

3 Differenzierbare Funktionen

In diesem Kapitel entwickeln wir die Differentialrechnung für Funktionen und Abbildungen in mehreren Variablen. Da sich alle im Folgenden betrachteten Beispiele und Anwendungen auf reellwertige Funktionen beziehen, beschränken wir uns auf diesen Fall, d. h. auf $\mathbb{K} = \mathbb{R}$.

3.1 Partielle und totale Ableitung

Definition 3.1 (Partielle Ableitung): *i) Sei $D \subset \mathbb{R}^n$ eine offene Menge. Eine Funktion $f : D \rightarrow \mathbb{R}$ heißt in einem Punkt $x \in D$ „partiell differenzierbar“ bzgl. der i -ten Koordinatenrichtung ($e^{(i)}$ der entsprechende kartesische Richtungsvektor), falls der Limes*

$$\lim_{h \rightarrow 0} \frac{f(x + he^{(i)}) - f(x)}{h} =: \frac{\partial f}{\partial x_i}(x) =: \partial_i f(x)$$

existiert; dieser heißt die „partielle Ableitung bzgl. x_i “ von f in x .

ii) Existieren in allen Punkten $x \in D$ alle partiellen Ableitungen so heißt f „partiell differenzierbar“. Sind alle partiellen Ableitungen stetige Funktionen auf D , so heißt f „stetig partiell differenzierbar“.

iii) Eine vektorwertige Funktion $f = (f_1, \dots, f_m) : D \rightarrow \mathbb{R}^m$ heißt „(stetig) partiell differenzierbar“, wenn alle ihre Komponenten f_i (stetig) partiell differenzierbar sind.

Die partielle Ableitung kann als gewöhnliche Ableitung interpretiert werden. Bei der Funktion $f(x) = f(x_1, \dots, x_n)$ seien alle bis auf das i -te Argument festgehalten und die Funktion $f(\xi) := f(x_1, \dots, x_{i-1}, \xi, x_{i+1}, \dots, x_n)$ als Funktion von ξ allein betrachtet. Die partielle Ableitung von f bzgl. x_i ist dann gerade die gewöhnliche Ableitung von \tilde{f} :

$$\frac{\partial f}{\partial x_i}(x) = \frac{d\tilde{f}}{d\xi}(\xi).$$

Deshalb gelten für die partielle Ableitung analoge Regeln wie für die gewöhnliche Ableitung, insbesondere die Produkt- und Quotientenregel:

$$\partial_i(fg) = g\partial_i f + f\partial_i g, \quad \partial_i\left(\frac{f}{g}\right) = \frac{g\partial_i f - f\partial_i g}{g^2}. \quad (3.1.1)$$

Für eine partiell differenzierbare Funktion $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ und eine differenzierbare Funktion $F : I \rightarrow \mathbb{R}$ mit $f(D) \subset I$ gilt die einfache Kettenregel

$$\partial_i F(f(x)) = F'(f(x))\partial_i f(x), \quad x \in D. \quad (3.1.2)$$

Eine weitere Verallgemeinerung der Kettenregel werden wir später kennenlernen.

Beispiel 3.1: Wir betrachten einige elementare Beispiele:

i) Das quadratische Polynom

$$p(x) := a_{20}x_1^2 + 2a_{11}x_1x_2 + a_{02}x_2^2 + a_{10}x_1 + a_{01}x_2 + a_{00}$$

ist auf ganz \mathbb{R}^n stetig partiell differenzierbar und hat die partiellen Ableitungen

$$\partial_1 p(x) = 2a_{20}x_1 + 2a_{11}x_2 + a_{10}, \quad \partial_2 p(x) = 2a_{11}x_1 + 2a_{02}x_2 + a_{01}.$$

ii) Die Abstandsfunktion

$$r(x) := \|x\|_2 = (x_1^2 + \dots + x_n^2)^{1/2}$$

ist in $D = \mathbb{R}^n \setminus \{0\}$ stetig partiell differenzierbar mit den partiellen Ableitungen (Anwendung der gewöhnlichen Kettenregel):

$$\partial_i r(\dots, x_i, \dots) = \frac{1}{2} \frac{2x_i}{(\dots + x_i^2 + \dots)^{1/2}} = \frac{x_i}{r(x)}.$$

iii) Sei $F : \mathbb{R}_+ \rightarrow \mathbb{R}$ eine beliebige, differenzierbare Funktion. Dann ist die zusammengesetzte Funktion $f(x) := F(r(x))$ auf ganz $D = \mathbb{R}^n \setminus \{0\}$ definiert und dort partiell differenzierbar. Ihre partielle Ableitungen erhält man mit der Kettenregel als

$$\partial_i f(x) = F'(r(x)) \frac{x_i}{r(x)}, \quad i = 1, \dots, n.$$

Z. B. hat die Funktion $f(x) = \ln(r(x))$ die partiellen Ableitungen

$$\partial_i f(x) = \frac{x_i}{r(x)^2}, \quad i = 1, \dots, n.$$

iv) Wir betrachten die folgende auf \mathbb{R}^2 definierte Funktion:

$$f(x) := \frac{x_1 x_2}{r(x)^4}, \quad x \neq 0, \quad f(0) := 0.$$

Diese ist partiell differenzierbar. Für $x \neq 0$ erhalten wir ihre partiellen Ableitungen wieder mit Hilfe der Produkt- und Kettenregel als

$$\begin{aligned} \partial_1 f(x) &= \partial_1(x_1 x_2) r(x)^{-4} + x_1 x_2 \partial_1(r(x)^{-4}) \\ &= x_2 r(x)^{-4} + x_1 x_2 (-4) r(x)^{-5} x_1 r^{-1} = x_2 r(x)^{-4} - 4x_1^2 x_2 r(x)^{-6} \end{aligned}$$

und analog für $i = 2$. Im Punkt $x = 0$ ist wegen $f(he^{(i)}) = f(0) = 0$:

$$\lim_{h \rightarrow 0} \frac{f(he^{(i)}) - f(0)}{h} = 0, \quad i = 1, 2.$$

Die Funktion f ist aber in $x = 0$ nicht stetig, denn für die Punkte $x_\varepsilon := (\varepsilon, \varepsilon)$ gilt $\|x_\varepsilon\| \rightarrow 0$ ($\varepsilon \rightarrow 0$) aber wegen $r(x_\varepsilon) = \sqrt{2}\varepsilon$:

$$f(x_\varepsilon) = \frac{\varepsilon^2}{4\varepsilon^4} \rightarrow \infty \quad (\varepsilon \rightarrow 0).$$

Wir sehen, dass für Funktionen in mehreren Variablen, im Gegensatz zum Fall $n = 1$, die partielle Differenzierbarkeit (im obigen Sinne) nicht notwendig die Stetigkeit erfordert. In diesem Beispiel sind aber die partiellen Ableitungen nicht gleichmäßig beschränkt.

Satz 3.1: Sei $D \subset \mathbb{R}^n$ offen. Die Funktion $f : D \rightarrow \mathbb{R}$ habe in einer Kugelumgebung $K_r(x) \subset D$ eines Punktes $x \in D$ beschränkte partielle Ableitungen (oder f sei überhaupt in $K_r(x)$ stetig partiell differenzierbar):

$$\sup_{x \in K_r(x)} |\partial_i f(x)| \leq M, \quad i = 1, \dots, n.$$

Dann ist f stetig im Punkt x .

Beweis: Wir geben den Beweis nur für den Fall $n = 2$. Seine Übertragbarkeit auf den allgemeinen Fall $n \in \mathbb{N}$ ist offensichtlich. Zunächst gilt für $y = (y_1, y_2) \in K_r(x)$:

$$f(y_1, y_2) - f(x_1, x_2) = f(y_1, y_2) - f(x_1, y_2) + f(x_1, y_2) - f(x_1, x_2).$$

Nach dem Mittelwertsatz der Differentialrechnung existieren Zwischenstellen $\xi = \xi(y_2)$, $\eta = \eta(x_1)$ zwischen x_1 und y_1 bzw. x_2 und y_2 , so dass

$$f(y_1, y_2) - f(x_1, x_2) = \partial_1 f(\xi, y_2)(y_1 - x_1) + \partial_2 f(x_1, \eta)(y_2 - x_2).$$

Wegen der Beschränktheit der partiellen Ableitungen in $K_r(x)$ folgt

$$|f(y_1, y_2) - f(x_1, x_2)| \leq M(|y_1 - x_1| + |y_2 - x_2|).$$

Für beliebiges $\varepsilon \in (0, r]$ gilt also $|f(y) - f(x)| < \varepsilon$ für $\|y - x\|_1 < \delta := \varepsilon/M$, d.h.: f ist stetig in x (genauer sogar Lipschitz-stetig). Q.E.D.

Sind für eine partiell differenzierbare Funktion $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ die partiellen Ableitungen $\partial_i f : D \rightarrow \mathbb{R}$ wieder partiell differenzierbar, so heißt f „zweimal partiell differenzierbar“ mit den partiellen Ableitungen zweiter Ordnung

$$\partial_i \partial_j f(x) := \frac{\partial^2 f}{\partial x_i \partial x_j}(x) := \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j}(x) \right).$$

Allgemein kann man partielle Ableitungen k -ter Ordnung bilden, welche mit

$$\partial_{i_1} \dots \partial_{i_k} f(x) = \frac{\partial}{\partial x_{i_1}} \dots \frac{\partial}{\partial x_{i_k}} f(x)$$

bezeichnet werden. Eine Funktion heißt dann „ k -mal stetig partiell differenzierbar“, wenn alle partiellen Ableitungen k -ter Ordnung von f existieren und stetig sind.

Beispiel 3.2: Die durch

$$f(x_1, x_2) := \frac{x_1^3 x_2 - x_1 x_2^3}{x_1^2 + x_2^2}, \quad (x_1, x_2) \neq (0, 0), \quad f(0, 0) := 0,$$

definierte Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ zeigt, dass i. Allg. die Reihenfolge der partiellen Ableitungen nicht vertauschbar ist. In der Tat ist f überall zweimal partiell differenzierbar, es ist aber (Übungsaufgabe)

$$\partial_1 \partial_2 f(0, 0) \neq \partial_2 \partial_1 f(0, 0).$$

In $(x_1, x_2) = (0, 0)$ sind die zweiten partiellen Ableitungen aber nicht stetig.

Satz 3.2 (Vertauschbarkeit der Differentiationsreihenfolge): Sei $D \subset \mathbb{R}^n$ offen. Die Funktion $f : D \rightarrow \mathbb{R}$ sei in einer Umgebung $K_r(x) \subset D$ eines Punktes $x \in D$ zweimal stetig partiell differenzierbar. Dann gilt:

$$\partial_i \partial_j f(x) = \partial_j \partial_i f(x), \quad i, j = 1, \dots, n. \quad (3.1.3)$$

Allgemein ist für eine k -mal stetig partiell differenzierbare Funktion die Reihenfolge der partiellen Ableitungen vertauschbar.

Beweis: i) Wir führen den Beweis wieder nur für den Fall $n = 2$. Sei

$$A := f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2)$$

und $\varphi(x_1) := f(x_1, x_2 + h_2) - f(x_1, x_2)$. Dann ist

$$A = \varphi(x_1 + h_1) - \varphi(x_1).$$

Mit dem Mittelwertsatz bzgl. x_1 erhalten wir

$$A = h_1 \varphi'(x_1 + \theta_1 h_1), \quad \theta_1 \in (0, h_1).$$

Wegen $\varphi'(x_1) = \partial_1 f(x_1, x_2 + h_2) - \partial_1 f(x_1, x_2)$ folgt wieder mit dem Mittelwertsatz diesmal bzgl. x_2 :

$$\varphi'(x_1) = h_2 \partial_2 \partial_1 f(x_1, x_2 + \theta'_1 h_2), \quad \theta'_1 \in (0, h_2).$$

Dies impliziert dann

$$\varphi'(x_1 + \theta_1 h_1) = h_2 \partial_2 \partial_1 f(x_1 + \theta_1 h_1, x_2 + \theta'_1 h_2)$$

und somit

$$A = h_1 h_2 \partial_2 \partial_1 f(x_1 + \theta_1 h_1, x_2 + \theta'_1 h_2).$$

Wir verfahren analog mit x_2 und erhalten für $\psi(x_2) := f(x_1 + h_1, x_2) - f(x_1, x_2)$:

$$A = \psi(x_2 + h_2) - \psi(x_2) = h_2 \psi'(x_2 + \theta_2 h_2) = h_1 h_2 \partial_1 \partial_2 f(x_1 + \theta_2 h_1, x_2 + \theta'_2 h_2).$$

Damit wird

$$\partial_2 \partial_1 f(x_1 + \theta_1 h_1, x_2 + \theta'_1 h_2) = \frac{A}{h_1 h_2} = \partial_1 \partial_2 f(x_1 + \theta_2 h_1, x_2 + \theta'_2 h_2).$$

Wegen der Stetigkeit von $\partial_1 \partial_2 f$ und $\partial_2 \partial_1 f$ in $K_r(x)$ gilt für $h_1, h_2 \rightarrow 0$:

$$\partial_2 \partial_1 f(x_1, x_2) = \partial_1 \partial_2 f(x_1, x_2).$$

ii) Sei nun f k -mal stetig partiell differenzierbar. Die Identität

$$\partial_1 \dots \partial_k f(x) = \partial_{i_1} \dots \partial_{i_k} f(x)$$

für jede Permutation (i_1, \dots, i_k) von $(1, \dots, k)$ folgt mit Hilfe von (i) durch Induktion nach k . Q.E.D.

3.1.1 Begriffe der Vektoranalysis

Definition 3.2 (Gradient): Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine partiell differenzierbare Funktion. Der Vektor der ersten partiellen Ableitungen

$$\operatorname{grad}f(x) := (\partial_1 f(x), \dots, \partial_n f(x))^T \in \mathbb{R}^n$$

heißt der „Gradient“ von f im Punkt $x \in D$. Man schreibt auch $\operatorname{grad}f(x) = \nabla f(x)$ mit dem sog. „Nabla-Operator“ (vektorieller Differentialoperator erster Ordnung)

$$\nabla = (\partial_1, \dots, \partial_n)^T.$$

Definition 3.3 (Hesse-Matrix): Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine zweimal partiell differenzierbare Funktion. Die Matrix der zweiten partiellen Ableitungen

$$H_f(x) := (\partial_i \partial_j f(x))_{i,j=1}^n \in \mathbb{R}^{n \times n}$$

heißt die „Hesse¹-Matrix“ von f im Punkt $x \in D$. Man schreibt auch $H_f(x) = \nabla^2 f(x)$.

Definition 3.4 (Jacobi-Matrix): Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}^m$ eine partiell differenzierbare Vektorfunktion. Die Matrix der ersten partiellen Ableitungen

$$J_f(x) := (\partial_j f_i(x))_{i=1,j=1}^{m,n} \in \mathbb{R}^{m \times n}$$

heißt die „Funktionalmatrix“ (oder auch die „Jacobi²-Matrix“) von f im Punkt $x \in D$. Man schreibt auch $J_f(x) = \nabla f(x) = f'(x)$. Im Fall $m = n$ wird die Determinante $\det J_f(x)$ von $J_f(x)$ auch „Funktionaldeterminante“ oder „Jacobi-Determinante“ genannt. Die Zeilen von $J_f(x)$ werden also gerade durch die (transponierten) Gradienten der Funktionen $f_i(x)$ gebildet.

