2 Der Satz über implizite Funktionen

Im vorangegangenen Abschnitt beschäftigten wir uns mit der Lösbarkeit von nichtlinearen Gleichungssystemen, bei welchen die Anzahl der Gleichungen mit der Anzahl der Variablen übereinstimmte. Im Folgenden betrachten wir nun die Lösbarkeit von solchen Systemen, bei denen mehr Variablen als Gleichungen vorhanden sind. Genauer gesagt betrachten wir m Gleichungen für $m+k$ Unbekannte, d.h.

$$\begin{array}{rcl} f_1(x_1,\ldots,x_k,y_1,\ldots,y_m) & = & 0 \\ \vdots & & \vdots \\ f_m(x_1,\ldots,x_k,y_1,\ldots,y_m) & = & 0 \end{array}$$

und untersuchen die Frage unter welchen Bedingungen sich $y_1,\ldots,y_m$ durch $x_1,\ldots,x_n$ ausdrücken lassen, oder anders formuliert, ob sich das obige Gleichungssystem nach y auflösen lässt.
Betrachtet man den Spezialfall von *linearen Gleichungssystemen* der Form

$$Ax + By = 0$$

mit $A \in M_{mk}(\mathbb{R})$, $B \in M_m(\mathbb{R})$ und $x = (x_1,\ldots,x_k)^T$ und $y = (y_1,\ldots,y_m)^T$, so ist dies möglich, falls B invertierbar ist; es gilt dann

$$y = -B^{-1}Ax.$$

Im Folgenden wollen wir *lokal* ein analoges Resultat für stetig differenzierbare Funktionen $f = (f_1,\ldots,f_m) : U \subset \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$ herleiten. Hierzu definieren wir

$$\begin{array}{lll} a = (a_1,\ldots,a_k), & b = (b_1,\ldots,b_m), & (a,b) := (a_1,\ldots,a_k,b_1,\ldots,b_m), \\ x = (x_1,\ldots,x_k), & y = (y_1,\cdots,y_m), & (x,y) := (x_1,\ldots,x_k,y_1,\ldots,y_m), \end{array}$$

$$f(a,b) := f(a_1,\ldots,a_k,b_1,\ldots,b_m), \quad f(x,y) := f(x_1,\ldots,x_k,y_1,\ldots,y_m),$$

$$D_x f(a,b) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a,b) & \ldots & \frac{\partial f_1}{\partial x_k}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a,b) & \ldots & \frac{\partial f_m}{\partial x_k}(a,b) \end{pmatrix}_{m\times k}$$

und

$$D_y f(a,b) := \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_m}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1}(a,b) & \cdots & \frac{\partial f_m}{\partial y_m}(a,b) \end{pmatrix}_{m\times m}$$

Der folgende Satz über implizite Funktionen bildet den Hauptsatz dieses Abschnitts.

2.1 Theorem. (Satz über implizite Funktionen). *Es seien $U \subset \mathbb{R}^k \times \mathbb{R}^m$ eine offene Menge und $f : U \to \mathbb{R}^m$ eine stetig differenzierbare Funktion. Ferner sei $(a,b) \in U$ mit $f(a,b) = 0$ und*

$$\det(D_y f(a,b)) \neq 0.$$

Dann existiert eine offene Umgebung $W \subset \mathbb{R}^k$ von a und eine eindeutig bestimmte stetig differenzierbare Abbildung $\varphi : W \to \mathbb{R}^m$ mit $\varphi(a) = b$, $(x, \varphi(x)) \in U$ für alle $x \in W$ und

$$f(x, \varphi(x)) = 0 \quad \textit{für alle } x \in W.$$

Ferner gilt

$$D\varphi(x) = -[D_y f(x, \varphi(x))]^{-1} D_x f(x, \varphi(x)), \quad x \in W.$$

