Speziell für $k = 2$ ist also

$$f(X_0 + H) = f(X_0) + \langle \nabla f(X_0), H \rangle + \frac{1}{2} \sum_{i,j=1}^{n} \partial_i \partial_j f(X_0) h_i h_j + R(X_0; H)$$

mit

$$R(X_0; H) = \frac{1}{6} \sum_{i,j,m=1}^{n} \partial_i \partial_j \partial_m f(Y) h_i h_j h_m$$

und passendem $Y \in [X_0, X_0 + H]$. Wegen $|h_i| \leq \|H\|$ folgt

$$\lim_{H \to 0} \frac{R(X_0; H)}{\|H\|^2} = 0.$$

Analog zu Analysis I, Satz (18.3) (und daraus herleitbar) gilt diese Schlußfolgerung auch schon unter der Voraussetzung, daß $f \in C^1(M)$ ist und die partiellen Ableitungen von f in X_0 differenzierbar sind.

Man kann den durch

$$f(X_0 + H) = f(X_0) + \langle \nabla f(X_0), H \rangle + \frac{1}{2} \sum_{i,j=1}^{n} \partial_i \partial_j f(X_0) h_i h_j + R(X_0; H),$$

$$\lim_{H \to 0} \frac{R(X_0; H)}{\|H\|^2} = 0$$

ausgedrückten Sachverhalt auch folgendermaßen beschreiben: In zweiter Näherung verhält sich die Funktion f an der Stelle X_0 wie ein Polynom zweiten Grades. Anschaulich läßt sich folgendes sagen. Betrachten wir den Graphen von f, so gibt der erste Term $f(X_0)$ der Taylorformel die Höhe des Graphen über X_0 an, der zweite Term, der durch den Gradienten von f in X_0 bestimmt ist, legt die Tangentialebene an den Graphen über X_0 fest, und der nächste Term,

$$\frac{1}{2} \sum_{i,j=1}^{n} \partial_i \partial_j f(X_0) h_i h_j,$$

gibt eine erste Information über die Gestalt des Graphen bei X_0. Dies zeigt sich insbesondere bei der Untersuchung lokaler Extremwerte. Zuvor wollen wir für diesen Term eine besondere Bezeichnung einführen.

(10.3) **Definition.** Sei $f : M \to \mathbb{R}$ zweimal partiell differenzierbar in X_0. Dann heißt die durch

$$Q(f, X_0; H) := \sum_{i,j=1}^{n} \partial_i \partial_j f(X_0) h_i h_j \qquad \text{für } H = (h_1, \ldots, h_n) \in \mathbb{R}^n$$

definierte Funktion $Q(f, X_0; \cdot)$ die *Hesse-Form* von f in X_0, und die Matrix

$$\mathrm{Hess}(f)_{X_0} := \begin{pmatrix} \partial_1\partial_1 f(X_0) & \cdots & \partial_1\partial_n f(X_0) \\ \vdots & & \vdots \\ \partial_n\partial_1 f(X_0) & \cdots & \partial_n\partial_n f(X_0) \end{pmatrix}$$

heißt die *Hessesche Matrix* von f in X_0.

Im folgenden werden wir stets voraussetzen, daß f in einer Umgebung von X_0 zweimal stetig differenzierbar ist. Nach (10.1) ist dann die Hessesche Matrix von f symmetrisch.

Die Hesse-Form ist also eine quadratische Form, die Hessesche Matrix ist die ihr (bezüglich der Standardbasis) zugeordnete Matrix. Über quadratische Formen, deren Behandlung in der Linearen Algebra erfolgt, wollen wir hier nur das für die folgenden Betrachtungen Notwendige bereitstellen.