Beispiel 3.3: Die Abstandsfunktion $r(x) = \|x\|_2$ hat den Gradienten

$$\nabla r(x) = \left(\partial_i \left(\sum_{j=1}^n x_j^2 \right)^{1/2} \right)_{i=1}^n = \left(\frac{x_i}{r(x)} \right)_{i=1}^n$$

und die Hesse-Matrix

$$\nabla^2 f(x) = \left(\partial_j \frac{x_i}{r(x)} \right)_{i,j=1}^n = \left(\frac{\delta_{ij}}{r(x)} - \frac{x_i x_j}{r(x)^3} \right)_{i,j=1}^n = \left(\frac{\delta_{ij} r(x)^2 - x_i x_j}{r(x)^3} \right)_{i,j=1}^n.$$

¹Ludwig Otto Hesse (1811–1874): Deutscher Mathematiker; wirkte in Königsberg, Heidelberg und München; Beiträge zur Theorie der algebraischen Funktionen und Invarianten

²Carl Gustav Jakob Jacobi (1804–1851): Deutscher Mathematiker; schon als Kind hochbegabt; wirkte in Königsberg und Berlin; Beiträge zu vielen Bereichen der Mathematik: zur Zahlentheorie, zu elliptischen Funktionen, zu partiellen Differentialgleichungen, zu Funktionaldeterminanten und zur theoretischen Mechanik.

Diese Hesse-Matrix ist gerade die Jacobi-Matrix der Vektorfunktion $v(x) = x/r(x)$:

$$J_v(x) = (\partial_j v_i(x))_{i,j=1}^n = \left(\frac{\delta_{ij} r(x)^2 - x_i x_j}{r(x)^3} \right)_{i,j=1}^n.$$

Allgemein hat die partiell differenzierbare Funktion $f(x) = F(r(x))$ den Gradienten

$$\nabla f(x) = F'(r(x)) \frac{x}{r(x)}.$$

Als Folgerung aus der Produktregel für die partielle Differentiation folgt für den Gradienten die Produktregel

$$\nabla(fg) = g\nabla f + f\nabla g. \quad (3.1.4)$$

Im Folgenden verwenden wir für die sog. „innere“ Multiplikation zweier Vektoren $v, w \in \mathbb{R}^n$ je nach Situation die Bezeichnungen:

$$v \cdot w := v^T w = \sum_{i=1}^n v_i w_i = (v, w)_2.$$

Definition 3.5 (Divergenz): Sei $D \subset \mathbb{R}^n$ eine offene Menge und $v : D \rightarrow \mathbb{R}^n$ eine partiell differenzierbare Abbildung. Die skalare Funktion

$$\operatorname{div} v(x) := \partial_1 v_1(x) + \cdots + \partial_n v_n(x).$$

heißt die „Divergenz“ von v im Punkt $x \in D$. Mit dem Nabla-Operator schreiben wir auch (im Sinne der sog. „inneren“ Vektormultiplikation)

$$\operatorname{div} v(x) = \nabla \cdot v(x).$$

Als Folgerung aus der Produktregel für die partielle Differentiation folgt für die Divergenz des Produkts einer skalaren Funktion $f : D \rightarrow \mathbb{R}$ mit einer Vektorfunktion $g : D \rightarrow \mathbb{R}^n$:

$$\nabla \cdot (fg) = \nabla f \cdot g + f \nabla \cdot g. \quad (3.1.5)$$

Beispiel 3.4: Wir betrachten die Vektorfunktion $v(x) = x/r(x)$ auf $\mathbb{R}^n \setminus \{0\}$. Wegen

$$\operatorname{div} x = \sum_{i=1}^n \partial_i x_i = n, \quad x \cdot x = r(x)^2,$$

hat sie die Divergenz

$$\nabla \cdot v(x) = \frac{\operatorname{div} x}{r(x)} + x \cdot \nabla \left(\frac{1}{r(x)} \right) = \frac{n}{r(x)} - x \cdot \frac{x}{r(x)} \frac{1}{r(x)^2} = \frac{n-1}{r(x)}.$$

Definition 3.6 (Rotation): Sei $D \subset \mathbb{R}^3$ eine offene Menge und $v : D \rightarrow \mathbb{R}^3$ eine partiell differenzierbare Abbildung. Die Vektorfunktion

$$\operatorname{rot} v(x) := (\partial_2 v_3(x) - \partial_3 v_2(x), \partial_3 v_1(x) - \partial_1 v_3(x), \partial_1 v_2(x) - \partial_2 v_1(x))$$

heißt die „Rotation“ von v im Punkt $x \in D$. Mit dem Nabla-Operator schreiben wir auch (im Sinne der sog. „äußeren“ Vektormultiplikation)

$$\operatorname{rot} v(x) = \nabla \times v(x).$$

Beispiel 3.5: Wir berechnen die Rotation einiger einfacher Funktionen:

i) Die Rotation der Identitätsfunktion $v(x) = x$ ist $\operatorname{rot} x = 0$.

ii) Die Rotation der Funktion $v(x) = x/r(x)$ auf $\mathbb{R}^3 \setminus \{0\}$ erhalten wir über

$$\partial_i \frac{x_j}{r(x)} = \frac{r(x)\delta_{ij} - x_j x_i r(x)^{-1}}{r(x)^2} = \frac{r(x)^2 \delta_{ij} - x_j x_i}{r(x)^3}, \quad i, j = 1, 2, 3,$$

zu

$$\left(\operatorname{rot} \frac{x}{r(x)}\right)_1 = \frac{r(x)^2 \delta_{23} - x_3 x_2}{r(x)^3} - \frac{r(x)^2 \delta_{32} - x_2 x_3}{r(x)^3} = 0,$$

und analog für die anderen beiden Komponenten.

iii) Für eine zweimal stetig partiell differenzierbare Funktion $f : D \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ ist nach Satz 3.2 die Reihenfolge der partiellen Ableitungen vertauschbar, d. h.: Es gilt z. B.:

$$\partial_3 \partial_2 f(x) - \partial_2 \partial_3 f(x) = 0.$$

Dies impliziert

$$\operatorname{rot} \operatorname{grad} f(x) = (\partial_2 \partial_3 f(x) - \partial_3 \partial_2 f(x), \partial_3 \partial_1 f(x) - \partial_1 \partial_3 f(x), \partial_1 \partial_2 f(x) - \partial_2 \partial_1 f(x)) = 0.$$

Also ist die Rotation eines Gradienten Null. Dies bedeutet, dass eine Vektorfunktion v nur dann der Gradient einer skalaren Funktion sein kann, wenn ihre Rotation verschwindet.

Für eine zweimal stetig partiell differenzierbare Funktion $u : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ist

$$\operatorname{div} \operatorname{grad} u(x) = \sum_{i=1}^n \partial_i^2 u(x) =: \Delta u(x),$$

mit dem sog. „Laplace-Operator“ $\Delta := \operatorname{div} \operatorname{grad}$. Der Laplace-Operator spielt eine wichtige Rolle in den Differentialgleichungen der mathematischen Physik. Die partielle Differentialgleichung

$$\Delta u = 0 \tag{3.1.6}$$

heißt „Laplace-Gleichung“ oder „Potentialgleichung“; ihre Lösungen heißen „harmonische Funktionen“. Ihr „inhomogenes“ Gegenstück

$$\Delta u = f$$

mit einer gegebenen rechten Seite f wird „(inhomogene) Laplace-Gleichung“ oder „Poisson³-Gleichung“ genannt. Differentialgleichungen dieser Art werden im dritten Teil dieser Buchserie genauer besprochen.

Wir wollen eine „typische“ Klasse von Lösungen der Laplace-Gleichung in der Form $F(r(x))$ auf $\mathbb{R}^n \setminus \{0\}$ angeben. Es gilt

$$\operatorname{grad} F(r(x)) = F'(r(x)) \frac{x}{r}, \quad \Delta F(r(x)) = \operatorname{div}(\operatorname{grad} F(r(x)))$$

und folglich

$$\begin{aligned} \Delta F(r(x)) &= \operatorname{div}\left(F'(r(x)) \frac{x}{r}\right) = \sum_{i=1}^n \left\{ F''(r(x)) \frac{x_i}{r(x)} \frac{x_i}{r(x)} + F'(r(x)) \partial_i \left(\frac{x_i}{r}\right) \right\} \\ &= F''(r(x)) + \sum_{i=1}^n F'(r(x)) \left(\frac{1}{r(x)} - \frac{x_i^2}{r(x)^3} \right) \\ &= F''(r(x)) + \frac{n-1}{r(x)} F'(r(x)). \end{aligned}$$

Dies impliziert in zwei Dimensionen auf $\mathbb{R}^2 \setminus \{0\}$:

$$\Delta \ln(r(x)) = -\frac{1}{r(x)^2} + \frac{1}{r(x)^2} = 0,$$

und allgemein in $n \geq 3$ Dimensionen auf $\mathbb{R}^n \setminus \{0\}$:

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

Diese speziellen harmonischen Funktionen $\ln(r(x))$ in zwei und $r(x)^{2-n}$ in $n \geq 3$ Dimensionen heißen „Fundamentallösungen“ des Laplace-Operators. Sie spielen eine wichtige Rolle bei der expliziten Konstruktion von Lösungen der Laplace-Gleichung.

3.1.2 Totale Differenzierbarkeit

Die folgende Definition der „Ableitung“ einer Funktion in *mehreren* Variablen ist das Analogon der uns schon vertrauten Ableitung einer Funktion *einer* Variablen.

Definition 3.7: Sei $D \subset \mathbb{R}^n$ eine offene Menge. Eine Abbildung $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt in einem Punkt $x \in D$ „total differenzierbar“ (oder einfach „differenzierbar“), wenn sie im Punkt x linear approximierbar ist, d. h. wenn es eine lineare Abbildung $Df(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (das sog. „Differential“ von f) gibt, so dass in einer Umgebung von x gilt:

$$f(x+h) = f(x) + Df(x)h + \omega(h), \quad h \in \mathbb{R}^n, x+h \in D, \quad (3.1.7)$$

³Siméon Denis Poisson (1781–1840): Französischer Mathematiker und Physiker; Prof. in Paris; Beiträge zur mathematischen Formulierung der Physik, zum Magnetismus, zur Himmelsmechanik und zur Wahrscheinlichkeitsrechnung; einer der Begründer der Potentialtheorie.

mit einer Funktion $\omega : D \rightarrow \mathbb{R}^m$ mit der Eigenschaft

$$\lim_{x+h \in D, \|h\|_2 \rightarrow 0} \frac{\|\omega(h)\|_2}{\|h\|_2} = 0. \quad (3.1.8)$$

Die Beziehung (3.1.8) schreiben wir auch abgekürzt in der Form $\omega(h) = o(\|h\|_2)$.

Bemerkung 3.1: Für $n = m = 1$ stimmt die Definition der „totalen Ableitung“ mit der schon bekannten Definition der Ableitung von Funktionen einer Variablen überein.

Satz 3.3 (Differenzierbarkeit): Sei $D \subset \mathbb{R}^n$ offen. Für Abbildungen $f : D \rightarrow \mathbb{R}^m$ gilt:

- a) Ist f in $x \in D$ differenzierbar, so ist es in x auch partiell differenzierbar, und das Differential von f ist gerade die Funktional-Matrix $Df(x) = J_f(x)$.
- b) Ist f partiell differenzierbar in einer Umgebung von $x \in D$ und sind die partiellen Ableitungen in x stetig, so ist f auch in x differenzierbar.

Beweis: Wir geben den Beweis nur für $n = 2$ und $m = 1$.

a) Für differenzierbares f gilt für $i \in \{1, 2\}$:

$$\lim_{h_i \rightarrow 0} \frac{f(x + he^{(i)}) - f(x)}{h_i} = \lim_{h_i \rightarrow 0} (Df(x)e^{(i)} + h_i^{-1}\omega(h_i)) = Df(x)e^{(i)},$$

d. h.: f ist partiell differenzierbar.

b) Für stetig partiell differenzierbares f gilt mit $h = (h_1, h_2)$:

$$f(x + h) - f(x) = f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) + f(x_1 + h_1, x_2) - f(x_1, x_2).$$

Mit Hilfe des Mittelwertsatzes der Differentialrechnung ergibt sich

$$\begin{aligned} f(x + h) - f(x) &= h_2 \partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) + h_1 \partial_1 f(x_1 + \theta_1 h_1, x_2) \\ &= h_2 (\partial_2 f(x_1, x_2) + \omega_2(h_1, h_2)) + h_1 (\partial_1 f(x_1, x_2) + \omega_1(h_1, h_2)) \end{aligned}$$

mit den Abkürzungen

$$\begin{aligned} \omega_1(h_1, h_2) &= \partial_1 f(x_1 + \theta_1 h_1, x_2) - \partial_1 f(x_1, x_2) \\ \omega_2(h_1, h_2) &= \partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) - \partial_2 f(x_1, x_2). \end{aligned}$$

Wegen der Stetigkeit der partiellen Ableitungen gilt

$$\lim_{h_1, h_2 \rightarrow 0} \omega_1(h_1, h_2) = \lim_{h_1, h_2 \rightarrow 0} \omega_2(h_1, h_2) = 0.$$

Also ist f differenzierbar mit der totalen Ableitung $Df(x) := \nabla f(x)$.

Q.E.D.

Bemerkung 3.2: Aufgrund der obigen Resultate gelten die folgenden Implikationen:

$$\text{stetig partiell differenzierbar} \Rightarrow (\text{total}) \text{ differenzierbar} \Rightarrow \text{partiell differenzierbar}.$$

Die umgekehrten Implikationen gelten i. Allg. nicht. Die erste Implikation erlaubt es, bei der k -maligen stetigen *partiellen* Differenzierbarkeit von Funktionen den Zusatz „partiell“ wegzulassen, da die Stetigkeit der k -ten partiellen Ableitungen die *totale* Differenzierbarkeit der $(k-1)$ -ten Ableitungen impliziert. Wir sprechen daher im Folgenden kurz von „ k -maliger stetiger Differenzierbarkeit“, wenn eine Funktion k -mal stetig partiell differenzierbar ist. Differentiale höherer Ordnung werden wir bei der Ableitung der n -dimensionalen Taylor-Formel kennenlernen.

Lemma 3.1 (Richtungsableitung): Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ im Punkt $x \in D$ differenzierbar. Dann existiert für jeden Vektor $v \in \mathbb{R}^n$ mit $\|v\|_2 = 1$ die Ableitung in Richtung v (sog. „Richtungsableitung“)

$$\frac{\partial f}{\partial v}(x) := \lim_{t \searrow 0} \frac{f(x + tv) - f(x)}{t}.$$

und lässt sich schreiben als

$$\partial_v f(x) := \frac{\partial f}{\partial v}(x) = \nabla f(x) \cdot v.$$

Beweis: Für $x \in D$ definieren wir die Funktion $\xi(t) := x + tv$. Für genügend kleines $\varepsilon > 0$ ist $\xi(t) \in D$ für $t \in [0, \varepsilon]$. Also ist die Komposition $h := f \circ \xi : [0, \varepsilon] \rightarrow \mathbb{R}$ definiert. Nach Definition der Richtungsableitung ist

$$\frac{\partial f}{\partial v}(x) = \lim_{t \searrow 0} \frac{f(x + tv) - f(x)}{t} = \left. \frac{d}{dt} f(x + tv) \right|_{t=0} = \left. \frac{dh}{dt}(0) \right|_{t=0} = h'(0).$$

Mit Hilfe der Kettenregel folgt

$$h'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\xi(t)) \xi'_i(t).$$

Beachtung von $\xi'_i(t) = v_i$ ergibt dann

$$h'(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) v_i = \nabla f \cdot v,$$

was zu zeigen war. Q.E.D.