Bevor wir den Beweis beginnen, wollen wir uns zunächst die Aussage des Satzes im Spezialfall $m = 1 = k$ veranschaulichen. Hierzu sei $f : \mathbb{R}^2 \to \mathbb{R}$ in Umgebung von $(x_0, y_0) \in \mathbb{R}^2$ stetig differenzierbar und es gelte $f(x_0, y_0) = 0$, sowie

$$D_y f(x_0, y_0) \neq 0.$$

Nach dem Satz über implizite Funktionen existiert dann ein $\delta > 0$ und eine auf dem Intervall $I := (x_0 - \delta, x_0 + \delta)$ eindeutig bestimmte stetig differenzierbare Abbildung φ mit $\varphi(x_0) = y_0$ und

$$f(x, \varphi(x)) = 0 \quad \text{für alle } x \in I.$$

Beweis. Betrachte die Abbildung $F : U \subset \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^k \times \mathbb{R}^m$ definiert durch

$$F(x,y) := (x, f(x,y)).$$

Dann ist F stetig differenzierbar und es gilt

$$\text{(2.1)} \quad DF(x,y)(h,k) = (h, D_x f(x,y) \cdot h + D_y f(x,y) \cdot k), \quad (x,y) \in U, (h,k) \in \mathbb{R}^k \times \mathbb{R}^m.$$

Weiter ist $DF(a,b) \in \mathcal{L}(\mathbb{R}^k \times \mathbb{R}^m)$ invertierbar, denn (2.1) impliziert

$$DF(a,b)(h,k) = 0 \Longrightarrow h = 0 \Longrightarrow \underbrace{D_y f(a,b) k}_{\text{invertierbar n. Vor.}} = 0 \Longrightarrow k = 0.$$

Somit ist $DF(a,b)$ injektiv, also auch bijektiv und daher invertierbar.
Wir können also den Satz 1.3 über die Umkehrabbildung auf F in (a,b) anwenden und

erhalten eine offene Umgebung $\tilde{U} \subset U$ von (a,b) mit der Eigenschaft, dass $\tilde{F} := F|_{\tilde{U}} : \tilde{U} \to F(\tilde{U}) =: V$ eine stetig differenzierbare Umkehrabbildung $G : V \to \tilde{U}$ besitzt.
Da F in den ersten Koordinaten wie die Identität wirkt, gilt dies auch für G, d.h. es existiert eine stetig differenzierbare Abbildung $g : V \to \mathbb{R}^m$ mit

$$G(\xi,\eta) = (\xi, g(\xi,\eta)) \quad \text{für alle } (\xi,\eta) \in V.$$

Für $(x,y) \in \tilde{U}$ gilt somit

$$\begin{aligned} f(x,y) = 0 \quad &\Longleftrightarrow \quad F(x,y) = (x,0) \Longleftrightarrow (x,y) = G(x,0) = (x, g(x,0)) \\ &\Longleftrightarrow \quad y = g(x,0) \end{aligned} \tag{2.2}$$

und insbesondere $b = g(a,0)$. Setzt man

$$W := \{x \in \mathbb{R}^k : \begin{pmatrix} x \\ 0_{\mathbb{R}^m} \end{pmatrix} \in V\},$$

so ist mit V auch W offen und

$$V \ni F(a,b) = (a, f(a,b)) = (a,0),$$

d.h. W ist eine Umgebung von a. Wir definieren schließlich die Funktion

$$\varphi : W \to \mathbb{R}^m, \ \varphi(x) := g(x,0),$$

welche wegen (2.2) die behauptete lokale Auflösung der Gleichung $f(x,y) = 0$ nach y in einer Umgebung des Punkts (a,b) liefert.
Die Ableitung der Funktion φ lässt sich via der Kettenregel wie folgt bestimmen: aus $f(x,\varphi(x)) = 0$ folgt

$$D_x f(x,\varphi(x)) \cdot \mathrm{Id}_{\mathbb{R}^k} + D_y f(x,\varphi(x)) \cdot D\varphi(x) = 0.$$

Für $(x,y) = (a,b)$ gilt wegen $\varphi(a) = b$

$$D\varphi(a) = -[D_y f(a,b)]^{-1} \cdot D_x f(a,b).$$

□

Wir bemerken an dieser Stelle noch, dass sich die Ableitung $D\varphi(a)$ von φ wie in Satz 2.1 angegeben *ohne* explizite Kenntnis von φ bestimmen lässt!

