(\mathbf{h}_1) < 0 < q(\mathbf{h}_2).
\end{aligned}$$

In der Linearen Algebra wird gezeigt:

Ist $A \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix, so gibt es eine orthogonale Matrix $S \in \mathrm{GL}(n, \mathbb{R})$, so daß $S^{-1} \cdot A \cdot S$ eine Diagonalmatrix ist. Die Einträge in der Diagonalmatrix sind die Eigenwerte von A. (Satz von der Hauptachsentransformation).

Daß S orthogonal ist, bedeutet, daß $S^t S = E_n$, also $S^{-1} = S^t$ ist. Setzen wir $\mathbf{v} := \mathbf{h} \cdot S$, so ist

$$\begin{aligned} q_A(\mathbf{h}) := \mathbf{h} \cdot A \cdot \mathbf{h}^t &= (\mathbf{v} \cdot S^t) \cdot A \cdot (\mathbf{v} \cdot S^t)^t \\ &= \mathbf{v} \cdot (S^{-1} \cdot A \cdot S) \cdot \mathbf{v}^t \\ &= \mathbf{v} \cdot \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \cdot \mathbf{v}^t \\ &= \sum_{i=1}^{n} \lambda_i (v_i)^2, \end{aligned}$$

wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A sind. Also folgt:

$$\begin{aligned} q_A \text{ positiv definit} &\iff \mathbf{h} \cdot A \cdot \mathbf{h}^t > 0 \text{ für alle } \mathbf{h} \neq \mathbf{0} \\ &\iff \sum_{i=1}^{n} \lambda_i (v_i)^2 > 0 \text{ für alle } \mathbf{v} \neq \mathbf{0} \\ &\iff \lambda_1, \ldots, \lambda_n > 0. \end{aligned}$$

Genauso ist q_A negativ definit, wenn alle Eigenwerte von A negativ sind. Und q_A ist positiv semidefinit (bzw. negativ semidefinit), wenn alle Eigenwerte von $A \geq 0$ (bzw. ≤ 0) sind. Gibt es wenigstens einen negativen und einen positiven Eigenwert, so ist q_A indefinit.

Im Falle $n = 2$ gibt es noch ein einfacheres Kriterium:

2.7 Satz.

Sei $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in M_{2,2}(\mathbb{R})$ eine symmetrische Matrix. Dann ist

$$q_A(h_1, h_2) = a h_1^2 + 2b h_1 h_2 + d h_2^2,$$

und es gilt:

1. *Ist $\det(A) < 0$, so ist q_A indefinit.*
2. *Ist $\det(A) > 0$ und $a > 0$, so ist q_A positiv definit.*
3. *Ist $\det(A) > 0$ und $a < 0$, so ist q_A negativ definit.*

Beweis: Sei $\Delta := \det(A) = ad - b^2$. Zur Berechnung der Eigenwerte brauchen wir noch das charakteristische Polynom:

$$p_A(x) = \det \begin{pmatrix} a-x & b \\ b & d-x \end{pmatrix} = (a-x)(d-x) - b^2 = x^2 - (a+d)x + \Delta.$$

Die Eigenwerte λ_1, λ_2 von A sind die beiden Nullstellen dieses quadratischen Polynoms. Nach dem Satz von Vieta ist

$$\lambda_1 + \lambda_2 = a + d \quad \text{und} \quad \lambda_1 \cdot \lambda_2 = \Delta.$$

Ist $\Delta < 0$, so haben die beiden Eigenwerte verschiedenes Vorzeichen, und q_A ist indefinit. Ist $\Delta > 0$, so sind λ_1 und λ_2 beide $\neq 0$, und sie haben das gleiche Vorzeichen. Außerdem ist $ad = \Delta + b^2 > 0$. Ist nun $a > 0$, so ist auch $d > 0$ und damit $\lambda_1 + \lambda_2 > 0$. In diesem Fall ist q_A positiv definit. Genauso folgt aus $a < 0$, daß q_A negativ definit ist. ∎

Nun ergibt sich:

2.8 Satz (Hinreichendes Kriterium für Extremwerte).
Sei $B \subset \mathbb{R}^n$ offen, $f \in \mathcal{C}^2(B)$. Weiter sei $\mathbf{a} \in B$ ein stationärer Punkt von f, also $\nabla f(\mathbf{a}) = \mathbf{0}$.

1. *Ist $H_f(\mathbf{a})$ positiv definit, so besitzt f in $\mathbf{a}$ ein relatives Minimum.*
2. *Ist $H_f(\mathbf{a})$ negativ definit, so besitzt f in $\mathbf{a}$ ein relatives Maximum.*
3. *Ist $H_f(\mathbf{a})$ indefinit, so liegt in $\mathbf{a}$ ein Sattelpunkt vor.*

Beweis:
1) Sei $q(\mathbf{h}) := \mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^t$. Da f in $\mathbf{a}$ stationär ist, ergibt die Taylorformel:

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \frac{1}{2} q(\mathbf{h}) + R(\mathbf{h}).$$

Die Funktion q ist stetig und nach Voraussetzung > 0 außerhalb des Nullpunktes. Insbesondere nimmt sie auf der abgeschlossenen und beschränkten und daher kompakten Menge

$$S^{n-1} := \{\mathbf{x} \in \mathbb{R}^n \,:\, \|\mathbf{x}\| = 1\}$$

ein Minimum $m > 0$ an. Daher gilt für beliebiges $\mathbf{h} \in \mathbb{R}^n$:

$$q(\mathbf{h}) = \|\mathbf{h}\|^2 \cdot q\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right) \geq m \cdot \|\mathbf{h}\|^2.$$

Ist jetzt ein ε mit $0 < \varepsilon < \frac{m}{2}$ vorgegeben und dazu ein $r = r(\varepsilon)$ so gewählt, daß

$$|R(\mathbf{h})| \leq \varepsilon \cdot \|\mathbf{h}\|^2 \text{ für } \mathbf{h} \in B_r(\mathbf{0})$$

ist, so folgt für alle $\mathbf{h} \in B_r(\mathbf{0})$:

$$
\begin{aligned}
f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) &= \frac{1}{2}q(\mathbf{h}) + R(\mathbf{h}) \\
&\geq (\frac{m}{2} - \varepsilon) \cdot \|\mathbf{h}\|^2 \geq 0
\end{aligned}
$$

Also ist $f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a})$ für kleines $\mathbf{h}$, und es liegt ein relatives Minimum in $\mathbf{a}$ vor.

2) Der Fall des Maximums kann durch Übergang von f zu $-f$ auf (1) zurückgeführt werden.

3) Ist q indefinit, so gibt es in jeder Umgebung von $\mathbf{0}$ Vektoren $\mathbf{h}_1$ und $\mathbf{h}_2$ mit $q(\mathbf{h}_1) < 0 < q(\mathbf{h}_2)$. Die Funktionen

$$
\begin{aligned}
f_1(t) &:= f(\mathbf{a}+t\mathbf{h}_1) \\
\text{und} \quad f_2(t) &:= f(\mathbf{a}+t\mathbf{h}_2)
\end{aligned}
$$

sind dann nahe $t = 0$ definiert und zweimal differenzierbar, und es gilt:

$$(f_1)'(0) = (f_2)'(0) = 0, \quad (f_1)''(0) = q(\mathbf{h}_1) < 0 \quad \text{und} \quad (f_2)''(0) = q(\mathbf{h}_2) > 0.$$

Also besitzt f_1 in $t = 0$ ein isoliertes Maximum und f_2 in $t = 0$ ein isoliertes Minimum. Das bedeutet, daß f beliebig nahe bei $\mathbf{a}$ sowohl Werte $< f(\mathbf{a})$ als auch Werte $> f(\mathbf{a})$ annimmt. Damit liegt ein Sattelpunkt vor. ∎

Bemerkung. Ist $H_f(\mathbf{a})$ nur semidefinit, so kann man keine genaue Aussage machen!

Beispiele.

1. Sei $f(x,y) := x^2 + y^2$. Dann ist $\nabla f(x,y) = (2x, 2y)$, also $(0,0)$ der einzige stationäre Punkt von f. Da $f(0,0) = 0$ und allgemein $f(x,y) \geq 0$ ist, liegt ein absolutes Minimum vor. Tatsächlich ist

   $$H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

   Da $\det H_f(x,y) = 4 > 0$ ist, ist die Matrix positiv definit. Das hinreichende Kriterium sagt also auch, daß f im Nullpunkt ein lokales Minimum besitzt.

2. Sei $f(x,y) := 1 - x^2 - y^2$. Dann ist $\nabla f(x,y) = (-2x, -2y)$ und $H_f(x,y) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$ negativ definit. Hier liegt im Nullpunkt ein Maximum vor.

3. Sei $f(x,y) := x^2 - y^2$. Nun ist $\nabla f(x,y) = (2x, -2y)$ und $H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$. Da $\det H_f(x,y) < 0$ ist, hat f in $\mathbf{0}$ einen Sattelpunkt.

4. Sei $f(x,y) := e^{xy} + x^2 + \frac{1}{9}y^2$.

   Dann ist $\nabla f(x,y) = (ye^{xy} + 2x, xe^{xy} + \frac{2}{9}y)$. Für die Hesse-Matrix ergibt sich:

$$H_f(x,y) = \begin{pmatrix} 2+y^2e^{xy} & (1+xy)e^{xy} \\ (1+xy)e^{xy} & 2/9+x^2e^{xy} \end{pmatrix}.$$

Der Nullpunkt ist sicher ein stationärer Punkt. Ist (x,y) irgendein anderer stationärer Punkt, so muß gelten:

$$xye^{xy} = -2x^2 \quad \text{und} \quad xye^{xy} = -\frac{2}{9}y^2,$$

also $x = \pm\frac{1}{3}y$.

Wäre $x = \frac{1}{3}y$, so wäre $0 = f_y(x,y) = \frac{y}{3}(e^{xy}+\frac{2}{3})$, also $y = 0$ (und damit auch $x = 0$) oder $e^{xy} = -\frac{2}{3}$, was nicht möglich ist. So bleibt nur die Gleichung $x = -\frac{1}{3}y$. Wegen der Bedingung $0 = f_x(x,y) = y(e^{xy}-\frac{2}{3})$ muß dann $e^{xy} = \frac{2}{3}$ sein, also $e^{-y^2/3} = \frac{2}{3}$.