Anwendung 3.1.1: Im Fall $\nabla f(x) \neq 0$ ist der Winkel θ zwischen den Vektoren $v \in \mathbb{R}^n$ mit $\|v\|_2 = 1$ und $\nabla f(x) \in \mathbb{R}^n$ definiert durch

$$\cos(\theta) = \frac{\nabla f(x) \cdot v}{\|\nabla f(x)\|_2 \|v\|_2}.$$

Damit finden wir

$$\frac{\partial f}{\partial v}(x) = \nabla f(x) \cdot v = \|\nabla f(x)\|_2 \|v\|_2 \cos(\theta) = \|\nabla f(x)\|_2 \cos(\theta).$$

Folglich ist die Richtungsableitung maximal, wenn v und $\nabla f(x)$ die gleiche Richtung haben, d. h.: *Der Vektor $\nabla f(x)$ gibt die Richtung des stärksten Anstiegs von f im Punkt x an.*

Beispiel 3.6: Zur Illustration von Satz 3.3 geben wir die folgenden Beispiele:

i) *Eine Funktion, für die alle Richtungsableitungen existieren, die aber dennoch nicht total differenzierbar ist:* Sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch

$$f(x) := \frac{x_1^2 x_2}{x_1^4 + x_2^2}, \quad x \neq (0, 0), \quad f(0) := 0.$$

Die Funktion f ist in $\mathbb{R}^2 \setminus \{0\}$ beliebig oft stetig differenzierbar. Sei $v = (v_1, v_2)$, ein beliebiger Richtungseinheitsvektor. Auf der Geraden $\{x = tv : t \in \mathbb{R}\}$ hat f die Werte

$$f(tv) = \frac{t^3 v_1^2 v_2}{t^4 v_1^4 + t^2 v_2^2} = \frac{t v_1^2 v_2}{t^2 v_1^4 + v_2^2}.$$

Die Ableitung von f im Punkt $x = 0$ in Richtung v existiert also im Fall $v_2 \neq 0$ und ist:

$$\partial_v f(0) := \lim_{t \searrow 0} \frac{f(tv) - f(0)}{t} = \lim_{t \searrow 0} \frac{v_1^2 v_2}{t^2 v_1^4 + v_2^2} = \frac{v_1^2}{v_2}.$$

Im Fall $v_2 = 0$ ist $\partial_v = \partial_2$. Auf den Koordinatenachsen ist aber $f(x_1, 0) \equiv f(0, x_2) \equiv 0$, so dass f in 0 partiell differenzierbar ist und zwar mit Gradient $\nabla f(0) = (0, 0)$. In diesem Fall gilt also die Formel

$$\partial_v f(0) = \nabla f(0) \cdot v$$

für die Berechnung der Richtungsableitungen differenzierbarer Funktionen nicht, so dass f in 0 nicht total differenzierbar sein kann. Dies liegt an der Unstetigkeit der Funktion f und ihrer partiellen Ableitungen in $x = 0$:

$$\lim_{\varepsilon \searrow 0} f(\varepsilon, \varepsilon^2) = \lim_{\varepsilon \searrow 0} \frac{\varepsilon^4}{2\varepsilon^4} = \frac{1}{2} \neq 0 = f(0, 0).$$

ii) *Eine stetige und partiell differenzierbare, aber nicht total differenzierbare Funktion:*

Sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch

$$f(x) := \frac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}}, \quad x \neq 0, \quad f(0) := 0.$$

Die Funktion f ist in $\mathbb{R}^2 \setminus \{0\}$ beliebig oft stetig differenzierbar. Für $x \rightarrow 0$ gilt:

$$|f(x)| = \frac{|x_1 x_2|}{\sqrt{x_1^2 + x_2^2}} \leq \frac{1}{2} \frac{x_1^2 + x_2^2}{\sqrt{x_1^2 + x_2^2}} \rightarrow 0;$$

also ist f stetig in 0 . Auf den Koordinatenachsen ist $f(x_1, 0) = 0$ und $f(0, x_2) = 0$, so dass f in 0 partiell differenzierbar ist mit den Ableitungen $\partial_x f(0) = \partial_y f(0) = 0$. Als Differential von f in 0 kommt daher nur der Gradient $\nabla f(0) = (0, 0)$ in Frage. Es gilt aber z. B. für $x_1 = x_2 \rightarrow 0$:

$$\frac{|f(x) - f(0) - \nabla f(0) \cdot (x_1, x_2)|}{\|x\|_2} = \frac{|x_1 x_2|}{x_1^2 + x_2^2} = \frac{x_1^2}{2x_1^2} \not\rightarrow 0,$$

d. h.: f ist in 0 nicht total differenzierbar.

Satz 3.4 (Kettenregel): Seien $D_f \subset \mathbb{R}^n$ und $D_g \subset \mathbb{R}^m$ offene Mengen und $g : D_g \rightarrow \mathbb{R}^n$ und $f : D_f \rightarrow \mathbb{R}^r$ Abbildungen. Ist die Abbildung g im Punkt $x \in D_g$ und die Abbildung f im Punkt $y = g(x) \in D_f$ differenzierbar, so ist die Komposition $h = f \circ g$ im Punkt x differenzierbar, und für die Differentiale gilt:

$$D_x h(x) = D_y f(g(x)) \cdot D_x g(x). \quad (3.1.9)$$

Dabei ist $D_x h(x) \in \mathbb{R}^{r \times m}$, $D_y f(y) \in \mathbb{R}^{r \times n}$, $D_x g(x) \in \mathbb{R}^{n \times m}$, und der Punkt „ \cdot “ steht für die entsprechende Matrix-Matrix-Multiplikation.

Beweis: Nach Voraussetzung ist mit $x \in D_g$ und $y = g(x) \in D_f$:

$$g(x+h) = g(x) + D_x g(x)h + \omega_g(h), \quad f(y+\eta) = f(y) + D_y f(y)\eta + \omega_f(\eta),$$

mit

$$\lim_{x+h \in D, \|h\|_2 \rightarrow 0} \frac{\|\omega_g(h)\|_2}{\|h\|_2} = 0, \quad \lim_{x+\eta \in D, \|\eta\|_2 \rightarrow 0} \frac{\|\omega_f(\eta)\|_2}{\|\eta\|_2} = 0.$$

Setzen wir $\eta := D_x g(x)h + \omega_g(h)$, so ergibt sich mit $y = g(x)$:

$$\begin{aligned} (f \circ g)(x+h) &= f(g(x+h)) = f(y+\eta) \\ &= f(y) + D_y f(y)\eta + \omega_f(\eta) \\ &= f(y) + D_y f(y) \cdot D_x g(x)h + D_y f(y)\omega_g(h) + \omega_f(D_x g(x)h + \omega_g(h)) \\ &= (f \circ g)(x) + D_y f(y) \cdot D_x g(x)h + \omega_{f \circ g}(h) \end{aligned}$$

mit

$$\omega_{f \circ g}(h) := D_y f(y)\omega_g(h) + \omega_f(D_x g(x)h + \omega_g(h)).$$

Wir haben zu zeigen, dass $\omega_{f \circ g}(h) = o(\|h\|)$. Mit $\omega_g(h) = o(\|h\|)$ ist auch $D_y f(y)\omega_g(h) = o(\|h\|)$. Ferner gilt mit einer Konstante $c > 0$

$$\|\omega_g(h)\|_2 \leq c\|h\|_2,$$

und wegen $\omega_g(\eta) = o(\|\eta\|)$ ist

$$\omega_f(\eta) = \|\eta\|_2 \tilde{\omega}_f(\eta), \quad \lim_{\eta \rightarrow 0} \tilde{\omega}_f(\eta) = 0.$$

Damit folgt

$$\begin{aligned} \|\omega_f(D_x g(x)h + \omega_g(h))\|_2 &\leq \|D_x g(x)h + \omega_g(h)\|_2 \|\tilde{\omega}_f(D_x g(x)h + \omega_g(h))\|_2 \\ &\leq (\|D_x g(x)\|_2 + c)\|h\|_2 \|\tilde{\omega}_f(D_x g(x)h + \omega_g(h))\|_2. \end{aligned}$$

Zusammen mit dem vorher Gezeigten folgt also

$$\frac{\|\omega_{f \circ g}(h)\|_2}{\|h\|_2} \rightarrow 0 \quad (h \rightarrow 0),$$

was den Beweis vervollständigt.

Q.E.D.

Bemerkung 3.3: Die Matrixidentität (3.1.9) lautet komponentenweise für $i = 1, \dots, m$ und $j = 1, \dots, r$:

$$\frac{\partial h_j}{\partial x_i}(x_1, \dots, x_m) = \sum_{k=1}^n \frac{\partial f_j}{\partial y_k}(g_1(x), \dots, g_n(x)) \frac{\partial g_k}{\partial x_i}(x_1, \dots, x_m). \quad (3.1.10)$$

Im Spezialfall $m = r = 1$, d. h. $g : D_g \subset \mathbb{R} \rightarrow \mathbb{R}^n$ und $f : D_f \subset \mathbb{R}^n \rightarrow \mathbb{R}$, ist also

$$h'(x) = \frac{d}{dx} f(g(x)) = \sum_{k=1}^n \frac{\partial}{\partial y_k} f(g_1(x), \dots, g_n(x)) \frac{d}{dx} g_k(x) = \nabla_y f(g(x)) \cdot g'(x). \quad (3.1.11)$$

3.1.3 Mittelwertsatz

Für differenzierbare Funktionen $f : [a, b] \rightarrow \mathbb{R}$ auf einem Intervall $[a, b] \subset \mathbb{R}$ gilt nach dem Fundamentalsatz die Beziehung

$$f(x+h) - f(x) = \int_0^1 \frac{d}{ds} f(x+sh) ds = \int_0^1 f'(x+sh)h ds = \left(\int_0^1 f'(x+sh) ds \right) h$$

und weiter nach dem 1. Mittelwertsatz der Differentialrechnung

$$f(x+h) - f(x) = f'(x+\tau h)h$$

mit einer Zwischenstelle $\tau \in (0, 1)$. Die Verallgemeinerung dieser Beziehungen für Funktionen $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ und Abbildungen $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ erfordert die Verwendung von vektor- bzw. matrix-wertigen Integralen. Für eine matrix-wertige, stetige Funktion $A = (a_{ij})_{i=1, j=1}^{m, n} : [a, b] \rightarrow \mathbb{R}^{m \times n}$ ist z. B.:

$$\int_a^b A(s) ds := \left(\int_a^b a_{ij}(s) ds \right)_{i, j=1}^{m, n}.$$

Satz 3.5 (Mittelwertsatz): Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ stetig differenzierbar. Ferner sei $x \in D$ und $h \in \mathbb{R}^n$, so dass $x + sh \in D$ für $0 \leq s \leq 1$. Dann gilt:

$$f(x+h) - f(x) = \left(\int_0^1 \nabla f(x+sh) ds \right) \cdot h. \quad (3.1.12)$$

Ist $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ stetig differenzierbar mit Jacobi-Matrix $J_f(x)$, so gilt

$$f(x+h) - f(x) = \left(\int_0^1 J_f(x+sh) ds \right) h. \quad (3.1.13)$$

Beweis: Wir betrachten gleich den allgemeinen Fall einer Abbildung $f : D \rightarrow \mathbb{R}^m$. Für die durch $g_j(s) := f_j(x+sh)$ definierten Funktionen $g_j : [0, 1] \rightarrow \mathbb{R}$ gilt nach der Kettenregel:

$$f_j(x+sh) - f_j(x) = g_j(1) - g_j(0) = \int_0^1 g_j'(s) ds = \int_0^1 \sum_{i=1}^n \partial_i f_j(x+sh) h_i ds.$$

Dies ist für $m = 1$ die Beziehung (3.1.12) und für $m \geq 2$ die Beziehung (3.1.13). Q.E.D.

Bemerkung 3.4: Für eine stetig differenzierbare Funktion $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ folgt aus (3.1.12) mit Hilfe des 1. Mittelwertsatzes der Integralrechnung die Beziehung

$$f(x+h) - f(x) = \int_0^1 \nabla f(x+sh) \cdot h ds = \nabla f(x+\tau h) \cdot h. \quad (3.1.14)$$

mit einem Zwischenwert $\tau \in (0, 1)$. Die mehrdimensionale Mittelwertaussage (3.1.13) hat keine analoge differentielle Form. Für $m > 1$ ist die naheliegend erscheinende Beziehung $f(x+h) - f(x) = J_f(x+\tau h)h$ i. Allg. nicht gültig, denn der skalare Parameterwert $\tau \in [0, 1]$ kann nicht für alle Komponenten f_i als derselbe angenommen werden.

Lemma 3.2: Für stetige vektorwertige und matrixwertige Funktionen $v : [a, b] \rightarrow \mathbb{R}^n$ bzw. $A : [a, b] \rightarrow \mathbb{R}^{m \times n}$ gilt:

$$\left\| \int_a^b v(s) ds \right\|_2 \leq \int_a^b \|v(s)\|_2 ds, \quad (3.1.15)$$

$$\left\| \int_a^b A(s) ds \right\|_2 \leq \int_a^b \|A(s)\|_2 ds. \quad (3.1.16)$$

Beweis: Mit der Setzung $w := \int_a^b v(s) ds$ erhalten wir

$$\left\| \int_a^b v(s) ds \right\|_2^2 = \int_a^b (v(s), w)_2 ds \leq \int_a^b \|v(s)\|_2 ds \|w\|_2.$$

Dies impliziert (3.1.15). Zum Beweis von (3.1.16) setzen wir $w := \int_a^b A(s) ds$ und verfahren analog wie zuvor. Q.E.D.

Korollar 3.1: Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}^m$ stetig differenzierbar. Ferner sei $x \in D$ mit $K_r(x) \subset D$ für ein $r > 0$. Dann gilt:

$$\|f(y) - f(x)\|_2 \leq M\|y - x\|_2, \quad y \in K_r(x), \quad (3.1.17)$$

mit $M := \sup_{z \in K_r(x)} \|J_f(z)\|_2$, d. h.: Die Abbildung f ist in D lokal Lipschitz-stetig. Insbesondere gilt für konvexes D :

$$\|f(x) - f(y)\|_2 \leq M\|x - y\|_2, \quad x, y \in D, \quad (3.1.18)$$

mit $M := \sup_{z \in D} \|J_f(z)\|_2$, d. h.: Die Abbildung f ist auf ganz D Lipschitz-stetig.

Beweis: Nach dem Mittelwertsatz 3.5 gilt mit $h := y - x$:

$$\|f(y) - f(x)\|_2 = \|f(x+h) - f(x)\|_2 = \left\| \int_0^1 J_f(x+sh)h \, ds \right\|_2,$$

und mit Lemma 3.2 folgt

$$\begin{aligned} \left\| \int_0^1 J_f(x+sh)h \, ds \right\|_2 &\leq \int_0^1 \|J_f(x+sh)h\|_2 \, ds \leq \int_0^1 \|J_f(x+sh)\|_2 \|h\|_2 \, ds \\ &\leq \sup_{0 < s < 1} \|J_f(x+sh)\|_2 \|h\|_2. \end{aligned}$$

Dies beweist die erste Abschätzung. der Beweis der zweiten sei als Übungsaufgabe gestellt. Q.E.D.

3.2 Taylor-Entwicklung und Extremwerte

Für eine $r+1$ -mal stetig differenzierbare Funktion $f : (a, b) \rightarrow \mathbb{R}$ besteht um jeden Punkt $x \in (a, b)$ die Taylor-Approximation

$$f(x+h) = \sum_{k=0}^r \frac{f^{(k)}(x)}{k!} h^k + R_{r+1}^f(x; h), \quad (3.2.19)$$

mit dem Restglied $R_{r+1}^f(x; h)$ in differentieller Form

$$R_{r+1}^f(x; h) = \frac{f^{(r+1)}(x + \theta h)}{(r+1)!} h^{r+1}, \quad \theta \in (0, 1),$$

oder in integraler Form

$$R_{r+1}^f(x; h) = \frac{h^{r+1}}{r!} \int_0^1 f^{(r+1)}(x+th)(1-t)^r \, dt.$$

Diese Restglieddarstellungen ergeben sich aus den in Abschnitt 5.3 des Bandes Analysis 1 hergeleiteten durch geeignete Variablensubstitution (Übungsaufgabe). Im Folgenden wollen wir dies für Funktionen in mehreren Variablen verallgemeinern. Als Anwendung gewinnen wir dann auch wieder notwendige und hinreichende Bedingungen für die Existenz lokaler Extrema.