2.2 Beispiele.
a) *Höhenlinien*: Es sei $U \subset \mathbb{R}^2$ eine offene Menge, $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion und $c \in \mathbb{R}$. Dann bezeichnet

$$N_f(c) = \{(x,y) \in U : f(x,y) = c\}$$

die *Niveaumenge* von f. Die Niveaumenge wird desöfteren auch als *Höhenlinie* bezeichnet, wobei im Allgemeinen die Niveaumenge keine „Linie“ sein muss. Der Satz über implizite Funktionen im Fall $m = k = 1$ angewandt auf die Funktion

$$f_c : U \to \mathbb{R}, \ f_c(x, y) = f(x, y) - c$$

liefert folgendes Resultat. Ist $(a, b) \in U$ mit $f(a, b) = c$ und gilt

$$\operatorname{grad} f_c(a, b) = (f_x(a, b), f_y(a, b)) \neq (0, 0), \tag{2.3}$$

so kann man die Gleichung $f(x, y) = c$

1. in einer Umgebung von a nach y auflösen, falls $f_y(a, b) \neq 0$ gilt und
2. in Umgebung von b nach x auflösen, falls $f_x(a, b) \neq 0$ gilt.

Mit anderen Worten bedeutet dies, dass Niveaumengen durch Punkte (a, b), die (2.3) lokal erfüllen, sich als stetig differenzierbare Abbildungen der Form $x \mapsto (x, \varphi(x))$ im Fall 1. bzw. $y \mapsto (\psi(y), y)$ im Fall 2. darstellen lassen. Die Niveaumenge wird in diesem Fall tatsächlich durch *Höhenlinien* beschrieben.
b) *Gleichungssysteme*
Für $k = 1$ und $m = 2$ betrachten wir das Gleichungsystem

$$\begin{array}{lll} f_1(x, y_1, y_2) & = x^3 + y_1^3 + y_2^3 - 7 & = 0 \\ f_2(x, y_1, y_2) & = xy_1 + y_1y_2 + xy_2 + 2 & = 0 \end{array}$$

mit $f_1(2, -1, 0) = f_2(2, -1, 0) = 0$. Für $f = (f_1, f_2)$ gilt

$$D_y f(2, -1, 0) = \begin{pmatrix} 3y_1^2 & 3y_2^2 \\ x + y_2 & x + y_1 \end{pmatrix}_{|(2,-1,0)} = \begin{pmatrix} 3 & o \\ 2 & 1 \end{pmatrix}.$$

Da diese Matrix invertierbar ist, existieren in einer Umgebung W von $a = 2$ zwei stetig differenzierbare Funktionen $\varphi_1, \varphi_2 : W \to \mathbb{R}$ mit

$$(\varphi_1(2), \varphi_2(2)) = (-1, 0)$$

und

$$f_1(x, \varphi_1(x), \varphi_2(x)) = 0 \quad \text{sowie} \quad f_2(x, \varphi_1(x), \varphi_2(x)) = 0.$$

Für die Ableitungen von φ_1 und φ_2 gilt

$$\begin{pmatrix} \varphi_1'(2) \\ \varphi_2'(2) \end{pmatrix} = - \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x}(2, -1, 0) \\ \frac{\partial f_2}{\partial x}(2, -1, 0) \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & 0 \\ \frac{2}{3} & -1 \end{pmatrix} \begin{pmatrix} 3x^2 \\ y_1 + y_2 \end{pmatrix}_{|(2,-1,0)} = \begin{pmatrix} -4 \\ 9 \end{pmatrix}$$

3 Untermannigfaltigkeiten und Extremwerte unter Nebenbedingungen

Der obige Satz über implizite Funktionen führt in geometrischer Sicht zum Begriff der differenzierbaren Mannigfaltigkeit. Letzterer spielt in der modernen Mathematik eine wichtige Rolle und wir wollen hier erste Eigenschaften aufzeigen.

3.1 Definition. Eine Teilmenge M des $\mathbb{R}^n$ heißt d-dimensionale *differenzierbare Untermannigfaltigkeit* des $\mathbb{R}^n$, wenn es zu jedem $a \in M$ eine in $\mathbb{R}^n$ offene Umgebung U von a und einen Diffeomorphismus $\varphi : U \to V$ auf eine offene Teilmenge V des $\mathbb{R}^n$ gibt mit

$$\varphi(U \cap M) = V \cap (\mathbb{R}^d \times \{0\}).$$

Ein- bzw. zweidimensionale Untermannigfaltigkeiten des $\mathbb{R}^n$ heißen auch in $\mathbb{R}^n$ *eingebettete Kurven* bzw. *Flächen.* Differenzierbare Untermannigfaltigkeiten des $\mathbb{R}^n$ der Dimension $n-1$ heißen auch *Hyperflächen.* . Im Folgenden verstehen wir unter einer (Unter)Mannigfaltigkeit eine differenzierbare Untermannigfaltigkeit des $\mathbb{R}^n$.
Beispiele von Mannigfaltigkeiten lassen sich relativ einfach, wie der folgende Satz zeigt, durch Graphen von differenzierbaren Funktionen erhalten.