Das führt auf die Gleichung $y^2 = -3\ln(\frac{2}{3})$. Setzen wir $p := \sqrt{-3\ln(\frac{2}{3})}$ (der Radikand ist positiv, weil $\ln(\frac{2}{3}) < 0$ ist!), so sind die Punkte

$$\mathbf{x}_{1,2} := \pm(-\frac{1}{3}p, p)$$

weitere Kandidaten für stationäre Punkte, und mehr kann es nicht geben. Nun gilt:

$$\begin{aligned} f(\mathbf{0}) &= 1 \\ \text{und} \quad f(\mathbf{x}_{1,2}) &= e^{-p^2/3} + \frac{2}{9}p^2 \\ &= \frac{2}{3}\cdot(1-\ln(\frac{2}{3})). \end{aligned}$$

Daß $\nabla f(\mathbf{0}) = (0,0)$ ist, ist klar. Ist (x,y) einer der beiden Punkte $\mathbf{x}_1$ oder $\mathbf{x}_2$, so ist $xy = -\frac{p^2}{3} = \ln(\frac{2}{3})$, also

$$\nabla f(x,y) = (\frac{2}{3}y + 2x, \frac{2}{3}x + \frac{2}{9}y) = \pm(\frac{2}{3}p - \frac{2}{3}p, -\frac{2}{9}p + \frac{2}{9}p) = (0,0).$$

Da $H_f(0,0) = \begin{pmatrix} 2 & 1 \\ 1 & 2/9 \end{pmatrix}$ ist, also $\det H_f(0,0) = \frac{4}{9} - 1 < 0$, liegt im Nullpunkt ein Sattelpunkt vor! Und da $f(x,y) > \frac{1}{9}(x^2+y^2)$ ist, gilt für $\|(x,y)\| \ge 3$, daß $f(x,y) > 1$ ist. Auf der kompakten Menge $\overline{B_3(\mathbf{0})}$ muß f als stetige Funktion ein globales Minimum ≤ 1 annehmen, und das muß sogar in der offenen Kugel $B_3(\mathbf{0})$ liegen, weil f auf dem Rand der Kugel schon Werte > 1 annimmt. In einem solchen Minimum muß f aber einen stationären Punkt besitzen. Dafür kommen nur die beiden Punkten $\mathbf{x}_1$ und $\mathbf{x}_2$ in Frage, und weil f in diesen Punkten den gleichen Wert annimmt, können wir schließen:

f besitzt in $\mathbf{x}_1$ und in $\mathbf{x}_2$ jeweils ein (globales) Minimum.

§ 3 Der Umkehrsatz

Wir werden uns zunächst mit invertierbaren Matrizen beschäftigen.

3.1 Satz. *Die Menge $G := \mathrm{GL}_n(\mathbb{R})$ aller invertierbaren Matrizen ist offen in $M_{n,n}(\mathbb{R})$, und die Abbildung $i : G \to G$ mit $i(A) := A^{-1}$ ist stetig.*

Beweis: Versehen wir $M_{n,n}(\mathbb{R}) \cong \mathbb{R}^{n^2}$ mit der euklidischen Norm, so erhalten wir die selbe Topologie wie mit der Matrizen-Norm. Die Determinante ist als Polynomfunktion stetig, und $G = \{A \in M_{n,n}(\mathbb{R}) \,:\, \det(A) \neq 0\}$ ist daher offen.

Aus der Linearen Algebra weiß man: Ist $A^{-1} = (y_{ij})$, so ist

$$y_{ij} = (-1)^{i+j} \det S_{ji}(A) \cdot (\det A)^{-1},$$

wobei $S_{ji}(A)$ die *Streichungsmatrix* ist, die aus A entsteht, indem man die j-te Zeile und die i-te Spalte streicht. Offensichtlich sind die y_{ij} stetige Funktionen. ∎

3.2 Satz. *Die Abbildung $i : A \mapsto A^{-1}$ ist überall in $G = \mathrm{GL}_n(\mathbb{R})$ differenzierbar, und es gilt: $Di(A_0)(B) := -A_0^{-1} \cdot B \cdot A_0^{-1}$.*

Beweis: Setzen wir $M := M_{n,n}(\mathbb{R})$, so können wir $\delta : G \to L(M, M)$ definieren durch

$$\delta(A)(B) := -A^{-1} \cdot B \cdot A_0^{-1}\,.$$

Dann ist

$$\begin{aligned} i(A) - i(A_0) &= A^{-1} - A_0^{-1} = A^{-1} \cdot (A_0 - A) \cdot A_0^{-1} \\ &= \delta(A)(A - A_0). \end{aligned}$$

Da δ in A_0 stetig ist, ist i in A_0 differenzierbar, und $Di(A_0)(B) = \delta(A_0)(B) = -A_0^{-1} \cdot B \cdot A_0^{-1}$. ∎

3.3 Banachscher Fixpunktsatz. *Sei X ein vollständiger metrischer Raum und $f : X \to X$ eine stetige Abbildung. Wenn es ein λ mit $0 < \lambda < 1$ gibt, so daß*

$$d(f(x), f(y)) \leq \lambda \cdot d(x, y)$$

für alle $x, y \in X$ ist, so besitzt f einen „Fixpunkt", d.h. es gibt ein $x^* \in X$ mit $f(x^*) = x^*$. Dieser Fixpunkt ist eindeutig bestimmt.

Bemerkung. Man nennt die Abbildung f *kontrahierend*. Daß $\lambda < 1$ ist, ist entscheidend.

Beweis: Wir geben mit dem Beweis zugleich ein Konstruktionsverfahren an:

Sei $x_0 \in X$ beliebig gewählt. Die Punktfolge (x_n) sei induktiv durch

$$x_{n+1} := f(x_n)$$

definiert. Wir wollen zeigen, daß (x_n) gegen einen Fixpunkt konvergiert. Dazu schätzen wir ab:

$$\begin{aligned} d(x_n, x_{n+1}) &= d(f(x_{n-1}), f(x_n)) \\ &\leq \lambda \cdot d(x_{n-1}, x_n) \\ &\leq \cdots \\ &\leq \lambda^n \cdot d(x_0, x_1). \end{aligned}$$

Da $0 < \lambda < 1$ ist, strebt der letzte Ausdruck gegen Null. Also kommen sich die Folgeglieder immer näher, die (x_n) bilden eine Cauchyfolge. Da X vollständig ist, konvergiert (x_n) gegen einen Punkt $x^* \in X$.

Weiter gilt:

$$\begin{aligned} d(f(x^*), x^*) &\leq d(f(x^*), f(x_n)) + d(f(x_n), x^*) \\ &\leq \lambda \cdot d(x^*, x_n) + d(x_{n+1}, x^*), \end{aligned}$$

und dieser Ausdruck wird beliebig klein. Das ist nur möglich, wenn $f(x^*) = x^*$ ist.

Zur Eindeutigkeit: Seien x und y zwei Fixpunkte. Ist $x \neq y$, so ist $d(x, y) > 0$ und daher

$$d(x, y) = d(f(x), f(y)) \leq \lambda \cdot d(x, y) < d(x, y).$$

Das kann aber nicht sein. ∎

Definition. Sei $B \subset \mathbb{R}^n$ offen und $F : B \to \mathbb{R}^n$ stetig differenzierbar. F heißt in $\mathbf{x}_0 \in B$ *regulär*, falls $\det J_F(\mathbf{x}_0) \neq 0$ ist.

Bemerkung. Daß $\det J_F(\mathbf{x}_0) \neq 0$ ist, hat zur Folge, daß $J_F(\mathbf{x}_0)$ eine invertierbare Matrix und $DF(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^n$ ein Isomorphismus ist. Also kann man die Umkehrabbildung $(DF(\mathbf{x}_0))^{-1}$ bilden.

3.4 Satz. *Ist F in $\mathbf{x}_0$ regulär, so gilt:*

1. *Es gibt eine offene Umgebung $U = U(\mathbf{x}_0) \subset B$, so daß $F|_U$ injektiv ist.*
2. *$\mathbf{y}_0 = F(\mathbf{x}_0)$ ist innerer Punkt von $F(B)$.*
3. *Es gibt eine Konstante $C > 0$, so daß $\|\mathbf{x}_1 - \mathbf{x}_2\| \leq C \cdot \|F(\mathbf{x}_1) - F(\mathbf{x}_2)\|$ für $\mathbf{x}_1, \mathbf{x}_2 \in U$ ist.*

BEWEIS: a) Zunächst vereinfachen wir die Situation ein wenig.

Ist $\mathbf{a} \in \mathbb{R}^n$ ein fester Punkt, so ist die Translation $T_{\mathbf{a}}(\mathbf{x}) := \mathbf{x} + \mathbf{a}$ ein Homöomorphismus des $\mathbb{R}^n$ auf sich. Da wir F durch $T_{-F(\mathbf{x}_0)} \circ F \circ T_{\mathbf{x}_0}$ ersetzen können, dürfen wir annehmen, daß $\mathbf{x}_0 = \mathbf{0}$ und $F(\mathbf{0}) = \mathbf{0}$ ist.

Die lineare Abbildung $L = DF(\mathbf{0})$ ist nach Voraussetzung ebenfalls ein Homöomorphismus. Da wir F durch $L^{-1} \circ F$ ersetzen können, dürfen wir annehmen, daß $DF(\mathbf{0}) = \mathrm{id}$ ist.

b) Lokale Injektivität:

Sei $F = (F_1, \ldots, F_n)$ und

$$\widetilde{F}(\mathbf{u}_1, \ldots, \mathbf{u}_n) := \begin{pmatrix} \nabla F_1(\mathbf{u}_1) \\ \vdots \\ \nabla F_n(\mathbf{u}_n) \end{pmatrix} \in M_{n,n}(\mathbb{R}).$$

Dann ist $\det \circ \widetilde{F} : B \times \ldots \times B \to \mathbb{R}$ eine stetige Funktion und $\det(\widetilde{F}(\mathbf{u}, \ldots, \mathbf{u})) = \det J_F(\mathbf{u})$, und man kann eine konvexe Umgebung $U = U(\mathbf{0}) \subset B$ finden, so daß

$$\det \widetilde{F}(\mathbf{u}_1, \ldots, \mathbf{u}_n) \neq 0 \text{ für } \mathbf{u}_1, \ldots, \mathbf{u}_n \in U$$

ist. Insbesondere ist dann $\det J_F(\mathbf{u}) \neq 0$ für $\mathbf{u} \in U$.