3.2.1 Taylor-Entwicklung im \mathbb{R}^n

Definition 3.8 (Multiindex-Notation): Ein n -dimensionaler „Multiindex“ ist ein n -Tupel $\alpha = (\alpha_1, \dots, \alpha_n)$ mit Komponenten $\alpha_i \in \mathbb{N}_0$. Für Multiindizes sind eine „Ordnung“ $|\alpha|$ und die „Fakultät“ $\alpha!$ definiert durch

$$|\alpha| := \alpha_1 + \dots + \alpha_n, \quad \alpha! := \alpha_1! \cdot \dots \cdot \alpha_n!$$

Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ wird gesetzt:

$$x^\alpha := x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n}.$$

Für eine $|\alpha|$ -mal stetig differenzierbare Funktion wird gesetzt:

$$\partial^\alpha f := \partial_1^{\alpha_1} \cdot \dots \cdot \partial_n^{\alpha_n} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdot \dots \cdot \partial x_n^{\alpha_n}}.$$

Wegen der Stetigkeit der Ableitungen ist dieser Ausdruck unabhängig von der Reihenfolge der partiellen Ableitungen. Summen über multiindizierte Größen werden abgekürzt geschrieben in der Form

$$\sum_{|\alpha|=0}^r a_\alpha := \sum_{k=0}^r \sum_{\alpha \in \mathbb{N}_0^n, |\alpha|=k} a_\alpha.$$

Beispiel 3.7: Zur Illustration der Multiindex-Schreibweise betrachten wir den repräsentativen Fall $n = 3$. Dann sind die Multiindizes $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ der Ordnung $|\alpha| = 2$ gegeben durch

$$(2, 0, 0), (0, 2, 0), (0, 0, 2), (1, 1, 0), (1, 0, 1), (0, 1, 1).$$

Die Fakultäten dieser Multiindizes sind der Reihe nach ($0! := 1$) $\alpha! = 2, 2, 2, 1, 1, 1$. Die zugehörigen partiellen Ableitungen sind

$$\partial^\alpha f = \partial_1^2 f, \partial_2^2 f, \partial_3^2 f, \partial_1 \partial_2 f, \partial_1 \partial_3 f, \partial_2 \partial_3 f.$$

Schließlich ist

$$\sum_{|\alpha|=2} \partial^\alpha f = \partial_1^2 f + \partial_2^2 f + \partial_3^2 f + \partial_1 \partial_2 f + \partial_1 \partial_3 f + \partial_2 \partial_3 f.$$

Satz 3.6 (Taylor-Formel): Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine $(r+1)$ -mal stetig differenzierbare Funktion. Dann gilt für jeden Vektor $h \in \mathbb{R}^n$ mit $x + sh \in D$, $s \in [0, 1]$, die „Taylor-Formel“

$$f(x+h) = \sum_{|\alpha| \leq r} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + R_{r+1}^f(x; h), \quad (3.2.20)$$

mit dem Taylor-Restglied $R_{r+1}^f(x; h)$ in differentieller Form

$$R_{r+1}^f(x; h) = \sum_{|\alpha|=r+1} \frac{\partial^\alpha f(x + \theta h)}{\alpha!} h^\alpha, \quad \theta \in (0, 1),$$

oder in integraler Form

$$R_{r+1}^f(x; h) = (r+1) \int_0^1 \sum_{|\alpha|=r+1} \frac{\partial^\alpha f(x + th)}{\alpha!} h^\alpha (1-t)^r dt.$$

Beweis: i) Wir betrachten die durch $g(t) := f(x + th)$ definierte Funktion $g: [0, 1] \rightarrow \mathbb{R}$. Diese ist $(r+1)$ -mal stetig differenzierbar mit den k -ten Ableitungen

$$g^{(k)}(t) = \sum_{i_1, \dots, i_k=1}^n \partial_{i_k} \dots \partial_{i_1} f(x + th) h_{i_1} \dots h_{i_k}. \quad (3.2.21)$$

Wir zeigen dies durch Induktion nach k mit Hilfe der Kettenregel. Für $k = 1$ gilt zunächst:

$$g'(t) = \frac{d}{dt} f(x_1 + th_1, \dots, x_n + th_n) = \sum_{i=1}^n \partial_i f(x + th) h_i.$$

Sei die Behauptung als richtig angenommen für $k-1 \geq 1$. Dann gilt:

$$\begin{aligned} g^{(k)}(t) &= \frac{d}{dt} g^{(k-1)}(t) = \frac{d}{dt} \left(\sum_{i_1, \dots, i_{k-1}=1}^n \partial_{i_{k-1}} \dots \partial_{i_1} f(x + th) h_{i_1} \dots h_{i_{k-1}} \right) \\ &= \sum_{i=1}^n \partial_i \left(\sum_{i_1, \dots, i_{k-1}=1}^n \partial_{i_{k-1}} \dots \partial_{i_1} f(x + th) h_{i_1} \dots h_{i_{k-1}} \right) h_i \\ &= \sum_{i_1, \dots, i_k=1}^n \partial_{i_k} \dots \partial_{i_1} f(x + th) h_{i_1} \dots h_{i_k}. \end{aligned}$$

Kommt unter den Indizes i_1, \dots, i_k der Index $i \in \{1, \dots, n\}$ genau α_i -mal vor, so gilt wegen der Vertauschbarkeit der Ableitungen:

$$\partial_{i_k} \dots \partial_{i_1} f(x + th) h_{i_1} \dots h_{i_k} = \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f(x + th) h_1^{\alpha_1} \dots h_n^{\alpha_n}.$$

Die Anzahl der k -Tupel (i_1, \dots, i_k) von Zahlen $i_j \in \{1, \dots, n\}$, bei denen die Zahl $i \in \{1, \dots, n\}$ genau α_i -mal vorkommt mit $\alpha_1 + \dots + \alpha_n = k$ ist nach Lemma 3.3 (s. unten)

$$\frac{k!}{\alpha_1! \dots \alpha_n!}.$$

Dies ergibt

$$\begin{aligned} g^{(k)}(t) &= \sum_{i_1, \dots, i_k=1}^n \partial_{i_k} \dots \partial_{i_1} f(x+th) h_{i_1} \dots h_{i_k} \\ &= \sum_{|\alpha|=k} \frac{k!}{\alpha_1! \dots \alpha_n!} \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f(x+th) h_1^{\alpha_1} \dots h_n^{\alpha_n} \\ &= \sum_{|\alpha|=k} \frac{k!}{\alpha!} \partial^\alpha f(x+th) h^\alpha. \end{aligned}$$

ii) Als nächstes wenden wir die eindimensionale Taylor-Formel auf die Funktion $g(t)$ an. Es gibt ein $\theta \in [0, 1]$, so dass

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

Nach (i) ist

$$\frac{g^{(k)}(0)}{k!} = \sum_{|\alpha|=k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha$$

und

$$\begin{aligned} \frac{g^{(r+1)}(\theta)}{(r+1)!} &= \sum_{|\alpha|=r+1} \frac{\partial^\alpha f(x+\theta h)}{\alpha!} h^\alpha, \\ \frac{1}{r!} \int_0^1 g^{(r+1)}(t)(1-t)^r dt &= (r+1) \int_0^1 \sum_{|\alpha|=r+1} \frac{\partial^\alpha f(x+th)}{\alpha!} h^\alpha (1-t)^r dt. \end{aligned}$$

Dies impliziert die Taylor-Formel (3.2.20) mit den Restgliedern in differentieller und integraler Form. Q.E.D.

Lemma 3.3: Sei $\alpha = (\alpha_1, \dots, \alpha_n)$ mit $|\alpha| = k \geq 1$ gegeben. Dann ist die Anzahl $N_\alpha(k)$ der k -Tupel (i_1, \dots, i_k) von Zahlen $i_j \in \{1, \dots, n\}$, bei denen die Zahl $i \in \{1, \dots, n\}$ genau α_i -mal vorkommt, bestimmt durch

$$N_\alpha(k) = \frac{k!}{\alpha_1! \dots \alpha_n!}. \quad (3.2.22)$$

Beweis: Wir ordnen die Indizes in dem k -Tupel gemäß

$$(i_1, \dots, i_k) = (\underbrace{1, \dots, 1}_{\alpha_1\text{-mal}}, \underbrace{2, \dots, 2}_{\alpha_2\text{-mal}}, \dots, \underbrace{n, \dots, n}_{\alpha_n\text{-mal}}), \quad \alpha_1 + \dots + \alpha_n = k.$$

Die Anzahl der möglichen Permutationen der k Elemente des k -Tupels ist $k!$. Das k -Tupel bleibt aber unverändert, wenn bei einer Permutation die α_i Elemente i nur unter sich vertauscht werden. Die Anzahl dieser Permutationen ist $\alpha_1! \cdot \dots \cdot \alpha_n! = \alpha!$. Dabei wird im Fall $\alpha_i = 0$ die Konvention $0! := 1$ verwendet. Insgesamt gibt es also $N_\alpha(k) = k!/\alpha!$ verschiedene Permutationen des k -Tupels. Q.E.D.

Korollar 3.2: Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine $r + 1$ -mal stetig differenzierbare Funktion. Dann gilt für $x \in D$ und $h \in \mathbb{R}^n$ mit $x + sh \in D$, $s \in [0, 1]$:

$$f(x + h) = \sum_{|\alpha| \leq r+1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \omega_{r+1}(x; h), \quad (3.2.23)$$

mit Funktionen $\omega_{r+1}(x; \cdot)$ mit den Eigenschaften $\omega_{r+1}(x; 0) = 0$ und

$$\omega_{r+1}(x; h) = o(\|h\|_2^{r+1}).$$

Speziell im Fall $r = 0$ gilt mit dem Gradienten ∇f von f :

$$f(x + h) = f(x) + (\nabla f(x), h)_2 + \omega_1(x; h), \quad (3.2.24)$$

und im Fall $r = 1$ gilt weiter mit der Hesse-Matrix H_f von f :

$$f(x + h) = f(x) + (\nabla f(x), h)_2 + \frac{1}{2}(H_f(x)h, h)_2 + \omega_2(x; h). \quad (3.2.25)$$

Beweis: i) Unter Verwendung der Taylor-Formel (3.2.20) mit ihrem Restglied in differentieller Form ergibt sich

$$\begin{aligned} f(x + h) &= \sum_{|\alpha| \leq r} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=r+1} \frac{\partial^\alpha f(x + \theta h)}{\alpha!} h^\alpha \\ &= \sum_{|\alpha| \leq r+1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=r+1} r_\alpha(x; h) h^\alpha \end{aligned}$$

mit

$$r_\alpha(x; h) := \frac{\partial^\alpha f(x + \theta h) - D^\alpha f(x)}{\alpha!}.$$

Wegen der Stetigkeit von $D^\alpha f$ für $|\alpha| = r + 1$ gilt $\lim_{h \rightarrow 0} r_\alpha(x; h) = 0$. Setzen wir

$$\omega_{r+1}(x; h) := \sum_{|\alpha|=r+1} r_\alpha(x; h) h^\alpha,$$

so folgt wegen

$$\frac{|h^\alpha|}{\|h\|_2^{|\alpha|}} = \frac{|h_1^{\alpha_1} \cdots h_n^{\alpha_n}|}{\|h\|_2^{\alpha_1} \cdots \|h\|_2^{\alpha_n}} \leq 1, \quad |\alpha| = r + 1,$$

die postulierte Konvergenz

$$\lim_{h \rightarrow 0} \frac{\omega(h)}{\|h\|_2^{r+1}} = 0.$$

Dies vervollständigt den Beweis der Taylor-Formel (3.2.23).

ii) Für $r = 0$ gilt:

$$\begin{aligned} f(x + h) &= \sum_{|\alpha| \leq 1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \omega_1(x; h) = f(x) + \sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \omega_1(x; h) \\ &= f(x) + \sum_{i=1}^n \partial_i f(x) h_i + \omega_1(x; h) = f(x) + (\nabla f(x), h)_2 + \omega_1(x; h), \end{aligned}$$

und für $r = 1$ mit elementarer Rechnung:

$$\begin{aligned} f(x+h) &= \sum_{|\alpha| \leq 2} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \omega_2(x; h) = f(x) + \sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \omega_2(x; h) \\ &= f(x) + \sum_{i=1}^n \partial_i f(x) h_i + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(x) h_i h_j + \omega_2(x; h) \\ &= f(x) + (\nabla f(x), h)_2 + \frac{1}{2} (H_f(x)h, h)_2 + \omega_2(x; h). \end{aligned}$$

Dies vervollständigt den Beweis.

Q.E.D.

Definition 3.9: Für eine beliebig oft partiell differenzierbare Funktion $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ und einem Punkt $x \in D$ heißt die Reihe

$$T_\infty^f(x+h) = \sum_{|\alpha|=0}^{\infty} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha$$

die „Taylor-Reihe“ von f in x .

Korollar 3.3: Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine beliebig oft differenzierbare Funktion. Dann konvergiert die Taylor-Reihe von f und stellt f dar, wenn

$$R_{r+1}^f(x, h) \rightarrow 0 \quad (r \rightarrow \infty), \quad x \in D. \quad (3.2.26)$$

Hinreichend dafür ist, dass die partiellen Ableitungen von f gleichmäßig beschränkt sind:

$$\sup_{|\alpha| \geq 0} \sup_{x \in D} |\partial^\alpha f(x)| < \infty.$$

Beweis: Mit der differentiellen Darstellung des Taylor-Restglieds folgt

$$\begin{aligned} \|R_{r+1}^f(x; h)\|_\infty &\leq \sum_{|\alpha|=r+1} \frac{|\partial^\alpha f(x + \theta h)|}{\alpha!} \|h\|_\infty^{|\alpha|}, \quad \theta \in (0, 1), \\ &\leq M(f) \sum_{|\alpha|=r+1} \frac{1}{\alpha!} \|h\|_\infty^{|\alpha|} \rightarrow 0 \quad (r \rightarrow \infty). \end{aligned}$$

Man beachte, dass hier $\|\cdot\|_\infty$ je nach Zusammenhang die Supremumnorm auf $C_b(D)$ (Vektorraum der auf D stetigen und beschränkten Funktionen) oder die Maximumnorm auf \mathbb{R}^n bedeutet. Der detaillierte Beweis der letzten Konvergenzaussage wird als Übungsaufgabe gestellt.

Q.E.D.

Das folgende Lemma bietet ein handliches Kriterium zur expliziten Bestimmung von Taylor-Reihen.

Lemma 3.4: Wird eine beliebig oft stetig differenzierbare Funktion $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ in einer Umgebung $K_r(x) \subset D$ durch eine Reihe homogener Polynome $P_k(x; h) = \sum_{|\alpha|=k} a_\alpha(x) h^\alpha$ auf \mathbb{R}^n mit Grad $k \geq 0$ dargestellt,

$$f(x+h) = \sum_{k=0}^{\infty} P_k(x; h), \quad x+h \in K_r(x), \quad (3.2.27)$$

so ist dies die Taylor-Reihe von f in x .