3.2 Satz. *Ist O offen in $\mathbb{R}^d$ und $f : O \to \mathbb{R}^n$ eine stetig differenzierbare Funktion, so ist graph f eine d-dimensionale Untermannigfaltigkeit des $\mathbb{R}^{d+n}$.*

Beweis. Setzt man $U := O \times \mathbb{R}^n$ und

$$\varphi : U \to \mathbb{R}^{d+n} = \mathbb{R}^d \times \mathbb{R}^n,\ (x,y) \mapsto (x, y - f(x)),$$

so ist φ stetig differenzierbar mit $\mathrm{Im}(\varphi) = U$. Ferner ist $\varphi : U \to U$ bijektiv mit $\varphi^{-1}(x,z) = (x, z + f(x))$, also ein Diffeomorphismus von U auf sich, und es gilt

$$\varphi(U \cap \text{graph }(f)) = O \times \{0\} = U \cap (\mathbb{R}^d \times \{0\}).$$

□

Der folgende Satz vom regulären Wert stellt eines der nützlichsten Mannigfaltigkeitskriterien dar. Wir erinnern zunächst an den folgenden Satz aus der Linearen Algebra: Ist $f : \mathbb{R}^n \to \mathbb{R}^m$ eine reguläre, d.h. surjektive und lineare Abbildung, so ist für jedes $c \in \mathbb{R}^m$ der Lösungsraum der Gleichung $f(x) = c$ ein affiner Unterraum des $\mathbb{R}^n$ der Dimension $n - m$. Im Folgenden verallgemeinern wir dieses Ergebnis auf den Fall einer Gleichung $f(x) = c$ für eine stetig differenzierbare Funktion f und einem sogenannten regulären Wert c.

3.3 Definition. Ein Punkt $x \in U \subset \mathbb{R}^d$ heißt *regulärer Punkt* der stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, falls die Ableitung $Df(x) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ surjektiv ist. Ferner heißt ein Punkt $y \in \mathbb{R}^n$ *regulärer Wert* von f, wenn alle $x \in f^{-1}(y)$ reguläre Punkte von f sind.

3.4 Bemerkungen.
a) Gilt $d < n$, so besitzt f keine regulären Punkte.
b) Gilt $d \geq n$, so ist $x \in U$ genau dann ein regulärer Punkt von f, wenn die Ableitung $Df(x)$ den Rang n besitzt.
c) Gilt $n = 1$, so ist $x \in U$ genau dann ein regulärer Punkt von f, wenn $\nabla f(x) \neq 0$ ist.
d) Ein Punkt $y \in \mathbb{R}^n$ ist genau dann ein regulärer Wert von f, wenn die Matrix $Df(x)$ in allen Punkten $x \in f^{-1}(y)$ den Rang n hat. Gilt $n = 1$, so bedeutet dies, dass $f'(x) \neq 0$ ist in allen solchen Punkten.
e) Ist $a \in U$ ein regulärer Punkt einer stetig differenzierbaren Funktion f, so existieren nach dem Satz über implizite Funktionen (nach einer möglichen orthogonalen Transformation des $\mathbb{R}^n$) n Variablen, so dass das Gleichungssystem

$$\begin{aligned} f_1(x_1, \ldots, x_d) &= 0 \\ \vdots \qquad &\ \ \vdots \\ f_n(x_1, \ldots, x_d) &= 0 \end{aligned}$$

in einer Umgebung von a eindeutig nach diesen Variablen als Funktion der übrigen $d - n$ Variablen aufgelöst werden kann.
f) Ist $0 \in \mathrm{Im}(f)$ ein regulärer Wert einer stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, so existiert zu jedem $a \in f^{-1}(0)$ eine Umgebung V in $\mathbb{R}^d$, so dass sich $f^{-1}(0) \cap V$ als Graph einer stetig differenzierbaren Funktion in $d - n$ Variablen darstellen lässt.

Wir kommen nun zum angekündigten Satz vom regulären Wert.

3.5 Theorem. *Ist $U \subset \mathbb{R}^d$ eine offene Menge und c ein regulärer Wert einer stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, so ist $f^{-1}(c)$ eine $(d-n)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$.*

Nach den obigen Vorbereitungen ist der Beweis einfach und folgt direkt aus Satz 3.2 und Bemerkung 3.4 e).