Sind nun $\mathbf{a}, \mathbf{b} \in U$, so gibt es nach dem Mittelwertsatz Punkte $\mathbf{z}_\nu$ zwischen $\mathbf{a}$ und $\mathbf{b}$, so daß gilt:

$$F_\nu(\mathbf{b}) - F_\nu(\mathbf{a}) = (\mathbf{b} - \mathbf{a}) \cdot \nabla F_\nu(\mathbf{z}_\nu)^{\,t}, \text{ für } \nu = 1, \ldots, n,$$

also

$$F(\mathbf{b}) - F(\mathbf{a}) = (\mathbf{b} - \mathbf{a}) \cdot \widetilde{F}(\mathbf{z}_1, \ldots, \mathbf{z}_n)^{\,t}.$$

Weil die Matrix $\widetilde{F}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ invertierbar ist, folgt: Ist $\mathbf{a} \neq \mathbf{b}$, so muß auch $F(\mathbf{a}) \neq F(\mathbf{b})$ sein. Das bedeutet, daß F auf U injektiv ist.

c) Lokale Surjektivität:

Sei $g(\mathbf{x}) := \mathbf{x} - F(\mathbf{x})$ und $\Phi_{\mathbf{y}}(\mathbf{x}) := g(\mathbf{x}) + \mathbf{y}$. Offensichtlich ist

$$\Phi_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} \quad \Longleftrightarrow \quad (\mathbf{x} - F(\mathbf{x})) + \mathbf{y} = \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{y} = F(\mathbf{x}).$$

Wir müssen also für $\mathbf{y} \in \mathbb{R}^n$ nach Fixpunkten von $\Phi_{\mathbf{y}}$ suchen. Um den Banachschen Fixpunktsatz anwenden zu können, müssen wir zeigen, daß $\Phi_{\mathbf{y}}$ kontrahierend ist. Wegen $\Phi_{\mathbf{y}}(\mathbf{a}) - \Phi_{\mathbf{y}}(\mathbf{b}) = g(\mathbf{a}) - g(\mathbf{b})$ reicht es zu zeigen, daß g kontrahierend ist.

Da $Dg(\mathbf{0}) = 0$ ist, gibt es ein $r > 0$, so daß $\|Dg(\mathbf{x})\| < \frac{1}{2}$ für $\mathbf{x} \in K := \overline{B_r(\mathbf{0})} \subset U$ ist. Nach dem Schrankensatz ist dann

$$\|g(\mathbf{a}) - g(\mathbf{b})\| \leq \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|, \text{ für } \mathbf{a}, \mathbf{b} \in K.$$

Weil $g(\mathbf{0}) = \mathbf{0}$ ist, bedeutet das auch, daß $\|g(\mathbf{x})\| \le \frac{1}{2}\|\mathbf{x}\|$ für $\mathbf{x} \in K$ ist. Daraus folgt: Ist $\|\mathbf{y}\| \le r/2$, so ist $\|\Phi_{\mathbf{y}}(\mathbf{x})\| = \|\mathbf{y} + g(\mathbf{x})\| \le r$ für $\mathbf{x} \in K$. Das bedeutet, daß $\Phi_{\mathbf{y}}$ für jedes $\mathbf{y}$ mit $\|\mathbf{y}\| \le r/2$ die Menge K in sich überführt. K ist als kompakte Menge ein vollständiger metrischer Raum, und da $\Phi_{\mathbf{y}} : K \to K$ kontrahierend ist, besitzt jedes $\mathbf{y}$ mit $\|\mathbf{y}\| \le r/2$ ein Urbild in K unter F, es ist $B_{r/2}(\mathbf{0}) \subset F(B)$.

d) Ersetzen wir U durch die kleinere Umgebung $B_r(\mathbf{0})$, so bleibt F dort injektiv, und für $\mathbf{x}_1, \mathbf{x}_2 \in U$ gilt:

$$\begin{aligned} \|\mathbf{x}_1 - \mathbf{x}_2\| &= \|g(\mathbf{x}_1) + F(\mathbf{x}_1) - (g(\mathbf{x}_2) + F(\mathbf{x}_2))\| \\ &\le \|g(\mathbf{x}_1) - g(\mathbf{x}_2)\| + \|F(\mathbf{x}_1) - F(\mathbf{x}_2)\| \\ &\le \frac{1}{2}\|\mathbf{x}_1 - \mathbf{x}_2\| + \|F(\mathbf{x}_1) - F(\mathbf{x}_2)\|, \end{aligned}$$

also

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \le 2 \cdot \|F(\mathbf{x}_1) - F(\mathbf{x}_2)\|.$$

■

3.5 Satz von der Umkehrabbildung. *Sei $B \subset \mathbb{R}^n$ offen und $F : B \to \mathbb{R}^n$ stetig differenzierbar. Ist $\mathbf{x}_0 \in B$, $F(\mathbf{x}_0) = \mathbf{y}_0$ und F in $\mathbf{x}_0$ regulär, so gibt es offene Umgebungen $U(\mathbf{x}_0) \subset B$ und $V(\mathbf{y}_0) \subset \mathbb{R}^n$, so daß gilt:*

1. $\det J_F(\mathbf{x}) \ne 0$ *für alle* $\mathbf{x} \in U$.
2. $F : U \to V$ *ist bijektiv.*
3. $F^{-1} : V \to U$ *ist wieder differenzierbar.*
4. *Für* $\mathbf{x} \in U$ *und* $\mathbf{y} = F(\mathbf{x})$ *ist* $D(F^{-1})(\mathbf{y}) = (DF(\mathbf{x}))^{-1}$.

BEWEIS: Nach Voraussetzung gibt es eine offene Umgebung $U_0 = U_0(\mathbf{x}_0)$, so daß $F|_{U_0}$ injektiv und $\mathbf{y}_0 := F(\mathbf{x}_0)$ innerer Punkt von $F(U_0)$ ist. Sei $V = V(\mathbf{y}_0) \subset F(U_0)$ eine offene Umgebung und $U := (F|_{U_0})^{-1}(V)$. Dann ist auch U eine offene Umgebung von $\mathbf{x}_0$, und $F : U \to V$ bijektiv. (Es ist klar, daß $F : U \to F(U)$ bijektiv und $F(U) \subset V$ ist. Umgekehrt gibt es zu jedem $\mathbf{y} \in V$ ein $\mathbf{x} \in U_0$ mit $F(\mathbf{x}) = \mathbf{y}$. Aber dann muß $\mathbf{x}$ schon in U und daher $\mathbf{y}$ in $F(U)$ liegen).

Es gibt eine Konstante $C > 0$, so daß für $\mathbf{x}_1, \mathbf{x}_2 \in U$ gilt:

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \le C \cdot \|F(\mathbf{x}_1) - F(\mathbf{x}_2)\|.$$

Aber dann ist

$$\|F^{-1}(\mathbf{y}_1) - F^{-1}(\mathbf{y}_2)\| \le C \cdot \|\mathbf{y}_1 - \mathbf{y}_2\| \text{ für } \mathbf{y}_1, \mathbf{y}_2 \in V,$$

also $F^{-1} : V \to U$ stetig.

Da F in $\mathbf{x}_0$ differenzierbar ist, gibt es eine Darstellung

$$F(\mathbf{x}) = F(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \Delta(\mathbf{x})^t,$$

mit einer in $\mathbf{x}_0$ stetigen Abbildung $\Delta : B \to M_{n,n}(\mathbb{R})$. Dann ist $d(\mathbf{x}) := \det \Delta(\mathbf{x})$ in $\mathbf{x}_0$ stetig und $d(\mathbf{x}_0) \neq 0$. Es gibt daher eine offene Umgebung U von $\mathbf{x}_0$, so daß $d(\mathbf{x}) \neq 0$ für $\mathbf{x} \in U$ ist. Das bedeutet, daß $\Delta(\mathbf{x})$ dort invertierbar ist. Da $i(A) := A^{-1}$ stetig ist, ist auch $\Delta^*(\mathbf{y}) := i(\Delta(F^{-1}(\mathbf{y}))) = \Delta(F^{-1}(\mathbf{y}))^{-1}$ stetig in $\mathbf{y}_0 = F(\mathbf{x}_0)$. Aus der Gleichung $(F(\mathbf{x}) - F(\mathbf{x}_0)) \cdot (\Delta(\mathbf{x})^t)^{-1} = \mathbf{x} - \mathbf{x}_0$ folgt nun:

$$F^{-1}(\mathbf{y}) = F^{-1}(\mathbf{y}_0) + (\mathbf{y} - \mathbf{y}_0) \cdot \Delta^*(\mathbf{y})^t.$$

Das liefert die Differenzierbarkeit von F^{-1} und die Ableitung
$DF^{-1}(\mathbf{y}_0) = DF(F^{-1}(\mathbf{y}_0))^{-1}$. ∎

Definition. $U, V \subset \mathbb{R}^n$ seien offene Teilmengen. Eine Abbildung $F : U \to V$ heißt *Diffeomorphismus*, falls gilt:

1. F ist differenzierbar und bijektiv.
2. $F^{-1} : V \to U$ ist differenzierbar.

Der Satz von der Umkehrabbildung besagt: Ist F in $\mathbf{x}_0$ regulär, so gibt es Umgebungen $U = U(\mathbf{x}_0)$ und $V = V(F(\mathbf{x}_0))$, so daß $F|_U : U \to V$ ein Diffeomorphismus ist.

Gibt es zu jedem $\mathbf{x}$ solche Umgebungen, so nennt man F einen *lokalen Diffeomorphismus.*

3.6 Satz. *Sei $F : U \to V$ ein differenzierbarer Homöomorphismus. Ist F überall regulär, so ist F sogar ein Diffeomorphismus.*

BEWEIS: Trivial! ∎

Beispiele.

1. Durch $(x, y) = F(r, \varphi) = (r\cos(\varphi), r\sin(\varphi))$ für $(r, \varphi) \in \mathbb{R}_+ \times \mathbb{R}$ sind die ebenen Polarkoordinaten gegeben. Es ist

   $$J_F(r, \varphi) = \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix},$$

   also $\det J_F(r, \varphi) = r\cos^2(\varphi) + r\sin^2(\varphi) = r$. In jedem Punkt (r, φ) mit $r > 0$ und $\varphi \in \mathbb{R}$ ist F damit lokal umkehrbar.

   $$F : \mathbb{R}_+ \times [0, 2\pi) \to \mathbb{R}^2 \setminus \{(0, 0)\}$$

   ist sogar global umkehrbar, denn F ist injektiv:

Ist $F(r_1, \varphi_1) = F(r_2, \varphi_2)$, so ist auch $r_1 = \|F(r_1, \varphi_1)\| = \|F(r_2, \varphi_2)\| = r_2$. Da – wie im ersten Semester gezeigt – die Abbildung $\varphi \mapsto e^{\mathrm{i}\varphi}$ bijektiv ist (von $[0, 2\pi) \to S^1$), muß auch $\varphi_1 = \varphi_2$ sein.

$F^{-1} : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}_+ \times [0, 2\pi)$ ist gegeben durch

$$F^{-1}(x, y) := (\sqrt{x^2 + y^2}, \arg(x, y)),$$

wobei $\arg(x, y)$ der Winkel zwischen der positiven x-Achse und der Geraden durch $(0,0)$ und (x, y) ist. Ist z.B. $x, y > 0$, so ist $\arg(x, y) = \arctan(y/x)$.