(10.4) **Definition.** Eine Funktion $q : \mathbb{R}^n \to \mathbb{R}$ mit

$$q(X) = \sum_{i,j=1}^{n} a_{ij} x_i x_j \qquad \text{für } X = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

($a_{ij} \in \mathbb{R}$ für $i, j = 1, \ldots, n$) heißt *quadratische Form* über $\mathbb{R}^n$. Die quadratische Form q heißt *positiv definit*, wenn $q(X) > 0$ für $X \in \mathbb{R}^n \setminus \{0\}$ gilt, *positiv semidefinit*, wenn $q(X) \geq 0$ für $X \in \mathbb{R}^n$ gilt, *negativ definit* (*negativ semidefinit*), wenn $-q$ positiv definit (bzw. positiv semidefinit) ist, und *indefinit*, wenn q weder positiv noch negativ semidefinit ist.

Nach diesen Vorbereitungen können wir nun lokale Extremwerte differenzierbarer Funktionen untersuchen.

(10.5) **Definition.** Die Funktion $f : M \to \mathbb{R}$ *hat in* X_0 *ein lokales Maximum (lokales Minimum)*, wenn es eine Umgebung $U \subset M$ von X_0 gibt mit $f(X_0) \geq f(Y)$ (bzw. $f(X_0) \leq f(Y)$) für alle $Y \in U$. Gilt $f(X_0) > f(Y)$ (bzw. $f(X_0) < f(Y)$) für $Y \in U \setminus \{X_0\}$, so hat f in X_0 ein *strenges lokales Maximum (strenges lokales Minimum)*. Die Funktion f hat in X_0 ein lokales Extremum, wenn sie dort ein lokales Maximum oder ein lokales Minimum hat.

Zunächst stellen wir, analog wie in Analysis I, für das Vorliegen eines lokalen Extremums eine notwendige Bedingung auf.

(10.6) **Satz.** *Sei $M \subset \mathbb{R}^n$ offen und $f : M \to \mathbb{R}$ differenzierbar in X_0. Hat f in X_0 ein lokales Extremum, so ist $\nabla f(X_0) = 0$.*

Beweis. Hat f in X_0 ein lokales Extremum, so hat für jeden Vektor $H \in \mathbb{R}^n$ die durch

$$g(t) := f(X_0 + tH) \qquad \text{für } X_0 + tH \in M$$

erklärte Funktion g in 0 ein lokales Extremum. Nach dem Kriterium aus Analysis I (Satz (18.4)) gilt daher

$$\langle \nabla f(X_0), H \rangle = g'(0) = 0.$$

Da $H \in \mathbb{R}^n$ beliebig war, ist $\nabla f(X_0) = 0$. ■

(10.7) **Definition.** Ist f differenzierbar in einer Umgebung von X_0 und ist $\nabla f(X_0) = 0$, so heißt X_0 ein *kritischer Punkt* von f.

Wir wollen nun hinreichende Bedingungen angeben für das Vorliegen eines strengen lokalen Extremums.

(10.8) **Satz.** *Sei $M \subset \mathbb{R}^n$ offen und $f \in C^2(M)$. Sei X_0 kritischer Punkt von f. Ist die Hesse-Form $Q(f, X_0; \cdot)$ von f in X_0 positiv definit (negativ definit), so hat f in X_0 ein strenges lokales Minimum (strenges lokales Maximum).*

Beweis. Sei etwa $Q(f, X_0; \cdot)$ positiv definit. Wir zeigen zunächst die Existenz einer Zahl $\delta \in \mathbb{R}^+$ mit $U(X_0, \delta) \subset M$ und der Eigenschaft, daß für alle $H \in \mathbb{R}^n$ mit $\|H\| < \delta$ auch die Hesse-Form $Q(f, X_0 + H; \cdot)$ positiv definit ist.