3.6 Beispiele.
a) Die euklidische $(n-1)-$*Sphäre*

$$S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$$

ist eine $(n-1)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Um dies einzusehen, betrachten wir die stetig differenzierbare Funktion

$$f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \|x\|_2^2.$$

Da $\nabla f(x) = 2x$ für alle $x \in \mathbb{R}^n$ gilt, ist 1 ein regulärer Wert von f. Weiter, da $S = f^{-1}(1)$ folgt die Behauptung aus dem Satz vom regulären Wert.
b) Betrachte auf $U = \mathbb{R}^3 \setminus (\{0\} \times \{0\} \times \mathbb{R})$ die stetig differenzierbare Funktion

$$f : U \to \mathbb{R},\ f(x_1, x_2, x_3) := \left(\sqrt{x_1^2 + x_2^2} - 2\right)^2 + x_3^2.$$

Da 1 ein regulärer Wert von f ist, ist $T^2 := f^{-1}(1)$ eine zweidimensionale Untermannigfaltigkeit des $\mathbb{R}^3$; T^2 ist die zweidimensionale *Torusfläche*, welche durch Rotation des in $(x_1 - x_3)$-Ebene liegenden Kreises $(x_1 - 2)^2 + x_3^2 = 1$ um die x_3-Achse entsteht.
c) Ist A eine reelle, symmetrische $n \times n$-Matrix mit $\det A \neq 0$, so ist die *Quadrik*

$$Q := \{x \in \mathbb{R}^n : (x^T | Ax) = 1\}$$

eine $(n-1)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Für den Beweis wählen wir $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = (x^T | Ax)$ und notieren, dass $Q = f^{-1}(1)$ gilt. Da $Df(x) = 2x^T A \neq 0$ für alle $x \in \mathbb{R}^n$ mit $x \neq 0$ gilt, ist 1 ein regulärer Wert von f und die Behauptung folgt wiederum aus dem Satz vom regulären Wert.

Will man die Konzepte der Differentialrechnung auf Abbildungen zwischen Untermannigfaltigkeiten übertragen, so erweist es sich als sehr nützlich, lineare Strukturen wie den Tangential- und Normalenraum an Mannigfaltigkeiten im euklidischen $\mathbb{R}^n$ einzuführen.
Wir betrachten zunächst den Begriff des Tangentialraums.

3.7 Definition. Es sei M eine Teilmenge des $\mathbb{R}^n$. Ein Vektor $v \in \mathbb{R}^n$ heißt *Tangentialvektor an M im Punkt $a \in M$*, wenn es in M eine stetig differenzierbare Kurve $\gamma : (-\varepsilon, \varepsilon) \to M$, $\varepsilon > 0$, mit $\gamma(0) = a$ und $\gamma'(0) = v$ gibt. Die Menge aller Tangentialvektoren an M in a heißt *Tangentialkegel* von M in a und wird mit $T_a M$ bezeichnet. Ist $T_a M$ ein Vektorraum, so wird er auch Tangential*raum* genannt.

3.8 Satz. *Es sei M eine $d-$dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Dann gilt für jedes $a \in M$:*
a) $T_a M$ ist ein Vektorraum der Dimension d.
b) Gilt $M = f^{-1}(c)$ für eine stetig differenzierbare Funktion $f : U \to \mathbb{R}^m$ auf einer offenen Menge $U \subset \mathbb{R}^n$ zu einem regulären Wert c von f, so ist

$$T_a M = \operatorname{kern} Df(a) = \{v \in \mathbb{R}^n : Df(a)v = 0\}.$$

Beweis. a) Die Behauptung ist leicht einzusehen für die d-dimensionale Untermannigfaltigkeit $M = V \cap (\mathbb{R}^d \times \{0\})$, wobei V offen in $\mathbb{R}^n$ ist. In diesem Fall ist

$$T_a(V \cap (\mathbb{R}^d \times \{0\}) = \mathbb{R}^d \times \{0\}.$$

Für den allgemeinen Fall, betrachten wir die Abbildung $\varphi : U \to V$ aus der Definition 3.1 der Untermannigfaltigkeit M, welche jeder Kurve $\gamma : (-\varepsilon, \varepsilon) \to M \cap U$ die Bildkurve $\tilde{\gamma} := \varphi \circ \gamma$ in $V \cap (\mathbb{R}^d \times \{0\})$ zuordnet. Jede Kurve in $V \cap (\mathbb{R}^d \times \{0\})$ ist eine solche Bildkurve und für zwei Kurven γ und $\tilde{\gamma}$ gilt