2. Sei $F : \mathbb{R}^2 \to \mathbb{R}^2$ definiert durch

$$F(x, y) := (x^2 - y^2, 2xy).$$

Dann gilt:

$$J_F(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}, \text{ also } \det J_F(x, y) = 4(x^2 + y^2).$$

Damit ist $\det J_F(x, y) \neq 0$ für $(x, y) \neq (0, 0)$ und F überall außerhalb des Nullpunktes lokal umkehrbar.

F ist aber nicht global umkehrbar, denn es ist ja $F(-x, -y) = F(x, y)$

3. Räumliche Polarkoordinaten:

$$F(r, \varphi, \theta) := (r \cos\varphi \cos\theta, r \sin\varphi \cos\theta, r \sin\theta),$$

mit $r > 0$, $0 < \varphi < 2\pi$ und $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Dabei ist r der Abstand vom Nullpunkt, φ der Winkel gegen die positive x-Achse und θ der Winkel gegen die x-y-Ebene.

Dann ist

$$\begin{aligned}
\det J_F(r, \varphi, \theta) &= \det \begin{pmatrix} \cos\varphi \cos\theta & -r \sin\varphi \cos\theta & -r \cos\varphi \sin\theta \\ \sin\varphi \cos\theta & r \cos\varphi \cos\theta & -r \sin\varphi \sin\theta \\ \sin\theta & 0 & r \cos\theta \end{pmatrix} \\
&= \sin\theta \cdot r^2(\sin^2\varphi \sin\theta \cos\theta + \cos^2\varphi \sin\theta \cos\theta) \\
&\quad + r \cos\theta \cdot r(\cos^2\varphi \cos^2\theta + \sin^2\varphi \cos^2\theta) \\
&= r^2 \sin\theta \cos\theta + r^2 \cos^3\theta \\
&= r^2 \cos\theta.
\end{aligned}$$

Offensichtlich ist $\det J_F(r, \varphi, \theta) > 0$ im Definitionsbereich von F. Auch hier kann man zeigen, daß F bijektiv ist.

Man beachte, daß die Kugelkoordinaten in der Literatur nicht einheitlich definiert werden!

§ 4 Implizite Funktionen

Wir wollen uns mit nichtlinearen Gleichungen beschäftigen:

$$f(\mathbf{x}) = \mathbf{0},$$

wobei $f = (f_1, \dots, f_m)$ stetig differenzierbar auf einer offenen Menge $B \subset \mathbb{R}^n$ und $m < n$ ist. Dann hat man m skalare Gleichungen für $n = m + k$ Variable, $k > 0$, also ein „unterbestimmtes Sytem". In der Linearen Algebra führt das zu einem mindestens 1-dimensionalen Lösungsraum. Wie ein solcher Lösungsraum im nichtlinearen Fall aussehen könnte, ahnen wir noch nicht. Also betrachten wir zunächst ein einfaches Beispiel:

Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x, y) := x^2 + y^2 - 1$. Dann ist

$$\{(x, y) \in \mathbb{R}^2 \,:\, f(x, y) = 0\} = \{\mathbf{x} \in \mathbb{R}^2 \,:\, \|\mathbf{x}\| = 1\} =: S^1$$

der Einheitskreis.

Ist $(a, b) \in S^1$, $a \neq \pm 1$, so gilt in der Nähe von (a, b):

$$\begin{aligned} f(x, y) = 0 \quad &\Longleftrightarrow \quad y^2 = 1 - x^2 \\ &\Longleftrightarrow \quad y = g(x) := \pm\sqrt{1 - x^2}. \end{aligned}$$

Die Lösungsmenge sieht also in der Nähe von (a, b) wie ein Graph aus. Es gibt eine Umgebung $U = V \times W$ von (a, b), so daß gilt:

$$\{(x, y) \in U \mid f(x, y) = 0\} = \{(x, g(x)) \mid x \in V\}.$$

Man kann übrigens sehr leicht die Ableitung von g berechnen. Da $f(x, g(x)) \equiv 0$ ist, folgt mit der Kettenregel:

$$0 = \frac{\partial f}{\partial x}(x, g(x)) \cdot 1 + \frac{\partial f}{\partial y}(x, g(x)) \cdot g'(x),$$

also

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))} = -\frac{2x}{2g(x)} = -\frac{x}{g(x)}.$$

Bei dieser Umformung hätten wir natürlich erst einmal überprüfen müssen, ob $f_y(x, g(x))$ in der Nähe von $x = a$ nicht verschwindet. Tatsächlich ist $f_y(a, b) = 2b = 0$ für ein $(a, b) \in S^1$ genau dann, wenn $a^2 = 1$ ist, also $a = \pm 1$. Das ist gerade die Bedingung für die Auflösbarkeit nach y.

Der Kreis S^1 ist eine so symmetrische Figur, daß nicht einzusehen ist, warum die Punkte $(1, 0)$ und $(-1, 0)$ eine Ausnahme bilden sollten. Wie können wir uns da behelfen? Wenn wir das Koordinatensystem um 90° drehen, dann sieht der Kreis auch dort lokal wie ein Graph aus, allerdings wie der Graph einer Funktion $x = h(y)$. Tatsächlich ist dort $h(y) = \pm\sqrt{1 - y^2}$.

Wie kann man erkennen, nach welcher Variablen aufgelöst werden kann? Der Kreis kann überall dort als Graph einer Funktion $y = g(x)$ aufgefaßt werden, wo er keine vertikale Tangente besitzt, und er kann überall dort als Graph einer Funktion $x = h(y)$ aufgefaßt werden, wo er keine horizontale Tangente besitzt. Nun ist S^1 eine Niveaulinie von f, und der Gradient von f steht jeweils auf der Tangenten senkrecht. Damit haben wir:

$$\begin{array}{rrcl} & S^1 \text{ Graph von } y = g(x) & \iff & \text{Tangente nicht vertikal} \\ & & \iff & \nabla f(x,y) \text{ nicht horizontal} \\ & & \iff & \dfrac{\partial f}{\partial y}(x,y) \neq 0 \\ \text{und} & S^1 \text{ Graph von } x = h(y) & \iff & \text{Tangente nicht horizontal} \\ & & \iff & \nabla f(x,y) \text{ nicht vertikal} \\ & & \iff & \dfrac{\partial f}{\partial x}(x,y) \neq 0. \end{array}$$

Wir betrachten nun eine offene Menge $B \subset \mathbb{R}^k \times \mathbb{R}^m$ und eine stetig differenzierbare Abbildung $f : B \to \mathbb{R}^m$, also ein System von m nichtlinearen Gleichungen für $k+m$ Variable. Den Satz der ersten k Variablen $x_1, \ldots, x_k$ fassen wir zu einem Vektor $\mathbf{x}$, den der folgenden m Variablen $x_{k+1}, \ldots, x_{k+m}$ zu einem Vektor $\mathbf{y}$ zusammen. Dann definieren wir:

$$\frac{\partial f}{\partial \mathbf{x}} := \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_k} \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_k} \end{pmatrix} \quad \text{und} \quad \frac{\partial f}{\partial \mathbf{y}} := \begin{pmatrix} \dfrac{\partial f_1}{\partial x_{k+1}} & \cdots & \dfrac{\partial f_1}{\partial x_{k+m}} \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_{k+1}} & \cdots & \dfrac{\partial f_m}{\partial x_{k+m}} \end{pmatrix}.$$

Damit ist

$$J_f(\mathbf{x},\mathbf{y}) = \left(\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x},\mathbf{y}) \;\Big|\; \frac{\partial f}{\partial \mathbf{y}}(\mathbf{x},\mathbf{y}) \right).$$

4.1 Satz über implizite Funktionen.

Ist $f(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ und die Matrix $\dfrac{\partial f}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \in M_{m,m}(\mathbb{R})$ regulär, so gibt es Umgebungen $U(\mathbf{x}_0)$, $V(\mathbf{y}_0)$ mit $U \times V \subset B$ und eine stetig differenzierbare Abbildung $g : U \to V$, so daß gilt:

1. $g(\mathbf{x}_0) = \mathbf{y}_0$.
2. *Für $(\mathbf{x},\mathbf{y}) \in U \times V$ gilt: $f(\mathbf{x},\mathbf{y}) = \mathbf{0} \iff \mathbf{y} = g(\mathbf{x})$.*

   Also ist $f(\mathbf{x}, g(\mathbf{x})) \equiv \mathbf{0}$.
3. $J_g(\mathbf{x}) = -\left(\dfrac{\partial f}{\partial \mathbf{y}}(\mathbf{x}, g(\mathbf{x}))\right)^{-1} \cdot \dfrac{\partial f}{\partial \mathbf{x}}(\mathbf{x}, g(\mathbf{x}))$ *auf U.*

Beweis: Sei $n := k + m$. Der Trick besteht darin, den Raum um $(\mathbf{x}_0, \mathbf{y}_0)$ herum so differenzierbar zu verbiegen, daß aus der Nullstellenmenge

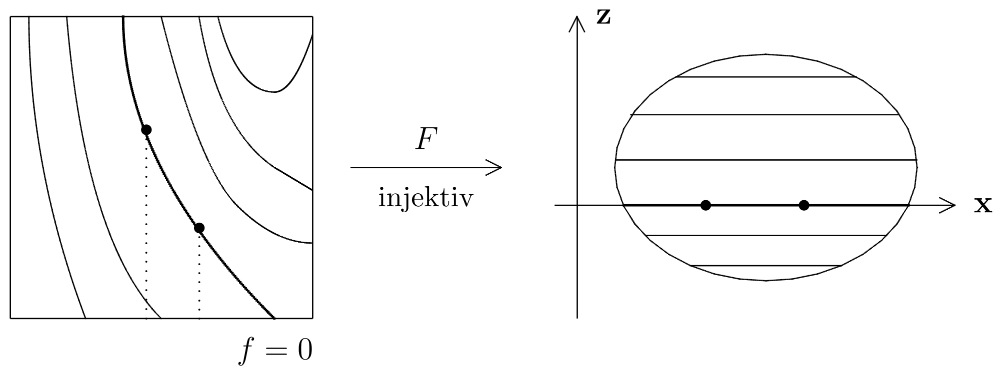
$$N := \{(x_1, \ldots, x_n) \mid f_1(x_1, \ldots, x_n) = \ldots = f_m(x_1, \ldots, x_n) = 0\}$$

ein Ebenenstück der Gestalt

$$E = \{(x_1, \ldots, x_k, z_{k+1}, \ldots, z_n) \,:\, z_{k+1} = \ldots = z_n = 0\}$$

wird. Zu diesem Zweck definieren wir $F : B \to \mathbb{R}^n$ durch

$$F(\mathbf{x}, \mathbf{y}) := (\mathbf{x}, f(\mathbf{x}, \mathbf{y})).$$



Dann ist

$$J_F(\mathbf{x}_0, \mathbf{y}_0) = \left(\begin{array}{c|c} E_k & \mathbf{0} \\ \hline \dfrac{\partial f}{\partial \mathbf{x}}(\mathbf{x}_0, \mathbf{y}_0) & \dfrac{\partial f}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \end{array}\right),$$

und daher

$$\det J_F(\mathbf{x}_0, \mathbf{y}_0) = \det \frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0) \neq 0.$$

Das bedeutet, daß F in $(\mathbf{x}_0, \mathbf{y}_0)$ lokal umkehrbar ist. Wir setzen $H := F^{-1}$ (in der Nähe von $F(\mathbf{x}_0, \mathbf{y}_0)$). Offensichtlich ist $H(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, h(\mathbf{u}, \mathbf{v}))$, mit einer differenzierbaren Abbildung h.