Beweis: Wir betrachten die durch $g(t) := f(x+th)$ definierte Funktion $g : [0, 1] \rightarrow \mathbb{R}$. Aufgrund der Homogenität der P_k gilt:

$$g(t) = \sum_{k=0}^{\infty} P_k(x; th) = \sum_{k=0}^{\infty} t^k P_k(x; h).$$

Die Ableitungen von $g(t)$ erhält man durch gliedweise Differentiation dieser Potenzreihe innerhalb ihres Konvergenzbereichs. Auswertung der k -ten Ableitung bei $t = 0$ ergibt $g^{(k)}(0) = k! P_k(x; h)$ und somit unter Verwendung des Arguments im Beweis von Satz 3.6:

$$\sum_{k=0}^{\infty} P_k(x; h) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} = \sum_{k=0}^{\infty} \sum_{|\alpha|=k} \frac{k!}{\alpha! k!} D^\alpha f(x) h^\alpha = \sum_{|\alpha|=0}^{\infty} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha.$$

Q.E.D.

Beispiel 3.8: Für die auf $D = \{x \in \mathbb{R}^2 : x \neq 0\}$ durch

$$f(x_1, x_2) = \frac{1}{1 - x_1 - x_2}$$

definierte Funktion $f : D \rightarrow \mathbb{R}$ in $x = 0$ erhält man mit der geometrischen Reihe

$$\frac{1}{1 - (x_1 + x_2)} = \sum_{k=0}^{\infty} (x_1 + x_2)^k.$$

Diese Reihe konvergiert genau in dem Streifen $S = \{x \in \mathbb{R}^2 : |x_1 + x_2| < 1\}$. Nach Lemma 3.4 ist dies die Taylor-Reihe von f , wie man vom Eindimensionalen her vielleicht vermuten könnte.

3.2.2 Extremwertaufgaben

Definition 3.10: Eine Funktion $f : D \rightarrow \mathbb{R}$ hat in einem Punkt $x \in D \subset \mathbb{R}^n$ ein „lokales Extremum“, wenn auf einer Kugelumgebung $K_\delta(x) \subset \mathbb{R}^n$ gilt:

$$f(x) = \sup_{y \in K_\delta(x) \cap D} f(y) \quad \text{oder} \quad f(x) = \inf_{y \in K_\delta(x) \cap D} f(y).$$

Das Extremum heißt „strikt“, wenn es in $K_\delta(x) \cap D$ nur im Punkt x angenommen wird. Das Extremum heißt „global“, wenn gilt

$$f(x) = \sup_{y \in D} f(y) \quad \text{oder} \quad f(x) = \inf_{y \in D} f(y).$$

Satz 3.7 (Notwendige Extremalbedingung): Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ stetig differenzierbar. Hat f in einem Punkt $\hat{x} \in D$ ein lokales Extremum, so gilt

$$\nabla f(\hat{x}) = 0. \quad (3.2.28)$$

Beweis: Die Funktion $f : D \rightarrow \mathbb{R}$ habe in $x \in D$ ein lokales Extremum. Mit den kartesischen Einheitsvektoren $e^{(i)}$ im \mathbb{R}^n betrachten wir die Funktionen

$$g_i(t) := f(\hat{x} + te^{(i)}), \quad i = 1, \dots, n.$$

Dann ist g_i auf einem nichtleeren Intervall $(-\delta_i, \delta_i) \subset \mathbb{R}$ definiert und differenzierbar. Ferner hat g_i in $t = 0$ ein lokales Extremum. Folglich muss $g_i'(0) = 0$ sein. Dies ist wegen der (totalen) Differenzierbarkeit von f nach der Kettenregel gleichbedeutend mit

$$0 = g_i'(0) = \sum_{j=1}^n \partial_j f(\hat{x}) \delta_{ij} = \partial_i f(\hat{x}), \quad i = 1, \dots, n.$$

Q.E.D.

Satz 3.8 (Hinreichende Extremalbedingung): Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar, und es sei in einem Punkt $\hat{x} \in D$

$$\nabla f(\hat{x}) = 0. \quad (3.2.29)$$

Ist die (symmetrische) Hesse-Matrix $H_f(x)$ in \hat{x} positiv definit, so liegt in \hat{x} ein striktes lokales Minimum, ist sie negativ definit, so liegt in \hat{x} ein striktes lokales Maximum, und ist sie indefinit, d. h. hat sie sowohl positive als auch negative Eigenwerte, so kann in \hat{x} überhaupt kein lokales Extremum vorliegen.

Beweis: Nach Korollar 3.2 gilt mit der Hesse-Matrix $H_f(x)$:

$$f(x+h) = f(x) + (\nabla f(x), h)_2 + \frac{1}{2}(H_f(x)h, h)_2 + \omega_2(x; h),$$

wobei $\lim_{h \rightarrow 0, h \neq 0} \omega_2(h) / \|h\|_2^2 = 0$. Wegen $\nabla f(\hat{x}) = 0$ gilt also:

$$f(\hat{x}+h) - f(\hat{x}) = \frac{1}{2}(H_f(\hat{x})h, h)_2 + \omega_2(\hat{x}; h).$$

i) Ist $H_f(\hat{x})$ positiv definit, so gilt mit seinem kleinsten Eigenwert $\lambda > 0$:

$$(H_f(\hat{x})h, h)_2 \geq \lambda \|h\|_2^2, \quad h \in \mathbb{R}^n.$$

Folglich ist

$$f(\hat{x}+h) - f(\hat{x}) \geq \frac{1}{2}\lambda \|h\|_2^2 + \omega_2(\hat{x}; h).$$

Für genügend kleines $\|h\|_2 < \delta$, $h \neq 0$, ist nach Voraussetzung $|\omega_2(h)| < \frac{1}{2}\lambda \|h\|_2^2$ und somit

$$f(\hat{x}+h) - f(\hat{x}) > \frac{1}{2}\lambda \|h\|_2^2 - \frac{1}{2}\lambda \|h\|_2^2 = 0,$$

d. h. in \hat{x} liegt also ein striktes lokales Minimum vor.

ii) Ist $H_f(\hat{x})$ negativ definit, so sehen wir mit einer analogen Argumentation, dass in \hat{x} ein striktes lokales Maximum vorliegt.

iii) Ist dagegen $H_f(\hat{x})$ indefinit, so gilt mit den Eigenvektoren z_+ und z_- zu einem positiven Eigenwert λ_+ und einem negativen Eigenwert λ_- von $H_f(\hat{x})$:

$$(H_f(\hat{x})z_+, z_+)_2 = \lambda_+ \|z_+\|_2^2 > 0, \quad (H_f(\hat{x})z_-, z_-)_2 = \lambda_- \|z_-\|_2^2 < 0.$$

Für genügend kleines $t > 0$ gilt dann

$$f(\hat{x} + tz_+) - f(\hat{x}) > \frac{1}{2}\lambda_+ t^2 \|z_+\|_2^2 - \frac{1}{2}\lambda_+ t^2 \|z_+\|_2^2 = 0.$$

sowie

$$f(\hat{x} + tz_-) - f(\hat{x}) < \frac{1}{2}\lambda_- t^2 \|z_-\|_2^2 - \frac{1}{2}\lambda_- t^2 \|z_-\|_2^2 = 0.$$

Also liegt in \hat{x} weder ein lokales Maximum noch ein lokales Minimum vor. Q.E.D.

Beispiel 3.9: Wir geben drei einfache Beispiele im Spezialfall $n = 2$. Die Funktionen

$$f_1(x) = a + x_1^2 + x_2^2, \quad f_2(x) = a - x_1^2 - x_2^2, \quad f_3(x) = x_1^2 - x_2^2,$$

haben die Gradienten

$$\nabla f_1(x) = (2x_1, 2x_2), \quad \nabla f_2(x) = (-2x_1, -2x_2), \quad \nabla f_3(x) = (2x_1, -2x_2),$$

und folglich möglicherweise in $\hat{x} = 0$ Extrema. Die zugehörigen Hesse-Matrizen sind

$$H_{f_1}(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad H_{f_2}(x) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}, \quad H_{f_3}(x) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Da $H_{f_1}(0)$ positiv und $H_{f_2}(0)$ negativ definit sind, liegt in $\hat{x} = 0$ für f_1 ein striktes lokales Minimum und für f_2 ein striktes lokales Maximum vor. Dagegen ist $H_{f_3}(0)$ indefinit, so dass f_3 in $\hat{x} = 0$ einen sog. „Sattelpunkt“ hat.

Beispiel 3.10: Ist die Hesse-Matrix in einer Nullstelle des Gradienten nur semidefinit, so lassen sich keine allgemeine Aussagen über lokale Extrema machen. Dies zeigen die folgenden Beispiele auf dem \mathbb{R}^2 :

$$f_1(x) = x_1^2 + x_2^4, \quad f_2(x) = x_1^2, \quad f_3(x) = x_1^2 + x_2^3.$$

Für alle drei Funktionen ist $\nabla f_i(0) = 0$ und die Hesse-Matrizen

$$H_{f_i}(0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

sind offenbar positiv semidefinit. Die drei Funktionen zeigen aber in $\hat{x} = 0$ unterschiedliches Verhalten. Die Funktion f_1 hat dort ein striktes lokales Minimum; die Funktion f_2 ein (nicht striktes) lokales Minimum, und die Funktion f_3 hat überhaupt kein lokales Extremum (Sattelpunkt).

Bemerkung 3.5: Zur praktischen Berechnung einer Extremalstelle, etwa eines Minimums, \hat{x} einer differenzierbaren Funktion $f : D \rightarrow \mathbb{R}$ kann man das sog. „Gradientenverfahren“ verwenden. Dabei wird ausgenutzt, dass in einem Punkt $x \in D$ die Funktion f den steilsten Abstieg in Richtung des negativen Gradienten $-\nabla f(x)$ hat. Ausgehend von einem Startpunkt $x^{(0)} \in D$ werden danach Iterierte $x^{(k)}$ erzeugt aus der Vorschrift

$$x^{(k)} = x^{(k-1)} - \lambda_k \nabla f(x^{(k-1)}), \quad (3.2.30)$$

wobei der Schrittweite $\lambda_k > 0$ aus der folgenden eindimensionalen Optimierungsbedingung (sog. „line search“) bestimmt wird:

$$f(x^{(k-1)} - \lambda_k \nabla f(x^{(k-1)})) = \min_{\lambda > 0} f(x^{(k-1)} - \lambda \nabla f(x^{(k-1)})).$$

Im Falle $\lambda_k \geq \lambda_* > 0$ und einer konvergenten Folge $(x^{(k)})_{k \in \mathbb{N}}$ gilt für deren Limes \hat{x} dann $\nabla f(\hat{x}) = 0$, d. h.: \hat{x} ist ein möglicher Extremalpunkt. Kriterien für die tatsächliche Konvergenz des Gradientenverfahrens werden in Texten zur Numerik abgeleitet.

3.2.3 Das Newton-Verfahren im \mathbb{R}^n

Als Anwendung der bisher bereitgestellten Begriffe und Resultate betrachten wir das sog. „Newton-Verfahren“ zur Lösung nicht linearer Gleichungssysteme im \mathbb{K}^n ,

$$f(x) = 0 \quad (3.2.31)$$

mit stetig differenzierbaren Abbildungen $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$. In Anlehnung an die Newton-Iteration im \mathbb{R}^1 ,

$$x_k = x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})},$$

welche geometrisch motiviert ist, wird die Newton-Iteration im \mathbb{R}^n wie folgt angesetzt:

$$x^{(k)} = x^{(k-1)} - J_f(x^{(k-1)})^{-1} f(x^{(k-1)}), \quad k \in \mathbb{N}, \quad (3.2.32)$$

mit der Jacobi-Matrix $J_f(\cdot)$ von f . In jedem Iterationsschritt ergibt sich ein lineares $(n \times n)$ -Gleichungssystem mit $J_f(x^{(t)})$ als Koeffizientenmatrix. Dies macht das Newton-Verfahren wesentlich aufwendiger als die einfache Fixpunktiteration; dafür konvergiert es aber in der Regel auch sehr viel schneller.

Zum Nachweis der Konvergenz des Newton-Verfahrens nehmen wir an, dass die Abbildung f auf einer offenen (konvexen) Teilmenge $D \subset \mathbb{R}^n$ definiert ist, und dort stetige und beschränkte partielle Ableitungen bis zur zweiten Ordnung besitzt. Ferner wird der Einfachheit halber angenommen, dass eine Nullstelle z von f in D existiert mit der Eigenschaft

$$f(z) = 0, \quad |J_f(z)| \neq 0, \quad (3.2.33)$$

d. h.: $J_f(z)$ ist regulär. Wir verwenden wieder die Taylor-Formel

$$f(y) = f(x) + J_f(x)(y - x) + R(x; y), \quad x, y, \in G, \quad (3.2.34)$$

mit dem Restglied $R(x; y) = (R_j(x; y))_{j=1, \dots, n}$, wobei

$$R_j(x; y) = \sum_{k, l=1}^n (y_k - x_k)(y_l - x_l) \int_0^1 \frac{\partial^2 f_j}{\partial x_k \partial x_l} (x + s(y - x))(1 - s) ds.$$

Aufgrund der Voraussetzungen an f gilt

$$\|R(x; y)\|_\infty \leq \frac{M}{2} \|y - x\|_\infty^2, \quad M := \max_{1 \leq j \leq n} \sup_{\zeta \in G} \sum_{k, l=1}^n \left| \frac{\partial^2 f_j}{\partial x_k \partial x_l}(\zeta) \right|. \quad (3.2.35)$$

Für jedes x aus einer Umgebung $K_r(z) = \{x \in \mathbb{R}^n : \|x - z\|_2 \leq r\} \subset D$ der Nullstelle z gilt aufgrund der Taylor-Formel

$$J_f(x) = J_f(z) + S(z; x) = J_f(z) \{I - J_f(z)^{-1} S(z; x)\},$$

mit dem Restglied $S(x; y) = (S_{jk}(x; y))_{j,k=1, \dots, n}$, wobei

$$S_{jk}(z; x) := \sum_{l=1}^n (x_l - z_l) \int_0^1 \frac{\partial^2 f_j}{\partial x_k \partial x_l} (z + s(x - z)) ds.$$

Mit der oben definierten Konstante M gilt

$$\|S(z; x)\|_\infty \leq \max_{1 \leq j \leq n} \sup_{\zeta \in G} \sum_{k, l=1}^n \left| \frac{\partial^2 f_j}{\partial x_k \partial x_l}(\zeta) \right| \|x - z\|_\infty \leq Mr.$$

Also ist für $x \in K_r(z)$:

$$\|J_f(z)^{-1} S(z; x)\|_\infty \leq \|J_f(z)^{-1}\|_\infty Mr.$$

Für hinreichend kleine Wahl von r ,

$$r < \frac{1}{\|J_f(z)^{-1}\|_\infty M}, \quad (3.2.36)$$

existiert dann nach Lemma 1.16 die Inverse $(I - J_f(z)^{-1} S(z; x))^{-1}$ und genügt der Abschätzung

$$\|[I - J_f(z)^{-1} S(z; x)]^{-1}\|_\infty \leq \frac{1}{1 - \|J_f(z)^{-1}\|_\infty Mr}.$$

Also existiert $J_f(x)^{-1} = [I - J_f(z)^{-1} S(z; x)]^{-1} J_f(z)^{-1}$ für alle $x \in K_r(z)$, r wie oben gewählt, und es gilt

$$\frac{1}{m} := \sup_{x \in K_r(z)} \|J_f(x)^{-1}\|_\infty \leq \frac{\|J_f(z)^{-1}\|_\infty}{1 - \|J_f(z)^{-1}\|_\infty Mr}.$$

Satz 3.9 (Newton-Verfahren): *Es seien die obigen Voraussetzungen erfüllt und r gemäß (3.2.36) bestimmt. Ferner sei $\rho \in (0, r]$ so gewählt, dass*

$$q := \frac{M}{2m} \rho < 1. \quad (3.2.37)$$

Dann sind für jeden Startwert $x^{(0)} \in K_\rho(z)$ die Newton-Iterierten $x^{(k)} \in K_\rho(z)$ wohl definiert und konvergieren gegen die Nullstelle z von f . Dabei gilt die Fehlerabschätzung

$$\|x^{(k)} - z\|_\infty \leq \frac{2m}{M} q^{(2^k)}, \quad k \in \mathbb{N}, \quad (3.2.38)$$

d. h.; Das Newton-Verfahren konvergiert „quadratisch“ (im Gegensatz zur nur „linear“ konvergierenden einfachen Fixpunktiteration).