$$\gamma'(t) = [D\varphi(a)]^{-1} \tilde{\gamma}'(t).$$

Daher folgt $T_a(M \cap U) = [D\varphi(a)]^{-1}(\mathbb{R}^d \times \{0\})$ und wegen $T_a M = T_a(M \cap U)$ folgt die Behauptung a).
b) Für $\gamma : (\varepsilon, \varepsilon) \to M$ gilt $f \circ \gamma = c$ und somit $Df(a)\gamma'(0) = 0$. Daher ist

$$T_a M \subset \text{kern } Df(a).$$

Da $Df(a) : \mathbb{R}^d \to \mathbb{R}^n$ nach Voraussetzung surjektiv ist, gilt nach der Dimensionsformel aus der Linearen Algebra dim kern $Df(a) = n - m =$ dim $M = d$. Daher ist $T_a M$ kein echter Untervektorraum von kern $Df(a)$ und die Behauptung ist bewiesen.

□

Unter einem *Normalenvektor* einer Untermannigfaltigkeit $M \subset \mathbb{R}^n$ im Punkt $a \in M$ versteht man einen Vektor $v \in \mathbb{R}^n$ der senkrecht zum Tangentialraum $T_a M$ steht; der *Normalenraum* $N_a M$ ist das orthogonale Komplement zu $T_a M$, d.h.

$$N_a M := (T_a M)^\perp.$$

Der obige Satz über den Tangentialraum ergibt folgendes Korollar.

3.9 Korollar. *Es sei $M = f^{-1}(c)$ die Niveaumenge einer stetig differenzieraren Funktion $f = (f_1, \ldots, f_{n-d}) : U \to \mathbb{R}^{n-d}$ zum regulären Wert $c \in \mathbb{R}^{n-d}$. Dann bilden die Gradienten $\nabla f_1(a), \ldots, \nabla f_{n-d}(a)$ in $a \in M$ eine Basis des Normalenraums, d.h. es gilt*

$$N_a M = [\nabla f_1(a), \ldots, \nabla f_{n-d}(a)].$$

Beweis. Die Zeilen der Matrix $Df(a)$ werden durch den obigen Gradienten beschrieben. Nach Satz 3.8 liegt $v \in \mathbb{R}^n$ genau dann in $T_a M$, wenn $(\nabla f_i(a)|v) = 0$ für $i = 1, \ldots, n-d$ gilt; daher stehen die Vektoren $\nabla f_i(a)$ für $i = 1, \ldots, n-d$ senkrecht auf $T_a M$. Die Matrix $Df(a)$ hat nach Voraussetzung den Rang $n - d$ und somit sind die Vektoren $\nabla f_i(a)$ für $i = 1, \ldots, n-d$ linear unabhängig und bilden eine Basis für $N_a M$.

□

3.10 Beispiele.
a) Für die Sphäre S^{n-1} in $\mathbb{R}^n$ gilt $S^{n-1} = f^{-1}(1)$ mit $f : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto \|x\|_2$. Wegen $\nabla f(a) = 2a$ gilt

$$N_a S^{n-1} = \{\lambda a : \lambda \in \mathbb{R}\}$$

für $a \in S^{n-1}$.
b) Ein Normalenvektor an die Torusfläche T^2 beschrieben in Beispiel 3.6 b) im Punkt (a_1, a_2, a_3) ist gegeben durch $2(a - h)$ mit

$$h = (\frac{2a_1}{(a_1^2 + a_2^2)^{1/2}}, \frac{2a_2}{(a_1^2 + a_2^2)^{1/2}}, 0).$$

c) Ist $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion, so ist die *Einheitsnormale* $\nu(x)$, d.h. ein Normalenvektor der Länge 1, an die Mannigfaltigkeit $M = \text{graph} f$ im Punkt $a = (x, f(x))$ gegeben durch

$$\nu(x) = \frac{(-\nabla f(x), 1)}{(1 + |\nabla f(x)|^2)^{1/2}}.$$

In vielen Anwendungen ist nicht nur das bloße Extremum einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ gesucht, sondern es sind Extremwerte unter Nebenbedingungen zu bestimmen. Genauer seien $f : U \subset \mathbb{R}^n \to \mathbb{R}$ und $g = (g_1, \ldots, g_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$ gegebene Funktionen und $M := \{x \in U : g(x) = 0\}$. Gesucht sind Punkte $x_0 \in M$ mit