Da die Funktion $Q(f, X_0; \cdot)$ stetig ist, nimmt ihre Einschränkung auf die wegen (6.13) kompakte Menge $\{E \in \mathbb{R}^n \mid \|E\| = 1\}$ nach (5.5) ein Minimum an. Da $Q(f, X_0; \cdot)$ positiv definit ist, ist dieses Minimum positiv. Es gibt also eine Zahl $a \in \mathbb{R}^+$ mit

$$Q(f, X_0; E) > 2a \quad \text{für alle } E \in \mathbb{R}^n \text{ mit } \|E\| = 1.$$

Da nach Voraussetzung die partiellen Ableitungen zweiter Ordnung der Funktion f in X_0 stetig sind (und M offen ist), gibt es ein $\delta \in \mathbb{R}^+$ mit $U(X_0, \delta) \subset M$ und

$$|\partial_i \partial_j f(X_0 + H) - \partial_i \partial_j f(X_0)| < \frac{a}{n^2} \quad \text{für alle} \quad H \in \mathbb{R}^n \quad \text{mit} \quad \|H\| < \delta$$

$(i, j = 1, \ldots, n)$. Für beliebiges $E = (e_1, \ldots, e_n) \in \mathbb{R}^n$ mit $\|E\| = 1$ und für $\|H\| < \delta$ folgt

$$\begin{aligned} & |Q(f, X_0 + H; E) - Q(f, X_0; E)| \\ \leq \; & \sum_{i,j=1}^{n} |\partial_i \partial_j f(X_0 + H) - \partial_i \partial_j f(X_0)| \, |e_i| \, |e_j| \\ \leq \; & \frac{a}{n^2} \sum_{i,j=1}^{n} 1 = a. \end{aligned}$$

Also ist $Q(f, X_0 + H; E) > a > 0$ für $\|E\| = 1$ und $\|H\| < \delta$; daraus folgt $Q(f, X_0 + H; Y) > 0$ für alle $Y \in \mathbb{R}^n \setminus \{0\}$. Für alle $H \in \mathbb{R}^n$ mit $\|H\| < \delta$ ist also $Q(f, X_0 + H; \cdot)$ positiv definit.

Nun gibt es nach (10.2) (für $k = 1$) und wegen der Voraussetzung $\nabla(X_0) = 0$ ein $c \in (0, 1)$ mit

$$f(X_0 + H) = f(X_0) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_i \partial_j f(X_0 + cH) h_i h_j .$$

Für $0 < \|H\| < \delta$ folgt

$$f(X_0 + H) > f(X_0).$$

Die Funktion f hat also in X_0 ein strenges lokales Minimum. ■

Umgekehrt läßt sich aus dem Vorliegen eines lokalen Extremums lediglich folgern, daß die Hesse-Form in diesem Punkt semidefinit ist:

(10.9) **Satz.** *Sei M offen und $f \in C^2(M)$. In X_0 habe f ein lokales Minimum (lokales Maximum). Dann ist die Hesse-Form $Q(f, X_0, \cdot)$ positiv (negativ) semidefinit.*

Beweis. Die Funktion f habe etwa in X_0 ein lokales Minimum. Angenommen, $Q(f, X_0; \cdot)$ wäre nicht positiv semidefinit. Dann gibt es ein $E = (e_1, \dots, e_n) \in \mathbb{R}^n$ mit $Q(f, X_0; E) < 0$. Analog wie im Beweis von (10.8) zeigt man die Existenz eines $\delta \in \mathbb{R}^+$ mit $U(X_0, \delta) \subset M$ und $Q(f, X_0 + H; E) < 0$ für alle $H \in \mathbb{R}^n$ mit $\|H\| < \delta$. Wählen wir $H := \lambda E$ mit hinreichend kleinem $\lambda \in \mathbb{R}^+$, so folgt $Q(f, X_0 + cH; H) < 0$ für $c \in (0, 1)$. Aus (10.2) ergibt sich dann $f(X_0 + H) < f(X_0)$. Da hier $\|H\| > 0$ beliebig klein gewählt werden kann, hat f in X_0 kein lokales Minimum, ein Widerspruch. ■

(10.10) **Korollar.** *Sei M offen und $f \in C^2(M)$. Ist die Hesse-Form $Q(f, X_0; \cdot)$ von f in X_0 indefinit, so hat f in X_0 kein lokales Extremum.*

Aus den vorstehenden Sätzen ergibt sich, daß man zur Untersuchung von lokalen Extrema Kriterien haben muß für den Definitheitscharakter quadratischer Formen. Wir geben ein solches Kriterium an für den Fall $n = 2$.