Beweis: i) Für jede Iterierte $x^{(k)} \in K_r(z)$ existiert $J_f(x^{(k)})^{-1}$ so dass auch $x^{(k+1)}$ definiert ist. Mit Hilfe der Taylor-Formel (3.2.34) für $x = x^{(k)}$ und $y = z$ sieht man

$$x^{(k+1)} - z = -J_f(x^{(k)})^{-1} \{f(x^{(k)}) + J_f(x^{(k)})(z - x^{(k)})\} = J_f(x^{(k)})^{-1} R(x^{(k)}; z)$$

und somit

$$\|x^{(k+1)} - z\|_\infty \leq \frac{M}{2m} \|x^{(k)} - z\|_\infty^2.$$

Im Falle $x^{(k)} \in K_\rho(z)$ folgt

$$\|x^{(k+1)} - z\|_\infty \leq \frac{M}{2m} \rho^2 \leq \rho,$$

d. h.: $x^{(k+1)} \in K_\rho(z)$.

ii) Mit der Abkürzung $\rho_k := \frac{M}{2m} \|x^{(k)} - z\|_\infty$ gilt wieder

$$\rho_k \leq \rho_{k-1}^2 \leq \dots \leq \rho_0^{(2^k)}.$$

Die Abschätzung

$$\rho_0 = \frac{M}{2m} \|x^{(0)} - z\|_\infty \leq \frac{M}{2m} \rho = q$$

ergibt dann die gewünschte Abschätzung sowie die Konvergenz $x^{(k)} \rightarrow z$ ($k \rightarrow \infty$).
Q.E.D.

Bemerkung 3.6: Bei der Durchführung des Newton-Verfahrens zur Lösung nichtlinearer Gleichungssysteme ist die Hauptschwierigkeit die Konstruktion eines „guten“ Startpunktes $x^{(0)}$. Zur Vergrößerung des Konvergenzbereiches des Newton-Verfahrens führt man eine „Dämpfung“ ein,

$$x^{(k+1)} = x^{(k)} - \lambda_k J_f(x^{(k)})^{-1} f(x^{(k)}), \quad (3.2.39)$$

wobei der Parameter $\lambda_k \in (0, 1]$ zu Beginn klein gewählt wird und dann nach endlich vielen Schritten gemäß einer geeigneten Dämpfungsstrategie $\lambda_k = 1$ gesetzt wird (\Rightarrow Numerische Mathematik).

Beispiel 3.11: Zur Bestimmung der Inversen $Z = A^{-1}$ einer regulären Matrix $A \in \mathbb{R}^{n \times n}$ wird gesetzt

$$f(X) := X^{-1} - A,$$

für $X \in \mathbb{R}^{n \times n}$ regulär. Eine Nullstelle dieser Abbildung $f(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ ist gerade die Inverse $Z = A^{-1}$. Diese soll mit dem Newton-Verfahren berechnet werden. Dazu ist zunächst eine Umgebung von A bzw. von A^{-1} zu bestimmen, auf der $f(\cdot)$ definiert und differenzierbar ist. Für $X \in K_\rho(A)$ mit $\rho < \|A^{-1}\|_2^{-1}$ folgt aus

$$X = A - A + X = A(I - A^{-1}(A - X))$$

die Beziehung

$$\|A^{-1}(A - X)\|_2 \leq \|A^{-1}\|_2 \|A - X\|_2 \leq \rho \|A^{-1}\|_2 < 1,$$

d. h.: $I - A^{-1}(A - X)$ und damit auch X sind regulär. Als nächstes ist die Jacobi-Matrix $f'(\cdot)$ von $f(\cdot)$ als Abbildung von $\mathbb{R}^{n \times n}$ in sich zu bestimmen. Für die Durchführung des Newton-Verfahrens genügt es offensichtlich, die Wirkung von $f'(\cdot)$ auf Matrizen $Y \in \mathbb{R}^{n \times n}$ zu bestimmen. Wir wollen zeigen, dass

$$J_f(X)Y = -X^{-1}YX^{-1}, \quad Y \in \mathbb{R}^{n \times n}.$$

Dies sieht man wie folgt: Aus $f(X) = X^{-1} - A$ folgt $Xf(X) = I - XA$. Für die Jacobi-Matrizen der rechten und linken Seite gilt

$$\begin{aligned} ([Xf(X)]'Y)_{j,k} &= \sum_{pq} \frac{\partial}{\partial x_{pq}} \sum_l x_{jl} f_{lk}(X) y_{pq} \\ &= \sum_{p,q} \sum_l \left\{ \underbrace{\frac{\partial x_{jl}}{\partial x_{pq}}}_{\delta_{jp} \cdot \delta_{lq}} f_{lk}(X) + x_{jl} \frac{\partial f_{lk}}{\partial x_{pq}}(X) \right\} y_{pq} \\ &= \sum_q f_{qk}(X) y_{jq} + \sum_{p,q} \sum_l x_{jl} \frac{\partial f_{lk}}{\partial x_{pq}}(X) y_{pq} = (Yf(X) + Xf'(X)Y)_{jk} \end{aligned}$$

und analog

$$[I - XA]'Y = -YA.$$

Also ist

$$-YA = Yf(X) + XJ_f(X)Y = YX^{-1} - YA - XJ_f(X)Y$$

bzw.

$$J_f(X)Y = -X^{-1}YX^{-1}.$$

Das Newton-Verfahren

$$J_f(X^{(t)})X^{(t+1)} = J_f(X^{(t)})X^{(t)} - f(X^{(t)})$$

erhält in diesem Fall also die Gestalt

$$-X^{(t)-1}X^{(t+1)}X^{(t)-1} = -X^{(t)-1} \underbrace{X^{(t)}X^{(t)-1}}_{=I} - X^{(t)-1} + A$$

bzw.

$$X^{(t+1)} = 2X^{(t)} - X^{(t)}AX^{(t)} = X^{(t)}\{2I - AX^{(t)}\}.$$

Diese Iteration ist das mehrdimensionale Analogon der Iteration $x_{t+1} = x_t(2 - ax_t)$ im skalaren Fall zur divisionsfreien Berechnung des Kehrwertes $1/a$ einer Zahl $a \neq 0$. Über die Identität

$$X^{(t+1)} - Z = 2X^{(t)} - X^{(t)}AX^{(t)} - Z = -(X^{(t)} - Z)A(X^{(t)} - Z)$$

gewinnt man die Fehlerabschätzung

$$\|X^{(t+1)} - Z\|_2 \leq \|A\|_2 \|X^{(t)} - Z\|_2^2.$$

Der Einzugsbereich der quadratischen Konvergenz für das Newton-Verfahren ist in diesem Fall also die Menge

$$\{X \in \mathbb{R}^{n \times n} \mid \|X - Z\|_2 < \|A\|_2^{-1}\}.$$

3.3 Implizite Funktionen und Umkehrabbildung

Häufig sind Funktionen nicht in der *expliziten* Form $y = f(x)$ gegeben sondern *implizit* durch eine Gleichung der Form $F(x, y) = 0$. Als illustrierendes Beispiel betrachten wir die Gleichung

$$F(x, y) := x^2 + y^2 - 1 = 0,$$

welche die Punkte des Einheitskreises $\partial K_1(0)$ in der Ebene \mathbb{R}^2 charakterisiert. Diese Gleichung lässt sich formal nach y auflösen:

$$y = f_{\pm}(x) := \pm\sqrt{1 - x^2}.$$

Offenbar gibt es in keiner ganzen Umgebung von $x = \pm 1 \in \mathbb{R}^1$ eine (reelle) Lösung $y = f(x)$. Anhand dieses Beispiels sehen wir, dass die Funktion $y = f(x)$ nicht überall zu existieren und, wenn sie existiert, nicht global eindeutig bestimmt zu sein braucht.

Die allgemeine Fragestellung lautet also wie folgt: Sei $D = D^x \times D^y$ eine offene Menge im Produktraum $\mathbb{R}^n \times \mathbb{R}^m$ und $F : D \rightarrow \mathbb{R}$ eine stetige Funktion. Wir wollen untersuchen, inwieweit durch die Gleichung

$$F(x, y) = 0, \quad x \in D^x, \quad y \in D^y,$$

implizit eine Funktion $f : D^x \rightarrow D^y$ definiert ist, so dass

$$F(x, f(x)) = 0, \quad x \in D^x.$$

Ein wichtiger Spezialfall dieser Situation ist die Gleichung

$$F(x, y) = g(y) - x = 0.$$

In diesem Fall bedeutet die Auflösung nach der Variable y gerade die Bestimmung der „Umkehrabbildung“ $y = g^{-1}(x)$.

3.3.1 Implizite Funktionen

Satz 3.10 (Implizite Funktionen): Seien $D^x \in \mathbb{R}^n$ und $D^y \in \mathbb{R}^m$ offene Mengen und $F : D^x \times D^y \rightarrow \mathbb{R}^m$ eine stetig differenzierbare Abbildung. Ferner sei $(\hat{x}, \hat{y}) \in D^x \times D^y$ ein Punkt, in dem

$$F(\hat{x}, \hat{y}) = 0 \quad (3.3.40)$$

gilt und die Jacobi-Matrix $D_y F(\hat{x}, \hat{y}) \in \mathbb{R}^{m \times m}$ regulär ist.

i) Dann gibt es eine offene Umgebung $U(\hat{x}) \times U(\hat{y}) \subset D^x \times D^y$ und eine stetige Funktion $f : U(\hat{x}) \rightarrow U(\hat{y})$, so dass

$$F(x, f(x)) = 0, \quad x \in U(\hat{x}). \quad (3.3.41)$$

ii) Die Funktion f ist eindeutig bestimmt, d. h.: Ist $(x, y) \in U(\hat{x}) \times U(\hat{y})$ ein Punkt mit $F(x, y) = 0$, so ist $y = f(x)$.

iii) Die Funktion f ist im Punkt \hat{x} stetig differenzierbar, und ihre Jacobi-Matrix $J_f(\hat{x}) = D_x f(\hat{x}) \in \mathbb{R}^{m \times n}$ ist gegeben durch

$$J_f(\hat{x}) = -D_y F(\hat{x}, \hat{y})^{-1} D_x F(\hat{x}, \hat{y}). \quad (3.3.42)$$

Beweis: Wir führen den Beweis in mehreren Schritten.

ia) O.B.d.A. sei $(\hat{x}, \hat{y}) = (0, 0)$. Die Matrix $J_y := D_y F(0, 0)$ ist gemäß Voraussetzung regulär, so dass durch

$$G(x, y) := y - J_y^{-1} F(x, y)$$

eine stetig differenzierbare Abbildung $G : D^x \times D^y \rightarrow \mathbb{R}^m$ definiert ist. Offenbar ist $G(0, 0) = 0$, und es gilt:

$$F(x, y) = 0 \quad \Leftrightarrow \quad G(x, y) = y.$$

Die Jacobi-Matrix von G bzgl. der Variable y (I die Einheitsmatrix von $\mathbb{R}^{m \times m}$) erfüllt

$$D_y G(0, 0) = I - J_y^{-1} D_y F(0, 0) = 0.$$

Wegen der vorausgesetzten stetigen Differenzierbarkeit von f ist die Matrix-Funktion $D_y G(x, y)$ stetig. Folglich gibt es eine Kugelumgebung $K_r^x(0) \times K_r^y(0) \subset D^x \times D^y$ mit Radius $r > 0$, so daß

$$\|D_y G(x, y)\|_2 \leq \frac{1}{2}, \quad (x, y) \in K_r^x(0) \times K_r^y(0). \quad (3.3.43)$$

Wegen $G(0, 0) = 0$ gibt es ferner eine weitere offene Kugelumgebung $K_s^x(0) \subset K_r^x(0)$ mit Radius $0 < s \leq r$, so dass

$$\|G(x, 0)\|_2 < \frac{1}{2}r, \quad x \in K_s^x(0). \quad (3.3.44)$$

ib) Wir wollen nun eine stetige Funktion $f : K_s^x(0) \rightarrow K_r^y(0)$ mit $G(x, f(x)) = f(x)$ bzw. $F(x, f(x)) = 0$ konstruieren. Dazu betrachten wir für beliebiges $x \in K_s^x(0)$ die Fixpunktgleichung

$$G(x, y) = y. \quad (3.3.45)$$

Für Punkte $(x, y_1), (x, y_2) \in K_s^x(0) \times K_r^y(0)$ folgt mit dem Mittelwertsatz

$$\|G(x, y_1) - G(x, y_2)\|_2 \leq \sup_{(x, y) \in K_s^x(0) \times K_r^y(0)} \|D_y G(x, y)\|_2 \|y_1 - y_2\|_2 \leq \frac{1}{2} \|y_1 - y_2\|_2.$$

Weiter gilt dann für $y \in K_r^y(0)$

$$\|G(x, y)\|_2 \leq \|G(x, y) - G(x, 0)\|_2 + \|G(x, 0)\|_2 \leq \frac{1}{2} \|y\|_2 + \frac{1}{2} r < r,$$

d. h.: $G(x, \cdot)$ ist eine Selbstabbildung der abgeschlossenen Kugel $K_r^y(0)$ und außerdem eine Kontraktion mit Lipschitz-Konstante $q = \frac{1}{2}$. Nach dem Banachschen Fixpunktsatz gibt es somit zu jedem $x \in K_s^x(0)$ genau einen Fixpunkt $y(x) \in K_r^y(0)$ von $G(x, \cdot)$. Dieser wird ausgehend von dem Startpunkt $y^{(0)}(x) := 0$ als Limes der Iteration

$$y^{(k)}(x) = G(x, y^{(k-1)}(x)), \quad k \in \mathbb{N},$$

gewonnen. Dabei gilt die Fehlerabschätzung

$$\|y(x) - y^{(k)}(x)\|_2 \leq 2^{-k} \|y^{(1)} - y^{(0)}\|_2 = 2^{-k} \|G(x, 0)\|_2 \leq 2^{-k-1} r, \quad x \in K_s^x(0). \quad (3.3.46)$$

Mit der Beziehung

$$y^{(k)}(x) = G(x, y^{(k-1)}(x)) = y^{(k-1)}(x) - J_y^{-1} F(x, y^{(k-1)}(x))$$

und der Stetigkeit von $F(x, y)$ erschließen wir induktiv, dass die $y^{(k)}$ stetige Funktionen von $x \in K_s^x(0)$ sind. Durch $f(x) := y(x)$ erhalten wir eine Funktion $f : K_s^x(0) \rightarrow K_r^y(0)$, für die nach Konstruktion gilt:

$$G(x, f(x)) = f(x), \quad x \in K_s^x(0).$$

Die Abschätzung (3.3.46) kann so interpretiert werden, dass die stetigen Funktionen $y^{(k)}$ auf $K_s^x(0)$ gleichmäßig gegen die Funktion f konvergieren, so dass letztere ebenfalls stetig ist. Die Richtigkeit der ersten Behauptung für den Fall $(\hat{x}, \hat{y}) = (0, 0)$ folgt also mit den offenen Umgebungen $U(\hat{x}) := K_s^x(0)$ und $U(\hat{y}) := K_r^y(0)$.