$$f(x_0) \leq f(x) \quad \text{für alle } x \in M \cap U_{x_0}$$

oder

$$f(x_0) \geq f(x) \quad \text{für alle } x \in M \cap U_{x_0}$$

für eine Umgebung $U_{x_0} \subset U$ von x_0. Ein solcher Punkt heißt *lokales Extremum* von f unter der *Nebenbedingung* $g = 0$.
Die folgende Multiplikatorregel von Lagrange beschreibt ein notwendiges Kriterium für das Vorliegen von solchen Extremwerten unter Nebenbedingungen.

3.11 Satz. (Multiplikatorregel von Lagrange). *Es seien $U \subset \mathbb{R}^n$ offen, $m < n$, $f : U \to \mathbb{R}$ sowie $g = (g_1, \ldots, g_m) : U \to \mathbb{R}^m$ stetig differenzierbare Funktionen. Ferner sei 0 ein regulärer Wert von g und es gelte $M = g^{-1}(0) \neq \emptyset$. Besitzt f in x_0 unter der Nebenbedingung $g = 0$ ein lokales Extremum, so existieren $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ mit*

$$Df(x_0) = \sum_{j=1}^{m} \lambda_j Dg_j(x_0).$$

Beweis. Nach dem Satz vom regulären Wert ist M eine $(n - m)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$ und es gibt es zu $v \in T_{x_0}M$ eine stetig differenzierbare

Kurve $\gamma : (-\varepsilon, \varepsilon) \to M$ mit $\gamma(0) = x_0$ und $\gamma'(0) = v$. Die durch $F : (-\varepsilon, \varepsilon) \to \mathbb{R}$, $F(t) := f(\gamma(t))$ definierte Funktion hat in $t = 0$ ein lokales Extremum. Daher ist $F'(0) = 0$, welches $(\mathrm{grad} f(x_0)|v) = 0$ und somit $\mathrm{grad} f(x_0) \in N_{x_0}M$ bedeutet. Nach Satz 3.9 gibt es daher eindeutig bestimmte $\lambda_1, \ldots, \lambda_m$ mit der behaupteten Eigenschaft.

$\square$

3.12 Bemerkungen.
a) Die Zahlen $\lambda_1, \ldots, \lambda_m$ heißen *Lagrange-Multiplikatoren.*
b) Im speziellen Fall $m = 1$ gilt: seien $U \subset \mathbb{R}^n$ offen, $f, g : U \to \mathbb{R}$ stetig differenzierbare Funktionen. Ist $Dg(x_0) \neq 0$ und besitzt f in x_0 ein lokales Extremum unter der Nebenbedingung $g = 0$, so existiert ein $\lambda \in \mathbb{R}$ mit

$$Df(x_0) = \lambda \cdot Dg(x_0).$$

Um dies einzusehen, nummerieren wir die Koordinaten so, dass

$$Dg(x_0) = \underbrace{(\xi_1, \ldots, \xi_{n-1}, \eta)}_{D_x g(x_0)} \qquad \text{mit } \eta \neq 0$$

gilt und wenden Satz 3.11 an.
c) Für $f, g : U \subset \mathbb{R}^2 \to \mathbb{R}$ und $c \in \mathbb{R}$ setze $N_c := f^{-1}(c)$ und $M := g^{-1}(0)$. Dann gilt $T_{x_0}N = T_{x_0}M$ und besitzt f in x_0 ein Minimum unter der Nebenbedingung $g = 0$ mit $f(x_0) = c$, so ergibt sich das folgende Bild.

In einem Schnittpunkt x von N_c und M kann kein Extremwert von $f_{|M}$ liegen, da durch Verschieben von x längs M sowohl kleinere als auch größere Funktionswerte von f auf M erreicht werden können.