(10.11) **Satz.** *Die durch*

$$q(x, y) = a_{11} x^2 + 2a_{12} xy + a_{22} y^2 \quad \text{für } (x, y) \in \mathbb{R}^2$$

definierte quadratische Form q ist genau dann positiv definit, wenn

$$a_{11} > 0 \qquad \text{und} \qquad a_{11} a_{22} - a_{12}^2 > 0$$

ist. Im Fall $a_{11}a_{22} - a_{12}^2 < 0$ *ist sie indefinit.*

Beweis. Sei q positiv definit. Dann ist $a_{11} = q(1,0) > 0$. Durch Ausrechnen bestätigt man die Identität

$$q(x,y) = \frac{1}{a_{11}}\left[(a_{11}x + a_{12}y)^2 + (a_{11}a_{22} - a_{12}^2)y^2\right].$$

Daraus folgt

$$a_{11}a_{22} - a_{12}^2 = a_{11}q\left(-\frac{a_{12}}{a_{11}}, 1\right) > 0.$$

Umgekehrt folgt aus $a_{11} > 0$ und $a_{11}a_{22} - a_{12}^2 > 0$ wegen der obigen Identität sofort $q(x,y) > 0$ für $(x,y) \neq (0,0)$. ■

11 Differenzierbare Abbildungen

Im ersten Teil dieses Abschnitts befassen wir uns mit differenzierbaren Abbildungen zwischen Mengen gleicher Dimension. Solche Abbildungen treten in den Anwendungen unter anderem als sogenannte Koordinatentransformationen auf. Wir erläutern dies zunächst an einem speziellen Beispiel.

Gegeben sei eine offene Menge $M \subset \mathbb{R}^2$ und eine Funktion $f : M \to \mathbb{R}$. Bei konkreten Fragestellungen ist es manchmal nicht zweckmäßig, einen Punkt aus M durch seine Koordinaten bezüglich der Standardbasis zu beschreiben, sondern man kann oft mit Vorteil andere Zahlenpaare wählen, durch die die Punkte ebenso festgelegt werden, die aber der Problemstellung besser angepaßt sind („krummlinige Koordinaten"). Nehmen wir etwa an, die Funktion f sei rotationssymmetrisch, das heißt $f(X)$ hänge nur vom „Radius" $\|X\|$ ab. Dann empfiehlt es sich, in $M \setminus \{0\}$ Polarkoordinaten einzuführen: Jeder Punkt X ist eindeutig festgelegt durch seinen Abstand r vom Nullpunkt und (im Fall $X \neq 0$) durch den Winkel φ, den der Vektor X mit E_1 bildet. Für $X = (x,y)$ gilt dann

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi, \end{aligned}$$

und man schreibt etwa

$$f(x,y) = f(r\cos\varphi, r\sin\varphi) = \tilde{f}(r,\varphi)$$

und sagt, man habe damit f auf Polarkoordinaten bezogen. Dies ist folgendermaßen zu präzisieren. Wir setzen

$$U := \{(r,\varphi) \in \mathbb{R}^2 : r > 0,\ \ 0 \leq \varphi < 2\pi\}$$

und definieren eine Abbildung $\tau : U \to \mathbb{R}^2$ durch

$$\tau(r, \varphi) := (r \cos\varphi, r \sin\varphi).$$

Offenbar ist $\tau(U) = \mathbb{R}^2 \setminus \{0\}$, und τ ist injektiv. Die Abbildung τ ist differenzierbar, ihre Umkehrabbildung τ^{-1} ebenfalls, denn sie ist explizit gegeben durch