Ist $f(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, so ist $F(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{0})$, also $(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}, \mathbf{0}) = (\mathbf{x}, h(\mathbf{x}, \mathbf{0}))$. Deshalb setzen wir

$$g(\mathbf{x}) := h(\mathbf{x}, \mathbf{0}).$$

Offensichtlich ist g stetig differenzierbar. Ist $f(\mathbf{x}, \mathbf{y}) = \mathbf{0}$, so ist nach Konstruktion $\mathbf{y} = g(\mathbf{x})$. Und umgekehrt ist

$$f(\mathbf{x}, g(\mathbf{x})) = f(\mathbf{x}, h(\mathbf{x}, \mathbf{0})) = \mathbf{0},$$

denn es ist ja $(\mathbf{x}, f(\mathbf{x}, h(\mathbf{x}, \mathbf{0}))) = F \circ H(\mathbf{x}, \mathbf{0}) = (\mathbf{x}, \mathbf{0})$. Und weil $f(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ ist, ist $\mathbf{y}_0 = g(\mathbf{x}_0)$.

Aus der Beziehung $f(\mathbf{x}, g(\mathbf{x})) \equiv \mathbf{0}$ folgt schließlich mit der Kettenregel:

$$\mathbf{0} = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}, g(\mathbf{x})) \cdot E_k + \frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}, g(\mathbf{x})) \cdot J_g(\mathbf{x}),$$

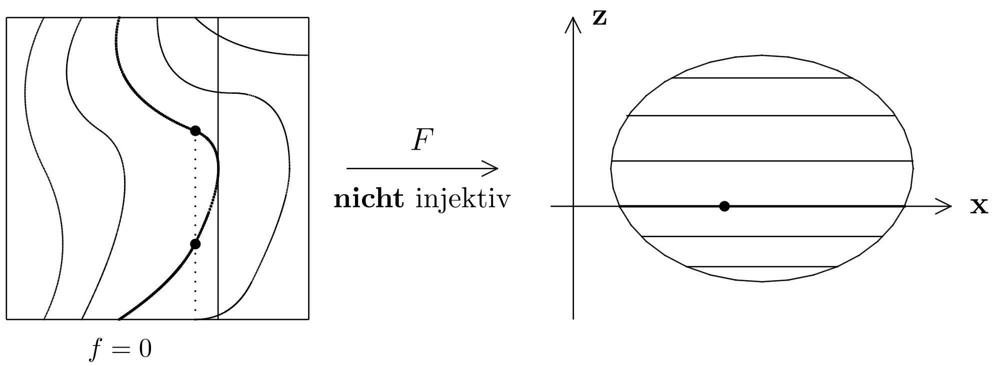
also

$$J_g(\mathbf{x}) = -\left(\frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}, g(\mathbf{x}))\right)^{-1} \cdot \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}, g(\mathbf{x})).$$

Man beachte hier die Reihenfolge bei der Matrizenmultiplikation!

Das Ganze gilt in der Nähe von $(\mathbf{x}_0, \mathbf{y}_0)$. Wählt man die Umgebungen U und V klein genug, so ist alles gezeigt. ■

Bemerkung. Das Verfahren kann nicht funktionieren, wenn man die Regularität von $\dfrac{\partial f}{\partial \mathbf{y}}(\mathbf{x}_0, \mathbf{y}_0)$ nicht fordert. Man kann anschaulich sehen, daß die Abbildung F dann nicht mehr injektiv ist.



Häufig lassen sich dann aber die Koordinaten so vertauschen, daß anschließend doch die Voraussetzungen erfüllt sind.

Beispiele.

1. Betrachten wir noch einmal den Kreis

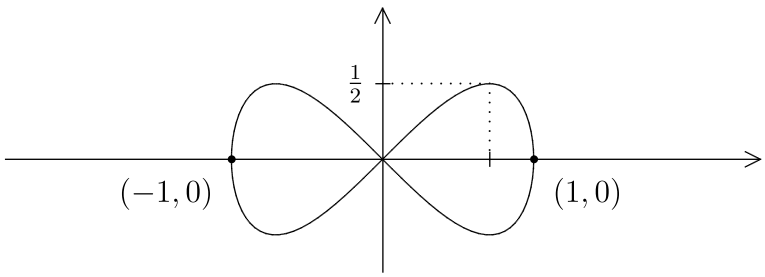
$$S^1 = \{(x, y) \mid f(x, y) := x^2 + y^2 - 1 = 0\}.$$

Für $y \neq 0$ (also $x \neq \pm 1$) ist $\dfrac{\partial f}{\partial y}(x, y) = 2y \neq 0$. Also kann man den Satz über implizite Funktionen anwenden und die Gleichung $f(x, y) = 0$ lokal nach y auflösen: $y = g(x)$. Die Formel für die Ableitung von g ergibt hier:

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))} = -\frac{x}{g(x)} = -\frac{x}{\sqrt{1 - x^2}}.$$

Leider ist die Auflösung nicht immer so schön konkret durchführbar!

2. Sei $f(x,y) := x^2(1-x^2) - y^2$. Die Kurve $C := \{(x,y) \mid f(x,y) = 0\}$ ist eine sogenannte „Lemniskate“:



Wir berechnen die partiellen Ableitungen:

$$\begin{aligned} f_x(x,y) &= 2x - 4x^3 = 2x(1-2x^2), \\ f_y(x,y) &= -2y. \end{aligned}$$

Im Nullpunkt ist die Gleichung überhaupt nicht auflösbar. Das liegt anschaulich daran, daß der dort auftretende Kreuzungspunkt aus keiner Richtung wie ein Graph aussieht.

In den Punkten $(1,0)$ und $(-1,0)$ ist jeweils $f_y(x,y) = 0$, also keine Auflösung nach y möglich. Allerdings ist dort $f_x(x,y) \neq 0$, wir können also lokal nach x auflösen. Das ist hier sogar konkret möglich, die Gleichung $x^4 - x^2 + y^2 = 0$ führt auf

$$x = \pm\frac{1}{2}\sqrt{2 \pm 2\sqrt{1-4y^2}}.$$

Läßt man y gegen Null gehen, so muß x^2 gegen 1 streben. Das schließt unter der ersten Wurzel das Minus–Zeichen aus, und man bekommt:

$$\begin{aligned} x &= +\frac{1}{2}\sqrt{2+2\sqrt{1-4y^2}} \quad \text{bei } (1,0) \\ \text{und} \quad x &= -\frac{1}{2}\sqrt{2+2\sqrt{1-4y^2}} \quad \text{bei } (-1,0). \end{aligned}$$

In allen anderen Punkten ist $f_y(x,y) \neq 0$, denn wenn $y = 0$ und $f(x,y) = 0$ ist, dann kann nur $x = 0$ oder $x = \pm 1$ sein. Dann ist

$$y = \pm\sqrt{x^2(1-x^2)},$$

wobei das Vorzeichen davon abhängt, ob man sich gerade in der oberen oder in der unteren Halbebene befindet.

Rechnen wir noch im Falle der oberen Halbebene die Ableitung von $y = g(x)$ aus:

$$g'(x) = -\frac{f_x(x,g(x))}{f_y(x,g(x))} = -\frac{2x-4x^3}{-2g(x)} = \frac{x(1-2x^2)}{\sqrt{x^2(1-x^2)}}\,.$$

Diese Beziehung gilt natürlich nicht bei $x = 0$. Für $0 < x < 1$ ist $g'(x) = 0$ genau dann erfüllt, wenn $1 - 2x^2 = 0$ ist, also $x = \frac{1}{2}\sqrt{2}$. Dort ist $y = \frac{1}{2}$. Offensichtlich liegt ein Maximum vor, und mit dieser Information kann man schon eine recht gute Skizze der Lemniskate erstellen.

Wir wollen im folgenden Teilmengen des $\mathbb{R}^n$ untersuchen, die überall die Bedingungen des Satzes über implizite Funktionen erfüllen, also lokal wie der Graph einer differenzierbaren Abbildung aussehen.

Ist $B \subset \mathbb{R}^n$ eine offene Teilmenge, so heißt eine Menge $M \subset B$ *abgeschlossen in* B, falls $B \setminus M$ offen ist. Dann braucht M allerdings nicht abgeschlossen im $\mathbb{R}^n$ zu sein.

Definition. Sei $B \subset \mathbb{R}^n$ offen, $0 \le d < n$. Eine in B abgeschlossene Teilmenge $M \subset B$ heißt d–codimensionale *(Unter–)Mannigfaltigkeit*, wenn es zu jedem Punkt $\mathbf{x}_0 \in M$ eine offene Umgebung $U(\mathbf{x}_0) \subset B$ und stetig differenzierbare Funktionen $f_1, \ldots, f_d : U \to \mathbb{R}$ gibt, so daß gilt:

1. $U \cap M = \{\mathbf{x} \in U \,:\, f_1(\mathbf{x}) = \ldots = f_d(\mathbf{x}) = 0\}$.
2. $\nabla f_1(\mathbf{x}), \ldots, \nabla f_d(\mathbf{x})$ sind linear unabhängig für alle $\mathbf{x} \in U$.

Die Zahl $n - d$ nennt man die *Dimension* von M in $\mathbf{x}_0$.

Bemerkung. Wir können die Bedingung in der Definition auch kurz so beschreiben: $\exists\, f : U \to \mathbb{R}^d$ mit $U \cap M = f^{-1}(\mathbf{0})$, so daß $Df(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^d$ für jedes $\mathbf{x} \in U$ surjektiv ist.