3.13 Beispiel.
Es sei $A \in M_n(\mathbb{R})$ eine symmetrische Matrix,

$$f : \mathbb{R}^n \to \mathbb{R}, \; f(x) := (x|Ax)$$

und $S := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ die $(n-1)$-Sphäre. Da S kompakt und f stetig aud S ist, nimmt f ein Maximum auf S an. Wir bestimmen die Maximalstelle x_0 mit Hilfe der Multiplikatorenregel von Lagrange. Hierzu beobachten wir, dass $S = g^{-1}(0)$ gilt für

$$g : \mathbb{R}^n \to \mathbb{R}, \; g(x) = (x|x) - 1.$$

Wie müssen daher f unter der Nebenbedingung g maximieren. Nach Beispiel 1.3 sind die Funktionen f und g stetig differenzierbar und es gilt

$$Df(x_0) = 2Ax_0, \; Dg(x_0) = 2x_0 \neq 0.$$

Nach der Lagrangeschen Multiplikatorenregel existiert ein $\lambda \in \mathbb{R}$ mit

$$Ax_0 = \lambda x_0,$$

d.h. x_0 ist ein Eigenvektor der Matrix A. Da $\|x_0\| = 1$ gilt, folgt

$$(x_0|Ax_0) = (x_0|\lambda x_0) = \lambda.$$

Wir haben damit den folgenden Satz aus der Linearen Algebra bewiesen.

3.14 Satz. *Ist $A \in M_n(\mathbb{R})$ eine symmetrische Matrix, so ist*

$$M := \max_{\|x\|=1} (x|Ax) \in \mathbb{R}$$

ein Eigenwert von A und jedes $x_0 \in \mathbb{R}$ mit $\|x_0\| = 1$ und $(x_0|Ax_0) = M$ ist ein zu M gehöriger Eigenvektor.

Als weitere Anwendung beweisen wir den den folgenden Spektralsatz für symmetrische Matrizen aus der Linearen Algebra.

3.15 Satz. (Spektralsatz für symmetrische Matrizen). *Ist $A \in M_n(\mathbb{R})$ eine symmetrische Matrix, so existieren*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n, \quad \lambda_j \in \mathbb{R} \quad \textit{und} \quad x_1, \ldots, x_n \in S^{n-1}$$

mit $Ax_j = \lambda x_j$ für alle $j = 1, \ldots, n$. Ferner bilden die Vektoren $x_1, \ldots, x_n$ eine Orthonormalbasis des $\mathbb{R}^n$ und bezüglich dieser Basis besitzt A Diagonalgestalt.

Beweis. Nach dem Satz 3.14 existieren $x_1 \in S$ und $\lambda_1 \in \mathbb{R}$ mit $Ax_1 = \lambda_1 x_1$. Wir konstruieren einen weiteren Vektor $x_2 \in S$ wie folgt: definiert man $g := (g_0, g_1)$ durch

$$g_0(x) := \|x\|^2 - 1, \quad g_1(x) = 2(x_1|x),$$

so ist $g^{-1}(0) = S \cap \{x_1\}^\perp =: K$ eine kompakte Menge. Für die Funktion $f : \mathbb{R}^n \to \mathbb{R}, x \mapsto (x|Ax)$ existiert nach dem Satz vom Maximum III.3.9 ein $x_2 \in K$ mit $f(x) \leq f(x_2)$ für alle $x \in K$. Ferner ist $Dg(x)$ invertierbar für alle $x \in K$. Nach der Multiplikatorenregel von Lagrange, Satz 3.11, existieren daher $\mu_0, \mu_1 \in \mathbb{R}$ mit

$$(3.1) \qquad \nabla f(x_2) = 2Ax_2 \overset{\text{Lagrange}}{=} \mu_0 \nabla g_0(x_2) + \mu_1 \nabla g_1(x_2) = \mu_0 2x_2 + \mu_1 2x_1.$$

Aufgrund von Satz 3.14 und da nach Konstruktion $(x_2|x_1) = 0$ gilt, folgt

$$(Ax_2|x_1) = (x_2|Ax_1) \stackrel{3.14}{=} (x_2|\lambda x_1) = \lambda(x_2|x_1) = 0.$$

Zusammen mit Gleichung (3.1) liefert dies

$$0 = (Ax_2|x_1) = \mu_0 \underbrace{(x_2|x_1)}_{=0} + \mu_1 \underbrace{(x_1|x_1)}_{=1} = \mu_1.$$

Gleichung (3.1) impliziert daher $Ax_2 = \mu_0 x_2$, welches bedeutet, dass μ_0 ein Eigenwert von A zum Eigenvektor x_2 ist. Der Wert von μ_0 berechnet sich schließlich zu

$$\mu_0 = \mu_0(x_2|x_2) = (Ax_2|x_2) = f(x_2).$$

Iterieren wir dieses Verfahren, so folgt die Behauptung.

□