$$\begin{aligned} r &= \sqrt{x^2 + y^2}, \\ \varphi &= \operatorname{arccot} \frac{x}{y} + k\pi \qquad \text{für } y \neq 0, \\ \varphi &= \arctan \frac{y}{x} + k\pi \qquad \text{für } x \neq 0, \end{aligned}$$

wo $k \in \{0, 1, 2\}$ jeweils so zu wählen ist, daß $\varphi(1, 0) = 0$ und φ für $(x, y) \neq (a, 0)$, $a > 0$, stetig ist. Die oben mit $\tilde{f}$ bezeichnete Funktion ist die Komposition $\tilde{f} = f \circ \tau$. Aus ihr gewinnt man f in der Form $f = \tilde{f} \circ \tau^{-1}$. Dabei sind jeweils noch die Definitionsbereiche zu beachten sowie die Tatsache, daß $(0, 0)$ nicht im Bild von τ liegt.

Allgemein bezeichnet man als *Koordinatentransformation* eine Abbildung $F : M \to \mathbb{R}^n$ mit $M \subset \mathbb{R}^n$ (M offen) mit den Eigenschaften: F ist injektiv, F und die Umkehrabbildung F^{-1} (definiert auf $F(M)$) sind differenzierbar. Sei F eine Abbildung mit diesen Eigenschaften. Setzen wir $X = (x_1, \ldots, x_n)$, $F(X) = Y = (y_1, \ldots, y_n)$, so ist also

$$(*) \qquad \begin{cases} y_1 = f_1(x_1, \ldots, x_n), \\ \vdots \\ y_n = f_n(x_1, \ldots, x_n), \end{cases}$$

wo $f_1, \ldots, f_n$ die Koordinatenfunktionen von F sind. Sind $g_1, \ldots, g_n$ die Koordinatenfunktionen der Umkehrabbildung $F^{-1} : F(M) \to M$, so ist also $(*)$ äquivalent mit

$$(**) \qquad \begin{cases} x_1 = g_1(y_1, \ldots, y_n), \\ \vdots \\ x_n = g_n(y_1, \ldots, y_n). \end{cases}$$

Im allgemeinen wird es nicht möglich sein, bei explizit gegebenen Funktionen $f_1, \ldots, f_n$ die Funktionen $g_1, \ldots, g_n$ explizit zu berechnen (also das Gleichungssystem $(*)$ „aufzulösen nach $x_1, \ldots, x_n$“). Häufig benötigt man jedoch nur die partiellen Ableitungen der Funktionen $g_1, \ldots, g_n$, und diese kann man stets explizit berechnen aus den partiellen Ableitungen der Funktionen $f_1, \ldots, f_n$. Dies ergibt sich in folgender Weise aus bekannten Resultaten.

Wir setzen voraus, daß die Abbildung $F : M \to \mathbb{R}$ differenzierbar ist in einem Punkt $X \in M$, daß die Umkehrabbildung F^{-1} existiert und differenzierbar ist im

Punkt $Y := F(X)$. Da die Abbildung $F^{-1} \circ F$ auf M die Identität ist, gilt nach der Kettenregel

$$D(F^{-1})_Y \circ DF_X = \text{Identität},$$

oder in Koordinatenschreibweise

$$JF^{-1}(Y)JF(X) = \text{Einheitsmatrix}.$$

Es folgt, daß das Differential DF_X und damit die Funktionalmatrix $JF(X)$ regulär sind und daß

$$JF^{-1}(Y) = JF(X)^{-1}$$

ist. Ausgeschrieben lautet diese Gleichung

$$\begin{pmatrix} \partial_1 g_1(Y) & \cdots & \partial_n g_1(Y) \\ \vdots & & \vdots \\ \partial_1 g_n(Y) & \cdots & \partial_n g_n(Y) \end{pmatrix} = \begin{pmatrix} \partial_1 f_1(X) & \cdots & \partial_n f_1(X) \\ \vdots & & \vdots \\ \partial_1 f_n(X) & \cdots & \partial_n f_n(X) \end{pmatrix}^{-1}.$$