ii) Die Eindeutige Bestimmtheit der Funktion $y = f(x)$ folgt aus der Tatsache, dass für $x \in K_s(\hat{x})$ der Fixpunkt $y = f(x)$ der Gleichung $G(x, y) = y$ eindeutig bestimmt ist.

iii) Aus der Definition der Differenzierbarkeit von $F(\cdot, \cdot)$ in $(0, 0)$ folgt $(J_y := D_y F(0, 0))$

$$F(x, y) = D_x F(0, 0)x + J_y y + \omega(x, y)$$

mit einer Funktion $\omega : K_r^x(0) \times K_r^y(0) \rightarrow \mathbb{R}^m$ mit der Eigenschaft $\|\omega(x, y)\|_2 = o(\|(x, y)\|_2)$. Nach dem eben Gezeigten gilt $F(x, f(x)) = 0$ auf $K_s^x(0)$ und folglich

$$f(x) = -J_y^{-1} D_x F(0, 0)x - J_y^{-1} \omega(x, f(x)).$$

Setzen wir $\psi(x) := -J_y^{-1}\omega(x, f(x))$, so ergibt sich

$$f(x) = -J_y^{-1}D_x F(0, 0)x + \psi(x). \quad (3.3.47)$$

Gemäß der Definition der totalen Differenzierbarkeit bedeutet dies, dass f in $x = 0$ differenzierbar ist, wenn wir zeigen können, dass $\psi(x) = o(\|x\|_2)$, d. h.:

$$\lim_{x \rightarrow 0} \frac{\psi(x)}{\|x\|_2} = 0.$$

Um dies zu zeigen, verwenden wir

$$\|\psi(x)\|_2 \leq \|J_y^{-1}\|_2 \|\omega(x, f(x))\|_2 = o(\|(x, f(x))\|_2).$$

Die Behauptung ist also bewiesen, wenn wir zeigen können, dass

$$\|f(x)\|_2 \leq \gamma \|x\|_2, \quad x \in K_\delta(0),$$

für hinreichend kleines $\delta > 0$. Dazu setzen wir

$$c_1 := \|J_y^{-1}D_x F(0, 0)\|_2, \quad c_2 := \|J_y^{-1}\|_2.$$

Es gibt Konstanten $\delta_i \in (0, r)$, so dass aus $\|x\| \leq \delta_1$ und $\|y\| \leq \delta_2$ folgt

$$\|\omega(x, y)\|_2 \leq \frac{1}{2c_2} \|(x, y)\|_2 \leq \frac{1}{2c_2} (\|x\|_2 + \|y\|_2).$$

Wegen der Stetigkeit von f gibt es ein $\delta \in (0, \delta_1]$, so dass für $\|x\|_2 \leq \delta$ gilt:

$$\|f(x)\|_2 \leq \delta_2.$$

Deshalb ist für $\|x\|_2 \leq \delta$

$$\|\omega(x, f(x))\|_2 \leq \frac{1}{2c_2} (\|x\|_2 + \|f(x)\|_2).$$

Die Gleichung (3.3.47) liefert nun

$$\|f(x)\|_2 \leq c_1 \|x\|_2 + c_2 \|\omega(x, f(x))\|_2 \leq (c_1 + \frac{1}{2}) \|x\|_2 + \frac{1}{2} \|f(x)\|_2.$$

Daraus folgt $\|f(x)\|_2 \leq (2c_1 + 1) \|x\|_2$, d. h. die gewünschte Abschätzung mit $\gamma := 2c_1 + 1$.

iiib) Es bleibt, die Existenz der Ableitung $D_x f(x)$ in einer Umgebung $K_\delta^x(0) \subset K_s^x(0)$ und ihre Stetigkeit in $x = 0$ zu zeigen. Dies ergibt sich aus den vorausgesetzten Eigenschaften der Abbildung $F(x, y)$ und dem bereits bekannten Störungssatz für reguläre Matrizen. Für hinreichend kleines $\delta > 0$ gilt für Punkte $(x, y) \in K_\delta^x(0) \times K_\delta^y(0) \subset K_s^x(0) \times K_r^y(0)$

$$\|D_y F(x, y) - D_y F(0, 0)\|_2 < \frac{1}{\|D_y F(0, 0)^{-1}\|_2}.$$

Nach Lemma 1.16 und dem nachfolgenden Korollar ist dann auch die Jacobi-Matrix $D_y F(x, y)$ regulär bzw. ihre Determinante $\det D_y F(x, y) \neq 0$. Die Elemente der inversen Matrix $D_y F(x, y)^{-1}$ sind daher nach der Cramerschen Regel stetige Funktionen der Elemente von $D_y F(x, y)$. Also konvergiert für $\|(x, y)\|_2 \rightarrow 0$:

$$\|D_y F(x, y)^{-1} D_x F(x, y) - D_y F(0, 0)^{-1} D_x F(0, 0)\|_2 \rightarrow 0.$$

Mit dem obigen Argument, diesmal angewendet für Punkte $(x, y) \in K_\delta^x(0) \times K_\delta^y(0)$ anstelle von $(0, 0)$, ergibt sich die Beziehung

$$f(x+h) - f(x) = -D_y F(x, y)^{-1} D_x F(x, y) h + \omega(x; h)$$

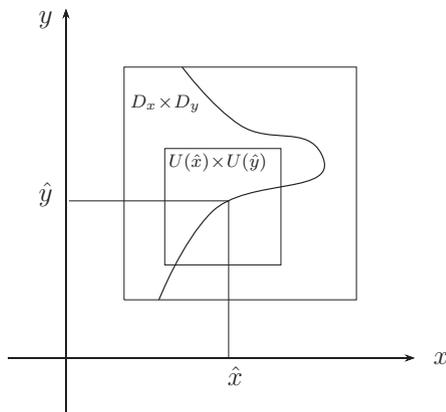
mit einem Restglied $\omega(x; h) = o(\|h\|_2)$. Die Ableitung

$$D_x f(x) = -D_y F(x, f(x))^{-1} D_x F(x, f(x))$$

von f ist also stetig in $x = 0$.

Q.E.D.

Bemerkung 3.7: Die Funktion $f(x)$ entsteht quasi durch „Auflösung“ der Gleichung $F(x, y) = 0$ nach y . Für die Gültigkeit von Satz 3.10 ist wesentlich, dass die Umgebung $D^x \times D^y$ verkleinert wird; in ganz $D^x \times D^y$ könnte es zu einem gegebenen x mehrere y -Werte (oder auch gar keine) geben, die der Gleichung $F(x, y) = 0$ genügen (s. Abbildung).



Beispiel 3.12: 1) Wir betrachten wieder die Gleichung

$$F(x, y) := x^2 + y^2 - 1 = 0$$

vom Anfang dieses Kapitels. Es ist $D_y F(x, y) = 2y$, so dass sich nach Satz 3.10 die Gleichung in der Umgebung eines jeden Punktes (\hat{x}, \hat{y}) mit $\hat{x}^2 + \hat{y}^2 - 1 = 0$ und $\hat{y} \neq 0$ (d. h.: $\hat{x} \neq \pm 1$) eindeutig durch $y = \sqrt{1 - x^2}$ oder $y = -\sqrt{1 - x^2}$ nach y auflösen lässt.

2) Wir betrachten die Gleichung

$$F(x, y) = (x - y)\left(\frac{1}{2}x - y\right) = 0.$$

Die Ableitung ist $D_y F(x, y) = (y - \frac{1}{2}x) + (y - x)$. Wegen $D_y F(0, 0) = 0$ ist Satz 3.10 nicht anwendbar. In der Tat ergibt formales Auflösen nach y :

$$y_{\pm} = \frac{3}{4}x \pm \frac{1}{4}x,$$

und wir sehen, dass die Funktion $y = y(x)$ bei $(x, y) = (0, 0)$ gerade ihren Zweig wechselt, d. h.: Sie ist dort nicht eindeutig bestimmt.

3.3.2 Reguläre Abbildungen

Wir beschäftigen uns jetzt mit der Invertierbarkeit von Abbildungen $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, d. h. mit der Existenz der Umkehrabbildung $f^{-1} : B_f \rightarrow \mathbb{R}^n$. In einer Dimension folgt für eine stetig differenzierbare Funktion $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ auf einem Intervall I aus der strikten Monotonie,

$$f'(x) \neq 0, \quad x \in D,$$

die Existenz der Umkehrfunktion $f^{-1} : B_f \rightarrow \mathbb{R}$, welche auch wieder stetig differenzierbar ist. Ein ähnliches Resultat gilt auch für Abbildungen in höheren Dimensionen.

Beispiel 3.13: Die (offene) Teilmenge $D := \{(r, \theta), r \in \mathbb{R}_+, \theta \in \mathbb{R}\}$ der (r, θ) -Ebene wird durch die stetige Abbildung

$$x_1 = f_1(r, \theta) := r \cos \theta, \quad x_2 = f_2(r, \theta) := r \sin \theta,$$

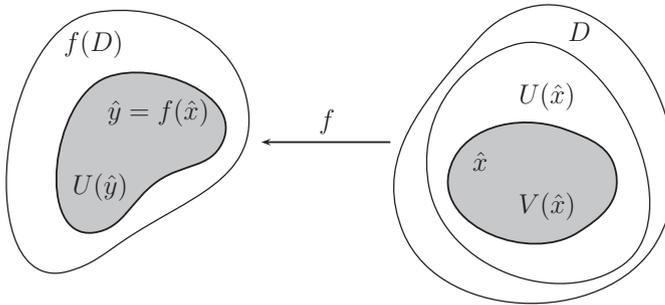
in die (x_1, x_2) -Ebene abgebildet. Diese Abbildung kann nur *lokal* injektiv sein (im Streifen $G := \{(r, \theta), r \in \mathbb{R}_+, \theta \in [0, 2\pi)\}$), aber nicht im Großen, da die (x_1, x_2) -Ebene wegen der Periodizität von Sinus und Kosinus mehrfach überdeckt wird. Für $r = 0$ tritt ein Problem auf, da offenbar alle Punkte der Form $(0, \theta)$, $\theta \in \mathbb{R}$, in den Nullpunkt $x = (0, 0)$ abgebildet werden, was der Injektivität widerspricht.

Definition 3.11: Sei $D \subset \mathbb{R}^n$ offen. Eine Abbildung $f : D \rightarrow \mathbb{R}^n$ heißt „regulär“ in einem Punkt $\hat{x} \in D$, wenn sie in einer Umgebung $K_{\delta}(\hat{x}) \subset D$ von \hat{x} stetig differenzierbar ist, und die Funktionalmatrix $J_f(\hat{x})$ regulär ist. Sie heißt „regulär in D “, wenn sie in jedem Punkt $\hat{x} \in D$ regulär ist.

Satz 3.11 (Umkehrabbildung): Sei $D \subset \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}^n$ regulär in einem Punkt $\hat{x} \in D$. Dann gibt es eine offene Umgebung $V(\hat{x}) \subset D$ von \hat{x} , die von f bijektiv auf eine offene Umgebung $U(\hat{y}) \subset \mathbb{R}^n$ von $\hat{y} := f(\hat{x})$ abgebildet wird. Die Umkehrabbildung $f^{-1} : U(\hat{y}) \rightarrow V(\hat{x})$ ist ebenfalls regulär in \hat{y} , und für ihre Funktionalmatrix und Funktionaldeterminante gilt:

$$J_{f^{-1}}(\hat{y}) = J_f(\hat{x})^{-1}, \quad \det J_{f^{-1}}(\hat{y}) = \frac{1}{\det J_f(\hat{x})}. \quad (3.3.48)$$

Beweis: Als Illustration des Beweises dient folgendes Bild:



Sei $\hat{x} \in D$ und $\hat{y} := f(\hat{x}) \in f(D)$. Wir betrachten die durch

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

definierte Abbildung $F : \mathbb{R}^n \times D \rightarrow \mathbb{R}^n$. Offenbar ist $F(\hat{y}, \hat{x}) = 0$. Ferner ist die Jacobi-Matrix $D_x F(y, x) = -J_f(x)$ gemäß Voraussetzung regulär in \hat{x} . Nach Satz 3.10 (angewendet mit vertauschten Rollen von x und y) gibt es also (offene) Umgebungen $U(\hat{y})$ von \hat{y} und $U(\hat{x})$ von \hat{x} sowie eine (eindeutig bestimmte) stetig differenzierbare Funktion $g : U(\hat{y}) \rightarrow U(\hat{x})$, so dass

$$0 = F(y, g(y)) = y - f(g(y)), \quad y \in U(\hat{y}).$$

Folglich gibt es zu jedem $y \in U(\hat{y})$ genau ein $x = g(y) \in U(\hat{x})$ mit $y = f(x)$. Wir setzen nun

$$V(\hat{x}) := U(\hat{x}) \cap f^{-1}(U(\hat{y})) = \{x \in U(\hat{x}) : f(x) \in U(\hat{y})\}.$$

Da $U(\hat{x})$ und $f^{-1}(U(\hat{y}))$ offen sind, ist $V(\hat{x})$ eine offene Umgebung von \hat{x} . Ferner wird $V(\hat{x})$ von f bijektiv auf $U(\hat{y})$ abgebildet. Die zugehörige Umkehrabbildung ist gerade $f^{-1} = g$. Wegen $J_{f \circ f^{-1}}(\cdot) = J_{\text{id}}(\cdot) = I$ folgt mit der Kettenregel

$$J_f(x) J_{f^{-1}}(f(x)) = I, \quad J_{f^{-1}}(f(x)) = J_f(x)^{-1}$$

und weiter mit dem Determinantensatz $\det J_{f^{-1}}(f(x)) = (\det J_f(x))^{-1}$. Q.E.D.

Korollar 3.4 (Offene Abbildung): *Ist die Abbildung $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ regulär, so ist für jede offene Menge $O \subset D$ auch die Bildmenge $f(O)$ offen. Solche Abbildungen werden auch als „offen“ bezeichnet.*

Beweis: Sei $O \subset D$ offen und $y \in f(O)$ beliebig mit $y = f(x)$, $x \in O$. Nach Satz 3.11 gibt es zu y Kugelumgebungen $K_r(y) \subset f(O)$ sowie $K_s(x) \subset O$, so dass $K_r(y) \subset f(K_s(x))$. Folglich ist $f(O)$ offen. Q.E.D.

Bemerkung 3.8: Satz 3.11 garantiert nur die *lokale* Umkehrbarkeit der Abbildung f . Für die *globale* Umkehrbarkeit gibt es nur wenig brauchbare Kriterien. Dazu bräuchte man, dass die implizite Funktion f im Satz 3.10 global eindeutig bestimmt ist, was nur unter sehr einschränkenden Bedingungen zu garantieren ist.

Bemerkung 3.9: Eine eineindeutige stetige Abbildung f einer offenen Menge $D \subset \mathbb{R}^n$ auf eine offene Menge $f(D) \subset \mathbb{R}^n$ mit stetiger Umkehrabbildung f^{-1} wird „Homöomorphismus“ genannt. Sind sowohl f als auch f^{-1} stetig differenzierbar, so spricht man von einem „Diffeomorphismus“.