4.2 Satz. *Sei $M \subset B$ eine abgeschlossene Teilmenge. Dann sind die folgenden drei Aussagen über M äquivalent:*

1. *M ist eine d–codimensionale Untermannigfaltigkeit von B.*
2. *Lokal sieht M nach geeigneter Numerierung der Koordinaten wie der Graph einer stetig differenzierbaren Abbildung von $\mathbb{R}^{n-d}$ nach $\mathbb{R}^d$ aus.*
3. *Zu jedem Punkt $\mathbf{x}_0 \in M$ gibt es eine offene Umgebung $U(\mathbf{x}_0) \subset B$ und eine in $\mathbf{x}_0$ reguläre Abbildung $F : U \to \mathbb{R}^n$, so daß gilt:*

$$F(U \cap M) = \{\mathbf{y} \in F(U) \,:\, y_{n-d+1} = \ldots = y_n = 0\}.$$

Beweis: (1) $\Longrightarrow$ (2)

Sei $M \cap U = \{f_1 = \ldots = f_d = 0\}$. Nach geeigneter Numerierung der Koordinaten ist $\mathbf{x} = (\mathbf{x}^*, \mathbf{x}^{**}) \in \mathbb{R}^{n-d} \times \mathbb{R}^d$, und die Funktionalmatrix von $f = (f_1, \ldots, f_d)$ hat die Gestalt

$$J_f(\mathbf{x}) = \left(\frac{\partial f}{\partial \mathbf{x}^*}(\mathbf{x}) \;\middle|\; \frac{\partial f}{\partial \mathbf{x}^{**}}(\mathbf{x}) \right),$$

mit $\det\left(\dfrac{\partial f}{\partial \mathbf{x}^{**}}(\mathbf{x})\right) \neq 0$. Aus dem Satz über implizite Funktionen folgt, daß die Nullstellenmenge von f (also M) lokal der Graph einer Funktion $\mathbf{x}^{**} = g(\mathbf{x}^*)$ ist.

(2) $\Longrightarrow$ (3)

Ist $M \cap (U \times V) = \{(\mathbf{x}^*, \mathbf{x}^{**}) \in U \times V \,:\, \mathbf{x}^{**} = g(\mathbf{x}^*)\}$, so definieren wir $F : U \times V \to \mathbb{R}^n$ durch

$$F(\mathbf{x}^*, \mathbf{x}^{**}) := (\mathbf{x}^*, \mathbf{x}^{**} - g(\mathbf{x}^*)).$$

Dann ist

$$J_F(\mathbf{x}^*, \mathbf{x}^{**}) = \left(\begin{array}{c|c} E_{n-d} & \mathbf{O} \\ \hline -J_g(\mathbf{x}^*) & E_d \end{array} \right).$$

Also ist F ein lokaler Diffeomorphismus, und

$$F(M \cap (U \times V)) = \{(\mathbf{y}^*, \mathbf{y}^{**}) \,:\, \mathbf{y}^{**} = \mathbf{0}\}.$$

(3) $\Longrightarrow$ (1)

Ist ein Diffeomorphismus $F = (F_1, \ldots, F_n) : U \to \mathbb{R}^n$ gegeben, mit

$$F^{-1}(\{\mathbf{y} \,:\, y_{n-d+1} = \ldots = y_n = 0\}) = U \cap M,$$

so setzen wir $f_i := F_{n-d+i}$ für $i = 1, \ldots, d$. Dann ist $M = \{f_1 = \ldots = f_d = 0\}$, und da J_F regulär ist, sind $\nabla f_1, \ldots, \nabla f_d$ linear unabhängig. ∎

Beispiele.

1. Die Kreislinie $S^1 \subset \mathbb{R}^2$ ist eine Untermannigfaltigkeit, die Lemniskate ist es nicht, denn in der Nähe des Nullpunktes kann sie auf keinerlei Weise als Graph geschrieben werden.

2. Sei $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch $f(x_1, \ldots, x_n) := x_1^2 + \cdots + x_n^2 - 1$. Dann ist

$$\{\mathbf{x} \in \mathbb{R}^n \,:\, f(\mathbf{x}) = 0\} = \{\mathbf{x} \in \mathbb{R}^n \,:\, \|\mathbf{x}\| = 1\} = S^{n-1}$$

   die $(n-1)$–Sphäre.

   $\nabla f(\mathbf{x}) = 2\mathbf{x}$ ist auf S^{n-1} überall $\neq 0$, hat also den Rang 1. Damit S^{n-1} eine $(n-1)$–dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$.

3. Sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x, y) := (x-1)^3 - y^2$. Dann nennt man $N := \{(x, y) \,:\, f(x, y) = 0\}$ eine Neilsche Parabel.

Der Gradient $\nabla f(x,y) = (3(x-1)^2, 2y)$ verschwindet in dem Punkt $(1,0)$, der leider auf N liegt. Also scheint N keine Untermannigfaltigkeit zu sein. Tatsächlich hat N in $(1,0)$ eine „Spitze“ und ist deshalb wirklich keine Untermannigfaltigkeit. Die Frage ist, wie man das exakt beweisen kann.

Generell gibt es dafür kein Kochrezept, aber in diesem Beispiel ist es einfach. Für jedes $\varepsilon > 0$ liegen die beiden Punkte $(1+\varepsilon, \pm\sqrt{\varepsilon^3})$ auf N. Deshalb kann N in der Nähe von $(0,0)$ nicht der Graph einer Funktion $y = g(x)$ sein. Zwar ist N in der Nähe des Ursprungs Graph der Funktion $x = 1 + \sqrt[3]{y^2}$, aber die ist bei $y = 0$ nicht differenzierbar. Sie kann auch nicht durch eine stetig differenzierbare Funktion mit gleichem Graph ersetzt werden. Durch $\alpha(t) := (1+t^2, t^3)$ wird N parametrisiert, und die Richtung der Tangente an N im Punkt $\alpha(t)$ ist durch den Vektor $\alpha'(t) = (2t, 3t^2) = t \cdot (2, 3t)$ gegeben. Läßt man t gegen 0 gehen, so erhält man den Richtungsvektor $(2,0)$, der parallel zur x-Achse ist. Jede Funktion $x = h(y)$, deren Graph nahe $(0,0)$ mit N übereinstimmt, hätte also bei $y = 0$ eine senkrechte Tangente und könnte nicht stetig differenzierbar sein.

Sei $B \subset \mathbb{R}^n$ offen, $M \subset B$ eine (abgeschlossene) Untermannigfaltigkeit. Weiter sei $\mathbf{x}_0 \in M$ und $W(\mathbf{x}_0) \subset B$ eine offene Umgebung der Gestalt $W = U \times V \subset \mathbb{R}^{n-d} \times \mathbb{R}^d$, so daß unter Verwendung der Koordinaten $(\mathbf{u}, \mathbf{z}) = \mathbf{x}$ gilt:

1. Es gibt eine stetig differenzierbare Abbildung $f : W \to \mathbb{R}^d$ mit $W \cap M = f^{-1}(\mathbf{0})$, so daß $Df(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^d$ für jedes $\mathbf{x} \in W$ surjektiv ist.

2. Es gibt eine stetig differenzierbare Abbildung $g : U \to V$ mit

$$M \cap W = \{(\mathbf{x}^*, \mathbf{x}^{**}) \in U \times V \,:\, \mathbf{x}^{**} = g(\mathbf{x}^*)\}.$$

Dann ist der affine Tangentialraum an M als Graph in $\mathbf{x}_0 = (\mathbf{u}_0, g(\mathbf{u}_0))$ gegeben durch

$$\begin{aligned} T^{\text{aff}}_{\mathbf{x}_0}(M) &= \{(\mathbf{u}, \mathbf{z}) \in \mathbb{R}^{n-d} \times \mathbb{R}^d \,:\, \mathbf{z} - g(\mathbf{u}_0) = Dg(\mathbf{u}_0)(\mathbf{u} - \mathbf{u}_0)\} \\ &= (\mathbf{u}_0, g(\mathbf{u}_0)) + \{(\mathbf{h}, \mathbf{v}) \,:\, \mathbf{v} = Dg(\mathbf{u}_0)(\mathbf{h})\}. \end{aligned}$$

Den Vektorraum $T_{\mathbf{x}_0}(M) := \{(\mathbf{h}, \mathbf{v}) \,:\, \mathbf{v} = Dg(\mathbf{u}_0)(\mathbf{h})\}$ nennen wir den *Tangentialraum* an M in $\mathbf{x}_0$.

4.3 Satz. *Es ist* $T_{\mathbf{x}_0}(M) = \operatorname{Ker}(Df(\mathbf{x}_0))$.

BEWEIS: Sei $T := T_{\mathbf{x}_0}(M)$ und $K := \operatorname{Ker}(Df(\mathbf{x}_0))$. Weil $f(\mathbf{u}, g(\mathbf{u})) \equiv 0$ ist, ist

$$\mathbf{0} = Df(\mathbf{u}, g(\mathbf{u}))(\mathbf{h}, Dg(\mathbf{u})(\mathbf{h})) \text{ für alle } \mathbf{h} \in \mathbb{R}^{n-d}.$$

Also liegt T in K. Nun ist $\dim(K) = n - \dim(\operatorname{Im} Df(\mathbf{x}_0)) = n - d$, und $\dim(T) = \dim(\mathbb{R}^{n-d}) = n - d$. Also ist $T = K$. ∎

Es gibt allerdings auch eine Beschreibung von $T_{\mathbf{x}_0}(M)$, die nicht von den Funktionen abhängt, durch die M gegeben ist.

4.4 Satz. *Ein Vektor $\mathbf{w} \in \mathbb{R}^n$ liegt genau dann in $T_{\mathbf{x}_0}(M)$, wenn es einen stetig differenzierbaren Weg $\alpha : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ gibt, so daß gilt:*

1. *Die Spur von α liegt ganz in M.*
2. *Es ist $\alpha(0) = \mathbf{x}_0$ und $\alpha'(0) = \mathbf{w}$.*

BEWEIS: Sei $M \cap U = f^{-1}(\mathbf{0})$, $\mathbf{x}_0 \in M \cap U$ und $Df(\mathbf{x}_0)$ surjektiv.

a) Sei $\alpha(0) = \mathbf{x}_0$ und $\alpha'(0) = \mathbf{w}$. Da $f \circ \alpha \equiv 0$ ist, ist $Df(\mathbf{x}_0)(\alpha'(0)) = 0$, also $\mathbf{w} \in \operatorname{Ker}(Df(\mathbf{x}_0)) = T_{\mathbf{x}_0}(M)$.

b) Sei $M \cap U = \{(\mathbf{u}, \mathbf{z}) : \mathbf{z} = g(\mathbf{u})\}$, und $\mathbf{w} \in T_{\mathbf{x}_0}(M)$. Dann ist $\mathbf{w} = (\mathbf{h}, Dg(\mathbf{u}_0)(\mathbf{h}))$. Wir setzen

$$\alpha(t) := (\mathbf{u}_0 + t\mathbf{h}, g(\mathbf{u}_0 + t\mathbf{h})).$$

Dann liegt die Spur von α in M, es ist $\alpha(0) = (\mathbf{u}_0, g(\mathbf{u}_0)) = \mathbf{x}_0$ und $\alpha'(0) = (\mathbf{h}, Dg(\mathbf{u}_0)(\mathbf{h})) = \mathbf{w}$. ∎

Beispiele.