Nach bekannten Regeln für die Berechnung einer inversen Matrix ergibt sich hieraus für jede Funktion $\partial_j g_i$ eine Darstellung als Quotient, wo Zähler und Nenner Polynome in den Funktionen $\partial_k f_m \circ F^{-1}$ $(k, m = 1, \ldots, n)$ sind und der Nenner nicht verschwindet. Durch weitere Differentiationen (nach der Kettenregel) ergibt sich dann: Sind $f_1, \ldots, f_n$ Funktionen der Klasse C^r, so sind $g_1, \ldots, g_n$ von der Klasse C^r.

Von besonderer Wichtigkeit ist nun naheliegenderweise die Frage, ob und gegebenenfalls wie man einer vorgelegten differenzierbaren Abbildung $F : M \to \mathbb{R}^n$ (mit $M \subset \mathbb{R}^n$) ansehen kann, ob sie eine differenzierbare Umkehrabbildung besitzt, also eine Koordinatentransformation definiert. Notwendig ist, wie wir eben gesehen haben, jedenfalls die Regularität (Umkehrbarkeit) des Differentials, also das Nichtverschwinden der Funktionaldeterminante. Für eine differenzierbare Funktion $f : (a, b) \to \mathbb{R}$ ist die Bedingung $f'(t) \neq 0$ für $t \in (a, b)$ bekanntlich auch hinreichend für die Existenz und Differenzierbarkeit der Umkehrfunktion, aber im Fall $n > 1$ ist die Sachlage nicht so einfach.

Beispiel. Sei $M = \mathbb{R}^2 \setminus \{0\}$ und $F : M \to \mathbb{R}^2$ definiert durch

$$F(x, y) := (x^2 - y^2, 2xy) \qquad \text{für } (x, y) \in \mathbb{R}^2.$$

Die Funktionalmatrix

$$JF(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$$

hat die Determinante $4(x^2 + y^2) \neq 0$, ist also regulär. Wegen $F(X) = F(-X)$ ist die Abbildung F aber nicht umkehrbar.

Die Regularität des Differentials an einer gegebenen Stelle reicht jedoch aus, um zumindest in einer hinreichend kleinen Umgebung dieser Stelle die Existenz und Differenzierbarkeit der Umkehrabbildung zu sichern. Diesen wichtigen Satz wollen wir nun beweisen.

Die Abbildung $F : M \to \mathbb{R}^n$ heißt *von der Klasse* C^r, wenn alle Koordinatenfunktionen von F r-mal stetig differenzierbar sind. Wir stellen zunächst einen Hilfssatz bereit.

(11.1) **Hilfssatz.** *Sei* $M \subset \mathbb{R}^n$ *offen, sei* $F : M \to \mathbb{R}^n$ *eine Abbildung der Klasse* C^1, *sei* $L : \mathbb{R}^n \to \mathbb{R}^n$ *eine lineare Abbildung. Sind* $X, Y \in M$ *Punkte mit* $[X, Y] \subset M$, *so gilt*

$$\|F(X) - F(Y) - L(X - Y)\| \leq \|X - Y\| \max\{\|DF_Z - L\| \mid Z \in [X, Y]\}.$$

Beweis. Setze $G(X) := F(X) - L(X)$ für $X \in M$; dann ist $DG_Z = DF_Z - L$ für $Z \in M$. Da F von der Klasse C^1 ist, ist die Funktion $Z \mapsto \|DF_Z - L\|$ stetig, daher existiert

$$\max\{\|DF_Z - L\| \mid Z \in [X, Y]\} =: c.$$

Aus dem Mittelwertsatz (8.6) folgt jetzt

$$\|G(X) - G(Y)\| \leq c\|X - Y\|,$$

also die Behauptung. ■

(11.2) **Satz** (über die Umkehrabbildung). *Sei* $M \subset \mathbb{R}^n$ *offen,* $X_0 \in M$, *sei* $F : M \to \mathbb{R}^n$ *eine Abbildung der Klasse* C^r (*für ein* $r \in \mathbb{N}$). *Das Differential* DF_{X_0} *von* F *in* X_0 *sei regulär* (*d.h. die Funktionaldeterminante von* F *in* X_0 *sei* $\neq 0$). *Dann gibt es eine offene Umgebung* $U \subset M$ *von* X_0 *mit folgenden Eigenschaften*:

(a) *die Einschränkung* $F|U$ *ist injektiv,*

(b) *die Bildmenge* $V := F(U)$ *ist offen,*

(c) *die Umkehrabbildung* $(F|U)^{-1} : V \to U$ *ist von der Klasse* C^r.

Beweis. Im folgenden bezeichnet I die identische Abbildung von $\mathbb{R}^n$ auf sich. Zur Abkürzung wird für $\alpha > 0$

$$U_\alpha := U(0, \alpha) = \{X \in \mathbb{R}^n \mid \|X\| < \alpha\}$$

gesetzt. Wir machen zunächst eine spezielle Voraussetzung.

Voraussetzung. $X_0 = 0$, $F(0) = 0$, $DF_0 = I$.

Nach Voraussetzung ist die Funktion $X \mapsto \|DF_X - I\|$ in M stetig, und es ist $\|DF_0 - I\| = 0$. Daher existiert zu vorgegebenem $\epsilon > 0$ ein $\alpha > 0$ mit $U_\alpha \subset M$ und

$$\|DF_X - I\| \le \epsilon \qquad \text{für alle } X \in U_\alpha.$$

Aus (11.1) folgt dann für alle $X, Y \in U_\alpha$

$$\|F(X) - F(Y) - (X - Y)\| \le \epsilon \|X - Y\| \tag{$*$}$$

und hieraus durch Anwendung der Dreiecksungleichung

$$(1 - \epsilon)\|X - Y\| \le \|F(X) - F(Y)\|. \tag{$**$}$$

Wir wählen nun ein positives $\epsilon < 1$ und dazu $\alpha > 0$ wie oben und so, daß $\overline{U_\alpha} \subset M$ ist.

<u>1. Behauptung.</u> $U_{(1-\epsilon)\alpha} \subset F(U_\alpha)$.

Beweis. Sei $Y \in U_{(1-\epsilon)\alpha}$. Es ist ein $X \in U_\alpha$ zu finden mit $Y = F(X)$. Hierzu benutzen wir den Banachschen Fixpunktsatz. Definiere $\Phi : \overline{U_\alpha} \to \mathbb{R}^n$ durch

$$\Phi(X) := Y - F(X) + X \qquad \text{für } X \in \overline{U_\alpha}.$$

Für $X \in \overline{U_\alpha}$ gilt wegen $F(0) = 0$ nach $(*)$

$$\|\Phi(X)\| \le \|Y\| + \|F(X) - X\| \le \|Y\| + \epsilon\|X\| < (1-\epsilon)\alpha + \epsilon\alpha = \alpha,$$

also $\Phi(X) \in U_\alpha$. Somit bildet Φ die Menge $\overline{U_\alpha}$ in sich ab. Für $X, Z \in \overline{U_\alpha}$ gilt nach $(*)$

$$\|\Phi(X) - \Phi(Z)\| = \|F(X) - F(Z) - (X - Z)\| \le \epsilon\|X - Z\|,$$

wegen $\epsilon < 1$ ist also Φ kontrahierend. Da $\overline{U_\alpha}$ nach (6.10) und (2.9) vollständig ist, gibt es nach dem Banachschen Fixpunktsatz (3.2) einen Punkt $X \in \overline{U_\alpha}$ mit $\Phi(X) = X$. Wie oben gezeigt, ist $X = \Phi(X) \in U_\alpha$. Es ist also ein Punkt $X \in U_\alpha$ gefunden mit $Y = F(X)$. Damit ist die 1. Behauptung bewiesen.