1. Die Sphäre $S^{n-1} = \{\mathbf{x} : \|\mathbf{x}\| = 1\}$ ist die Nullstellenmenge von $f(\mathbf{x}) := \|\mathbf{x}\|^2 - 1$. Für $\mathbf{x}_0 \in S^{n-1}$ ist

   $$T_{\mathbf{x}_0}(S^{n-1}) = \operatorname{Ker} Df(\mathbf{x}_0) = \{\mathbf{v} \in \mathbb{R}^n : \nabla f(\mathbf{x}_0) \bullet \mathbf{v} = 0\} = \{\mathbf{v} : \mathbf{x}_0 \bullet \mathbf{v} = 0\}.$$

2. Wir betrachten noch eine neue Mannigfaltigkeit:

   $$O(n) := \{A \in M_{n,n}(\mathbb{R}) : A \cdot A^t = E_n\}$$

   ist die sogenannte *orthogonale Gruppe*. Definiert man

   $$f : M_{n,n}(\mathbb{R}) \to \operatorname{Sym}(n) := \{B \in M_{n,n}(\mathbb{R}) : B^t = B\}$$

   durch $f(A) := A \cdot A^t - E_n$, so ist $O(n) = f^{-1}(0)$. Wir müssen noch die Ableitung von f berechnen. Die Abbildung $\Phi : M_{n,n}(\mathbb{R}) \times M_{n,n}(\mathbb{R}) \to M_{n,n}(\mathbb{R})$ mit $\Phi(A, B) := A \cdot B$ ist bilinear. Weil Φ differenzierbar mit $D\Phi(A_0, B_0)(X, Y) = \Phi(X, B_0) + \Phi(A_0, Y)$ und $A \mapsto A^t$ linear ist, folgt:

   $$Df(A_0)(Z) = \Phi(Z, A_0^t) + \Phi(A_0, Z^t) = Z \cdot A_0^t + A_0 \cdot Z^t.$$

   Nun müssen wir noch zeigen, daß $Df(A_0) : M_{n,n}(\mathbb{R}) \to \operatorname{Sym}(n)$ surjektiv ist. Ist eine symmetrische Matrix S gegeben, so setzen wir $Z := \frac{1}{2} S \cdot A_0$. Dann ist tatsächlich

$$Df(A_0)(Z) = \frac{1}{2}[S \cdot A_0 \cdot A_0^t + A_0 \cdot A_0^t \cdot S^t] = S\,.$$

Also ist $O(n)$ eine Untermannigfaltigkeit von $M_{n,n}(\mathbb{R})$.

Der Tangentialraum ist zumindest für den Spezialfall $A_0 = E_n$ leicht berechnet. Es ist

$$T_{E_n}(O(n)) = \operatorname{Ker} Df(E_n) = \operatorname{Ker}(Z \mapsto Z+Z^t) = \{Z \in M_{n,n}(\mathbb{R}) \,:\, Z^t = -Z\}.$$

Das ist der Raum der schiefsymmetrischen Matrizen. Speziell folgt nun:

$$\dim(O(n)) = \frac{n^2-n}{2}\,.$$

Wir wollen als nächstes einen vernünftigen Differenzierbarkeitsbegriff für Funktionen auf Mannigfaltigkeiten entwickeln.

Definition. Sei M eine k-dimensionale Untermannigfaltigkeit einer offenen Teilmenge des $\mathbb{R}^n$ und $\mathbf{x}_0$ ein Punkt von M. Eine *(lokale) Parametrisierung* von M in $\mathbf{x}_0$ ist eine stetig differenzierbare Abbildung $\varphi : U \to \mathbb{R}^n$ mit folgenden Eigenschaften:

1. $U \subset \mathbb{R}^k$ ist eine offene Teilmenge, und es gibt ein $\mathbf{u}_0 \in U$ mit $\varphi(\mathbf{u}_0) = \mathbf{x}_0$.
2. $D\varphi(\mathbf{u}_0) : \mathbb{R}^k \to \mathbb{R}^n$ ist injektiv,
3. Es gibt eine offene Umgebung $W(\mathbf{x}_0) \subset \mathbb{R}^n$, so daß $\varphi : U \to W \cap M$ ein Homöomorphismus ist.

4.5 Satz. *Eine abgeschlossene Teilmenge M einer offenen Menge $B \subset \mathbb{R}^n$ ist genau dann eine Untermannigfaltigkeit, wenn es zu jedem $\mathbf{x}_0 \in M$ eine lokale Parametrisierung von M in $\mathbf{x}_0$ gibt.*

BEWEIS: 1) Sei zunächst M eine Untermannigfaltigkeit. Es gibt eine Umgebung der Gestalt $W = U \times V \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$ von $\mathbf{x}_0$ und eine stetig differenzierbare Abbildung $g : U \to V$, so daß $W \cap M = \{(\mathbf{u},\mathbf{z}) \in U \times V \,:\, \mathbf{z} = g(\mathbf{u})\}$ ist. Wir setzen dann $\varphi(\mathbf{u}) := (\mathbf{u}, g(\mathbf{u}))$. Offensichtlich ist φ stetig differenzierbar und injektiv, und $D\varphi(\mathbf{u})$ hat Rang k, ist also injektiv.

Die Umkehrabbildung $\varphi^{-1}(\mathbf{u},\mathbf{z}) = \mathbf{u}$ ist stetig, also ist $\varphi : U \to (U \times V) \cap M$ ein Homöomorphismus.

2) Die umgekehrte Richtung ist etwas schwieriger. Sei $\varphi : U \to \mathbb{R}^n$ stetig differenzierbar, $\varphi(\mathbf{u}_0) = \mathbf{x}_0$ und $D\varphi(\mathbf{u}_0) : \mathbb{R}^k \to \mathbb{R}^n$ injektiv. Außerdem gebe es eine Umgebung $W(\mathbf{x}_0) \subset \mathbb{R}^n$, so daß $\varphi : U \to W \cap M$ ein Homöomorphismus ist.

O.B.d.A. sei $\det\left(\frac{\partial \varphi_i}{\partial u_j}(\mathbf{u}_0) \;\Big|\; i,j = 1,\ldots,k\right) \neq 0$. Wir benutzen die kanonische Projektion $\pi : \mathbb{R}^k \times \mathbb{R}^{n-k} \to \mathbb{R}^k$ mit $\pi(x_1,\ldots,x_n) := (x_1,\ldots,x_k)$. Dann ist

$\det J_{\pi\circ\varphi}(\mathbf{u}_0) \neq 0$, und es gibt offene Umgebungen $U_1 = U_1(\mathbf{u}_0) \subset \mathbb{R}^k$ und $U_2 = U_2(\pi \circ \varphi(\mathbf{u}_0))$, so daß $\pi \circ \varphi : U_1 \to U_2$ ein Diffeomorphismus ist. Offensichtlich ist dann $\pi|_{\varphi(U_1)}$ injektiv. Sei $\psi : U_2 \to U_1$ die Umkehrabbildung von $\pi \circ \varphi$ und

$$g(y_1, \ldots, y_k) := (\varphi_{k+1} \circ \psi(y_1, \ldots, y_k), \ldots, \varphi_n \circ \psi(y_1, \ldots, y_k)).$$

Dann gilt für $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$:

$$\begin{aligned}
\mathbf{x} \in \varphi(U_1) \quad &\iff \quad \exists\, \mathbf{u} \in U_1 \text{ mit } \mathbf{x} = \varphi(\mathbf{u}) \\
&\iff \quad \exists\, \mathbf{u} \in U_1 \text{ mit } (x_1, \ldots, x_k) = \pi \circ \varphi(\mathbf{u}) \\
&\qquad\quad \text{und } (x_{k+1}, \ldots, x_n) = (\varphi_{k+1}(\mathbf{u}), \ldots, \varphi_n(\mathbf{u})) \\
&\iff \quad \mathbf{u} := \psi(x_1, \ldots, x_k) \in U_1 \\
&\qquad\quad \text{und } (x_{k+1}, \ldots, x_n) = (\varphi_{k+1}(\mathbf{u}), \ldots, \varphi_n(\mathbf{u})) \\
&\iff \quad (x_{k+1}, \ldots, x_n) = g(x_1, \ldots, x_k).
\end{aligned}$$

Also ist $\varphi(U_1)$ der Graph von g und M damit eine Untermannigfaltigkeit. ■

4.6 Satz. *Ist $\varphi : U \to M$ eine lokale Parametrisierung und $\varphi(\mathbf{u}_0) = \mathbf{x}_0$, so ist* $\operatorname{Im} D\varphi(\mathbf{u}_0) = T_{\mathbf{x}_0}(M)$.

Beweis: Sei $\dim(M) = k$, $M = f^{-1}(0)$ und $Df(\mathbf{x}_0)$ surjektiv. Dann ist $f \circ \varphi(\mathbf{u}) \equiv 0$, und daher

$$0 = D(f \circ \varphi)(\mathbf{u}_0) = Df(\mathbf{x}_0) \circ D\varphi(\mathbf{u}_0).$$

Das bedeutet, daß $\operatorname{Im} D\varphi(\mathbf{u}_0) \subset \operatorname{Ker} Df(\mathbf{x}_0)$ ist. Aus Dimensionsgründen folgt die Behauptung. ■

4.7 Satz. *Sei $B \subset \mathbb{R}^n$ offen, $M \subset B$ eine k-dimensionale Untermannigfaltigkeit und $\mathbf{x}_0 \in M$. Sind W_1, W_2 zwei offene Umgebungen von $\mathbf{x}_0$ und $\varphi_1 : U_1 \to W_1 \cap M$ und $\varphi_2 : U_2 \to W_2 \cap M$ zwei lokale Parametrisierungen, so ist*

$$\varphi_2^{-1} \circ \varphi_1 : \varphi_1^{-1}(W_1 \cap W_2 \cap M) \to \varphi_2^{-1}(W_1 \cap W_2 \cap M)$$

ein Diffeomorphismus.

Beweis: Der Beweis erfordert einen kleinen Trick. Wir können annehmen, daß die ersten k Zeilen von $J_{\varphi_1}(\mathbf{u})$ linear unabhängig sind. Dann definieren wir $F : U_1 \times \mathbb{R}^{n-k} \to \mathbb{R}^n$ durch $F(\mathbf{u}, \mathbf{t}) := \varphi_1(\mathbf{u}) + (\mathbf{0}, \mathbf{t})$. Es ist

$$J_F(\mathbf{u}, \mathbf{t}) = \left(J_{\varphi_1}(\mathbf{u}) \;\middle|\; \frac{0}{E_{n-k}} \right),$$

also $\det J_F(\mathbf{u}, \mathbf{t}) \neq 0$. Damit ist F ein lokaler Diffeomorphismus.

$F^{-1} \circ \varphi_2$ ist demnach stetig differenzierbar. Ist $\varphi_1(\mathbf{u}) = \varphi_2(\mathbf{v})$, so ist

$$\begin{aligned} F^{-1} \circ \varphi_2(\mathbf{v}) &= F^{-1} \circ \varphi_1(\mathbf{u}) \\ &= F^{-1} \circ F(\mathbf{u}, \mathbf{0}) \;=\; (\mathbf{u}, \mathbf{0}) \\ &= (\varphi_1^{-1} \circ \varphi_2(\mathbf{v}), \mathbf{0}). \end{aligned}$$

Daraus folgt, daß $\varphi_1^{-1} \circ \varphi_2$ stetig differenzierbar ist. Umgekehrt zeigt man genauso, daß $\varphi_2^{-1} \circ \varphi_1$ stetig differenzierbar ist. ∎

Definition. Sei M eine Mannigfaltigkeit. Eine stetige Funktion $h : M \to \mathbb{R}$ heißt *differenzierbar*, falls $h \circ \varphi$ für jede Parametrisierung φ differenzierbar ist.

Diese Definition ist nach dem obigen Satz vernünftig. Ist nämlich $f \circ \varphi$ differenzierbar und ψ eine andere Parametrisierung, so ist auch $f \circ \psi = (f \circ \varphi) \circ (\varphi^{-1} \circ \psi)$ differenzierbar.

Sei $B \subset \mathbb{R}^n$ offen, $g : B \to \mathbb{R}^m$ stetig differenzierbar und $\operatorname{rg} J_g(\mathbf{x}) = m$ für alle $x \in B$. Weiter sei $M := \{\mathbf{x} \in B \,:\, g(\mathbf{x}) = \mathbf{0}\}$, $\mathbf{a} \in M$, $U(\mathbf{a}) \subset B$ eine offene Umgebung und $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion.

Definition. f hat in $\mathbf{a}$ ein *relatives Maximum* (bzw. *relatives Minimum*) *unter den Nebenbedingungen*

$$g_1(\mathbf{x}) = \ldots = g_m(\mathbf{x}) = 0,$$

falls $f(\mathbf{x}) \le f(\mathbf{a})$ (bzw. $f(\mathbf{x}) \ge f(\mathbf{a})$) für alle $\mathbf{x} \in U \cap M$ ist.

4.8 Satz (Methode der Lagrangeschen Multiplikatoren).

Hat f in $\mathbf{a}$ ein relatives Extremum unter den Nebenbedingungen

$$g_1(\mathbf{x}) = \ldots = g_m(\mathbf{x}) = 0,$$

so gibt es Zahlen $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, so daß gilt:

$$\nabla f(\mathbf{a}) = \lambda_1 \cdot \nabla g_1(\mathbf{a}) + \cdots + \lambda_m \cdot \nabla g_m(\mathbf{a}).$$

Die Zahlen $\lambda_1, \ldots, \lambda_m$ nennt man Lagrangesche Multiplikatoren. Man beachte, daß es sich hier wieder nur um ein notwendiges Kriterium handelt! Die Punkte, in denen die angegebene Bedingung erfüllt ist, können Extremwerte sein. Ob sie es wirklich sind, muß man mit anderen Mitteln feststellen.

Beweis: Sei M die Mannigfaltigkeit, die durch die Nebenbedingungen gegeben ist. Hat $f|_M$ in $\mathbf{a} \in M$ ein lokales Extremum und ist φ eine Parametrisierung für M in $\mathbf{a}$, so ist $f \circ \varphi$ differenzierbar und hat ein Extremum in $\mathbf{u}_0$ (mit $\varphi(\mathbf{u}_0) = \mathbf{a}$).

Dann ist $0 = D(f \circ \varphi)(\mathbf{u}_0) = Df(\mathbf{a}) \circ D\varphi(\mathbf{u}_0)$. Also liegt $T_{\mathbf{a}}(M) = \operatorname{Im} D\varphi(\mathbf{u}_0)$ im Kern von $Df(\mathbf{a})$. Nun ist

$$T_{\mathbf{a}}(M) = \operatorname{Ker} D(g_1, \ldots, g_m)(\mathbf{a}) = < \nabla g_1(\mathbf{a}), \ldots, \nabla g_m(\mathbf{a}) >^{\perp}$$

und Ker $Df(\mathbf{a}) =< \nabla f(\mathbf{a}) >^\perp$. Daraus folgt:[3]

$$< \nabla f(\mathbf{a}) > \subset < \nabla g_1(\mathbf{a}), \dots, \nabla g_m(\mathbf{a}) >,$$

und das bedeutet, daß es Zahlen $\lambda_1, \dots, \lambda_m$ gibt, so daß gilt:

$$\nabla f(\mathbf{a}) = \sum_{\mu=1}^{m} \lambda_\mu \cdot \nabla g_\mu(\mathbf{a}).$$

■

Beispiele.

1. Sei $g(x, y, z) := 1 - x^2 - y^2 - z$. Dann ist

   $$M := \{(x, y, z) \,:\, g(x, y, z) = 0\} = \{(x, y, z) \,:\, z = 1 - x^2 - y^2\}$$

   ein nach unten geöffnetes Paraboloid. Die „Höhenfunktion“ $h(x, y, z) := z$ hat im $\mathbb{R}^3$ keinerlei Extremwert. Anders ist die Situation, wenn man h auf M einschränkt. Es ist

   $$\nabla g(x, y, z) = (-2x, -2y, -1) \quad \text{und} \quad \nabla h(x, y, z) = (0, 0, 1).$$

   Besitzt h einen Extremwert auf M, so muß es ein $\lambda \in \mathbb{R}$ mit $\lambda x = \lambda y = 0$ und $-\lambda = 1$ geben. Dann folgt, daß $x = y = 0$ ist, und aus der Gleichung $g(x, y, z) = 0$ ergibt sich $z = 1$. Tatsächlich nimmt h auf M in $(x, y, z) = (0, 0, 1)$ sein Maximum an.

2. Sei $M = S^{n-1} = f^{-1}(0)$, für $f(\mathbf{x}) := \|\mathbf{x}\|^2 - 1$. Weiter sei $A \in M_{n,n}(\mathbb{R})$ eine symmetrische Matrix. Wir untersuchen $h(\mathbf{x}) := \mathbf{x} \cdot A \cdot \mathbf{x}^{-1}$ auf Extremwerte unter der Nebenbedingung $f(\mathbf{x}) = 0$. Es ist $\nabla h(\mathbf{x}) = 2\mathbf{x} \cdot A$ und $\nabla f(\mathbf{x}) = 2\mathbf{x}$.

   Ein Extremwert liegt höchstens dann vor, wenn es ein $\lambda \in \mathbb{R}$ mit $\mathbf{x} \cdot A^t = \lambda \cdot \mathbf{x}$ gibt (man beachte $A^t = A$). Da $\|\mathbf{x}\| = 1$ ist, ist $\mathbf{x}$ dann ein Eigenvektor zum Eigenwert λ.

   Weil S^{n-1} kompakt und h stetig ist, muß h auf S^{n-1} ein Maximum annehmen. Damit ist gezeigt: Jede symmetrische Matrix besitzt mindestens einen reellen Eigenwert.

3. Sei $h(\mathbf{x}) := x_1 \cdot x_2 \cdots x_n$ und

   $$M := \{\mathbf{x} \,:\, x_1 + \cdots + x_n = 1 \text{ und } x_\nu > 0 \text{ für alle } \nu\}.$$

   h nimmt auf der kompakten Menge $\overline{M}$ sein Maximum an, und dies muß schon in M liegen (denn h verschwindet auf dem Rand von M). Setzen wir

[3] Ist $U \subset \mathbb{C}^n$ ein Untervektorraum, so ist $U^\perp = \{\mathbf{x} \in \mathbb{R}^n \,:\, \mathbf{x} \bullet \mathbf{y} = 0 \text{ für alle } \mathbf{y} \in U\}$. Wir benutzen das folgende Ergebnis aus der Linearen Algebra: Sind $U, V \subset \mathbb{R}^n$ zwei Untervektorräume mit $U \subset V$, so ist $V^\perp \subset U^\perp$.

$f(\mathbf{x}) := x_1 + \cdots + x_n - 1$, so ist $\nabla f(\mathbf{x}) = (1, 1, \ldots, 1)$ in jedem Punkt $\mathbf{x}$, und es ist

$$\nabla h(\mathbf{x}) = (\prod_{i \neq 1} x_i, \ldots, \prod_{i \neq n} x_i).$$

Wenn h auf M in $\mathbf{a} = (a_1, \ldots, a_n)$ sein Maximum annimmt, dann muß es ein $\lambda \in \mathbb{R}$ geben, so daß $\prod_{i \neq j} a_i = \lambda$ ist, für alle j. Das ist nur möglich, wenn $a_1 = \ldots = a_n$ ist. Da außerdem $a_1 + \cdots + a_n = 1$ ist, folgt:

$$\mathbf{a} = (\frac{1}{n}, \ldots, \frac{1}{n}).$$

Dieses Ergebnis hat eine interessante Konsequenz. Es ist ja

$$h(\mathbf{x}) \leq h(\mathbf{a}) = \left(\frac{1}{n}\right)^n \text{ für alle } \mathbf{x} \in M.$$

Sind nun $t_1, \ldots, t_n$ beliebige positive reelle Zahlen, so gilt:

$$\mathbf{t} := \left(\frac{t_1}{t_1 + \cdots + t_n}, \ldots, \frac{t_n}{t_1 + \cdots + t_n}\right) \in M.$$

Dann ist $\dfrac{t_1 \cdots t_n}{(t_1 + \cdots + t_n)^n} \leq \left(\dfrac{1}{n}\right)^n$, und daher

$$\sqrt[n]{t_1 \cdots t_n} \leq \frac{t_1 + \cdots + t_n}{n}.$$

Das ist die bekannte Ungleichung zwischen dem *geometrischen* und *arithmetischen Mittel.*