Nun setzen wir $V := U_{(1-\epsilon)\alpha}$ und $U := U_\alpha \cap F^{-1}(V)$. Dann ist U eine offene Umgebung von 0 und $F(U) = V$. Aus $(**)$ folgt, daß $F|U$ injektiv ist. Für den Moment bezeichne $G : V \to U$ die Umkehrabbildung von $F|U$.

<u>2. Behauptung.</u> G ist in 0 differenzierbar.

Beweis. Sei $\epsilon' \in \mathbb{R}^+$ gegeben. Wie anfangs gezeigt ($(*)$ mit $Y = 0$) existiert ein $\alpha' \in \mathbb{R}^+$ mit $U_{\alpha'} \subset M$ und

$$\|F(X) - X\| \le \frac{\epsilon'}{1 + \epsilon'}\|X\| \qquad \text{für } \|X\| < \alpha'$$

und daher

$$\|X\| \leq (1+\epsilon')\|F(X)\| \qquad \text{für } \|X\| < \alpha'.$$

Sei $H \in V$ ein Vektor mit $\|H\| < \alpha'(1-\epsilon)$. Für $X := G(H)$ ist dann $X \in U$, also $\|X\| < \alpha$ und daher nach $(**)$

$$\|X\| \leq \frac{1}{1-\epsilon}\|F(X)\| = \frac{1}{1-\epsilon}\|H\| < \alpha',$$

also

$$\|G(H) - H\| = \|X - F(X)\| \leq \frac{\epsilon'}{1+\epsilon'}\|X\| \leq \epsilon'\|F(X)\| = \epsilon'\|H\|,$$

somit

$$\frac{\|G(H)-H\|}{\|H\|} \leq \epsilon' \quad \text{für } 0 < \|H\| < \alpha'(1-\epsilon).$$

Damit ist

$$\lim_{H\to 0} \frac{G(H)-H}{\|H\|} = 0$$

gezeigt, also (wegen $G(0) = 0$) die Differenzierbarkeit von G in 0 sowie $DG_0 = I$.

Wir lassen nun die spezielle Voraussetzung fallen, daß $X_0 = 0$, $F(0) = 0$ und $DF_0 = I$ sein soll.

Es ist also jetzt vorausgesetzt, daß $X_0 \in M$ ein Punkt ist derart, daß DF_{X_0} regulär ist. Durch passende Transformationen gewinnen wir eine Abbildung $\tilde{F}$, die die spezielle Voraussetzung erfüllt. Hierzu bezeichne $T_Z : \mathbb{R}^n \to \mathbb{R}^n$ die Translation um den Vektor Z, also $T_Z(X) := X + Z$ für $X \in \mathbb{R}^n$. Wir setzen

$$\begin{aligned} L &:= DF_{X_0}, \\ \tilde{M} &:= (L \circ T_{-X_0})(M) \\ \tilde{F}(X) &:= T_{-F(X_0)} \circ F \circ T_{X_0} \circ L^{-1}(X) \qquad \text{für } X \in \tilde{M}. \end{aligned}$$

Wie man leicht nachprüft, gilt $0 \in \tilde{M}$, $\tilde{F}(0) = 0$ und $D\tilde{F}_0 = I$. Anwendung des bereits Bewiesenen auf $\tilde{F}$ ergibt dann die lokale Umkehrbarkeit und Differenzierbarkeit von F. Hierzu beachte man, daß

$$F = T_{F(X_0)} \circ \tilde{F} \circ L \circ T_{-X_0}$$

ist. Die Differenzierbarkeit der Umkehrabbildung ergibt sich zunächst nur im Punkt $F(X_0)$. Nun ist aber in X_0 die Funktionaldeterminante von F verschieden von Null, und das gilt, da sie stetig ist, noch in einer ganzen Umgebung von X_0. Auf jeden Punkt dieser Umgebung kann man das bereits Bewiesene anwenden, erhält also die Differenzierbarkeit der Umkehrabbildung. Daß mit F auch die