Beispiel 3.14: Wir betrachten die durch

$$(x_1, x_2) = f(r, \theta) := (r \cos \theta, r \sin \theta)$$

definierte Abbildung $f : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}^2$. Dies ist die Transformation der sog. „Polarkoordinaten“ (r, θ) auf kartesischen Koordinaten (x_1, x_2) . Die Jacobi-Matrix von f ist

$$J_f(r, \theta) = \begin{pmatrix} \partial_r f_1 & \partial_\theta f_1 \\ \partial_r f_2 & \partial_\theta f_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix},$$

und die zugehörige Jacobi-Determinante ist $\det J_f(r, \theta) = r > 0$. Die Abbildung f ist also auf ganz $D = \mathbb{R}_+ \times \mathbb{R}$ regulär. Nach Satz 3.11 ist f also überall in D lokal umkehrbar. Die Jacobi-Matrix der lokalen Umkehrabbildung $f^{-1}(x_1, x_2)$ lautet

$$J_{f^{-1}}(x_1, x_2) = J_f(r, \theta)^{-1} = \begin{pmatrix} \cos \theta & \sin \theta \\ -r^{-1} \sin \theta & r^{-1} \cos \theta \end{pmatrix}.$$

Zur Umrechnung in die Variablen $(x_1, x_2) = (r \cos \theta, r \sin \theta)$ formen wir um:

$$r = \sqrt{x_1^2 + x_2^2}, \quad r^{-1}x = \cos \theta, \quad r^{-1}y = \sin \theta.$$

Damit ergibt sich

$$J_{f^{-1}}(x_1, x_2) = \frac{1}{\sqrt{x_1^2 + x_2^2}} \begin{pmatrix} x_1 \sqrt{x_1^2 + x_2^2} & x_2 \sqrt{x_1^2 + x_2^2} \\ -x_2 & x_1 \end{pmatrix}.$$

Wir kennen also die Jacobi-Matrix $J_{f^{-1}}(x_1, x_2)$ der Umkehrabbildung, ohne sie selbst explizit berechnet zu haben. Letzteres ist im vorliegenden Fall aber möglich. Bzgl. der Mengen

$$U := \mathbb{R}_+ \times \left(-\frac{1}{2}\pi, \frac{1}{2}\pi\right), \quad V := \mathbb{R}_+ \times \mathbb{R},$$

ist die Abbildung $f : U \rightarrow V$ bijektiv mit der Umkehrabbildung

$$f^{-1}(x, y) = \left(\sqrt{x_1^2 + x_2^2}, \arctan(x_1/x_2)\right).$$

Durch Berechnung der partiellen Ableitungen dieser Abbildung kann die obige Form ihrer Jacobi-Matrix bestätigt werden. Die Abbildung f bildet $\mathbb{R}_+ \times \mathbb{R}$ auf $\mathbb{R}^2 \setminus \{0\}$ ab, sie ist aber wegen $f(r, \theta) = f(r, \theta + 2k\pi)$, $k \in \mathbb{Z}$, nicht global injektiv.

3.3.3 Extremalaufgaben mit Nebenbedingungen

Im Folgenden betrachten wir sog. „restringierte“ Optimierungsaufgaben, speziell solche mit Gleichungsnebenbedingungen. Seien $f : D \rightarrow \mathbb{R}$ und $g : D \rightarrow \mathbb{R}$ differenzierbare Funktionen auf einer offenen Menge $D \subset \mathbb{R}^n$. Wir suchen einen Punkt $\hat{x} \in D$ mit der Eigenschaft

$$f(\hat{x}) = \inf \{f(x), x \in U(\hat{x}), g(\hat{x}) = 0\}, \quad (3.3.49)$$

oder analog $f(\hat{x}) = \sup \{f(x), x \in U(\hat{x}), g(\hat{x}) = 0\}$, für eine Umgebung $U(\hat{x})$. Analog kann auch die Maximierung von f betrachtet werden. Der folgende Satz gibt uns ein notwendiges Kriterium für eine Lösung dieser Aufgabe.

Satz 3.12 (Lagrange-Multiplikatoren): Sei $D \subset \mathbb{R}^n$ offen und $f, g : D \rightarrow \mathbb{R}$ zwei stetig differenzierbare Abbildungen. Ferner sei $\hat{x} \in D$ ein Punkt, in dem F ein lokales Extremum unter der Nebenbedingung $g(\hat{x}) = 0$ besitzt, d. h.: Mit der Menge

$$N_g := \{x \in D : g(x) = 0\}$$

gilt auf einer Umgebung $U(\hat{x}) \subset D$:

$$f(\hat{x}) = \inf_{x \in U \cap N_g} f(x) \quad \text{oder} \quad f(\hat{x}) = \sup_{x \in U \cap N_g} f(x). \quad (3.3.50)$$

Ist dann $\nabla g(\hat{x}) \neq 0$, so gibt es ein $\hat{\lambda} \in \mathbb{R}$, so dass

$$\nabla f(\hat{x}) = \hat{\lambda} \nabla g(\hat{x}). \quad (3.3.51)$$

Der Parameter $\hat{\lambda}$ wird „Lagrange-Multiplikator“ genannt.

Beweis: Wegen der Voraussetzung $\nabla g(\hat{x}) \neq 0$ können wir nach eventueller Umnummerierung der Koordinaten annehmen, dass $\partial_n g(\hat{x}) \neq 0$. Wir setzen

$$\hat{x} := (\hat{x}', \hat{x}_n) \in \mathbb{R}^n, \quad \hat{x}' = (\hat{x}_1, \dots, \hat{x}_{n-1}) \in \mathbb{R}^{n-1}.$$

Satz 3.10, angewendet auf die Gleichung

$$F(x', x_n) := g(x) = 0,$$

liefert die Existenz von Umgebungen $U(\hat{x}') \subset \mathbb{R}^{n-1}$ von \hat{x}' und $U(\hat{x}_n) \subset \mathbb{R}$ von \hat{x}_n mit $U(\hat{x}') \times U(\hat{x}_n) \subset D$, sowie einer (eindeutig bestimmten) stetig differenzierbaren Funktion $\varphi : U(\hat{x}') \rightarrow U(\hat{x}_n)$, so dass

$$F(x', \varphi(x')) = 0, \quad x' \in U(\hat{x}'), \quad (3.3.52)$$

und

$$N_g \cap (U(\hat{x}_n) \times U(\hat{x}')) = \{x \in U(\hat{x}_n) \times U(\hat{x}') : x_n = \varphi(x')\}.$$

Mit Hilfe der Kettenregel folgt aus (3.3.52):

$$\partial_i g(\hat{x}) + \partial_n g(\hat{x}) \partial_i \varphi(\hat{x}') = 0, \quad i = 1, \dots, n-1. \quad (3.3.53)$$

Da f auf N_g im Punkt \hat{x} ein lokales Extremum besitzt, hat die Funktion

$$\tilde{f}(x') := F(x', \varphi(x'))$$

auf $U(\hat{x}')$ im Punkt \hat{x}' ein lokales Extremum. Nach Satz 3.7 gilt also notwendigerweise

$$0 = \partial_i \tilde{f}(\hat{x}') = \partial_i f(\hat{x}) + \partial_n f(\hat{x}) \partial_i \varphi(\hat{x}'), \quad i = 1, \dots, n-1. \quad (3.3.54)$$

Definieren wir nun

$$\hat{\lambda} := \partial_n f(\hat{x}) \partial_n g(\hat{x})^{-1} \quad \text{bzw.} \quad \partial_n f(\hat{x}) = \hat{\lambda} \partial_n g(\hat{x}),$$

so ergibt sich zusammen mit (3.3.53) und (3.3.54)

$$\partial_i f(\hat{x}) = \hat{\lambda} \partial_i g(\hat{x}), \quad i = 1, \dots, n,$$

bzw.

$$\nabla f(\hat{x}) = \hat{\lambda} \nabla g(\hat{x}).$$

Q.E.D.

Bemerkung 3.10: Die Aussage von Satz 3.12 kann auch so interpretiert werden, dass jeder lokale Minimalpunkt \hat{x} der Funktion F unter der Nebenbedingung $g(\hat{x}) = 0$ notwendig zu einem sog. „stationären Punkt“ der Lagrange-Funktion

$$\mathcal{L}(x, \lambda) := f(x) - \lambda g(x), \quad (x, \lambda) \in D \times \mathbb{R},$$

korrespondiert, d. h. zu einem Punkt $(\hat{x}, \hat{\lambda})$ mit

$$\nabla_{(x,\lambda)} \mathcal{L}(\hat{x}, \hat{\lambda}) = \begin{pmatrix} \nabla_x f(\hat{x}) - \hat{\lambda} \nabla_x g(\hat{x}) \\ g(\hat{x}) \end{pmatrix} = 0. \quad (3.3.55)$$

Dieser Lösungsansatz für Optimierungsprobleme mit Gleichungsnebenbedingungen wird „Euler-Lagrange-Formalismus“ (oder auch „indirekte“ Lösungsmethode) genannt. Im Gegensatz dazu wird bei der „direkten“ Lösungsmethode die Nebenbedingung $g(x) = 0$ explizit z. B. nach $x_n = \varphi(x_1, \dots, x_{n-1})$ aufgelöst und dann die reduzierte, unrestringierte Optimierungsaufgabe

$$f(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1})) \rightarrow \min.$$

gelöst. In der Praxis ist diese explizite Auflösung aber häufig schwierig, so dass die indirekte Methode trotz der damit verbundenen Dimensionserhöhung meist vorgezogen wird.

Beispiel 3.15: Sei $A = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix und $a(\cdot)$ die zugehörige quadratische Form

$$f(x) := (x, Ax)_2 = \sum_{i,j=1}^n a_{ij}x_i x_j.$$

Wir wollen die Extrema von $a(x)$ unter der Nebenbedingung $\|x\|_2 = 1$ bestimmen. Wir definieren:

$$g(x) := \|x\|_2^2 - 1, \quad N_g := \{x \in \mathbb{R}^n : g(x) = 0\}.$$

Wegen $\nabla g(x) = 2x$ ist $\nabla g(x) \neq 0$ für $x \in N_g$. Weiter gilt wegen $a_{ij} = a_{ji}$:

$$\begin{aligned} \partial_k f(x) &= \sum_{i,j=1}^n a_{ij} \delta_{ik} x_j + \sum_{i,j=1}^n a_{ij} x_i \delta_{jk} \\ &= \sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n a_{ik} x_i = 2 \sum_{i=1}^n a_{ki} x_i. \end{aligned}$$

In kompakter Schreibweise heißt dies $\nabla f(x) = 2Ax$. Auf der kompakten Menge N_g nimmt die stetige Funktion f ihr Maximum an. Nach Satz 3.12 gibt es ein $\hat{\lambda} \in \mathbb{R}$, so dass

$$A\hat{x} = \hat{\lambda}\hat{x}.$$

Dies bedeutet, dass λ Eigenwert der Matrix A mit dem Eigenvektor \hat{x} ist. Wegen

$$f(\hat{x}) = (\hat{x}, A\hat{x})_2 = (\hat{x}, \hat{\lambda}\hat{x})_2 = \hat{\lambda}$$

wird das Minimum bei einem Eigenvektor zum kleinsten Eigenwert λ_{\min} angenommen. Für diesen gilt dann offenbar

$$\lambda_{\min} = \min_{x \in \mathbb{R}^n} \frac{(x, Ax)_2}{\|x\|_2^2},$$

d. h. er ist charakterisiert als das Minimum des sog. „Rayley-Quotienten“

$$\mathcal{R}(x) := \frac{(x, Ax)_2}{\|x\|_2^2}$$

der Matrix A auf \mathbb{R}^n . Analog sieht man, dass der maximale Eigenwert λ_{\max} entsprechend als das Maximum des Rayley-Quotienten charakterisiert ist.

Beispiel 3.16: Wir wollen das Maximum der auf \mathbb{R}^n definierten Funktion

$$f(x) := (x_1 \cdot \dots \cdot x_n)^2$$

auf der Sphäre $S_1 = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ bestimmen. Die Funktion f und die Funktion g in der Nebenbedingung

$$g(x) := \sum_{i=1}^n x_i^2 - 1 = 0 \tag{3.3.56}$$

sind offenbar Polynome und damit stetig differenzierbar. Da die Sphäre $S_1 \subset \mathbb{R}^n$, kompakt ist, nimmt f auf S_1 sein Maximum und Minimum an, wobei offenbar $\min_{x \in S_1} f(x) = 0$ und $\max_{x \in S_1} f(x) > 0$. Ferner ist $\nabla g(x) = 2x \neq 0$ für $x \in S_1$. Nach Satz 3.12 sind die Extrempunkte also unter den Lösungen $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}$ des Gleichungssystems

$$\partial_i f(x) = \lambda \partial_i g(x), \quad i = 1, \dots, n,$$

zu suchen. Dieses ist äquivalent zu

$$\frac{2(x_1 \cdot \dots \cdot x_n)^2}{x_i} = 2\lambda x_i \quad \text{bzw.} \quad (x_1 \cdot \dots \cdot x_n)^2 = \lambda x_i^2, \quad i = 1, \dots, n.$$

Wegen $\max_{x \in S_1} f(x) > 0$, können wir $x_i \neq 0$ annehmen und erhalten ebenfalls $\lambda \neq 0$. Summation über $i = 1, \dots, n$ und Berücksichtigung der Nebenbedingung (3.3.56) liefert daher

$$n(x_1 \cdot \dots \cdot x_n)^2 = \lambda \sum_{i=1}^n x_i^2 = \lambda,$$

und damit weiter

$$(x_1 \cdot \dots \cdot x_n)^2 = n(x_1 \cdot \dots \cdot x_n)^2 x_i^2, \quad i = 1, \dots, n,$$

bzw.

$$x_i^2 = \frac{1}{n}, \quad i = 1, \dots, n.$$

Dies beschreibt alle möglichen (nicht trivialen) Lösungen $(\hat{x}, \hat{\lambda})$ des Euler-Lagrange-Systems, d. h. möglicher lokaler Extrema, wobei die zugehörigen Funktionswerte $f(\hat{x}) = n^{-n}$ offenbar alle gleich sind. Damit ergibt sich folgendes Resultat: Das (globale) Maximum von f auf der Sphäre S_1 wird u. a. in dem Punkt

$$\hat{x} = (n^{-1/2}, \dots, n^{-1/2}), \quad f(\hat{x}) = n^{-n},$$

angenommen. Für alle Punkte $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ mit der Eigenschaft $x_1^2 + \dots + x_n^2 = 1$ gilt also die Ungleichung

$$(x_1 \cdot \dots \cdot x_n)^2 \leq n^{-n}. \quad (3.3.57)$$

Wendet man diese Ungleichung auf die zu einem beliebigen $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$ gebildeten Werte $\xi_i := \sqrt{x_i} / (\sum_{i=1}^n x_i)^{1/2}$ an, so erhält man die bekannte Beziehung zwischen geometrischem und arithmetischem Mittel positiver Zahlen $x_i > 0$:

$$\sqrt[n]{x_1 \cdot \dots \cdot x_n} \leq \frac{x_1 + \dots + x_n}{n}. \quad (3.3.58)$$

3.4 Übungen

Übung 3.1: Man berechne die partiellen Ableitungen $\partial_i f$ der folgenden Funktionen $f: \mathbb{R}^n \setminus \{0\}$, wobei $r(x) := \|x\|_2$ bezeichnet:

$$a) \quad f(x) = r(x)^{-n}, \quad b) \quad f(x) = e^{-1/r(x)^2}.$$

Übung 3.2: Die Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ sei definiert durch

$$f(x, y) := \frac{x^3y - xy^3}{x^2 + y^2}, \quad (x, y) \neq 0, \quad f(0, 0) := 0.$$

Man zeige:

- f ist auf ganz \mathbb{R}^2 zweimal partiell differenzierbar, d. h.: Die zugehörige Hesse-Matrix $\nabla^2 f$ existiert überall.
- f und ∇f sind stetig auf \mathbb{R}^2 .
- Die Hesse-Matrix $\nabla^2 f$ ist stetig auf $\mathbb{R}^2 \setminus \{0\}$; es ist aber $\partial_x \partial_y f(0, 0) \neq \partial_y \partial_x f(0, 0)$, d. h.: Die Hesse-Matrix ist in $(x, y) = (0, 0)$ nicht symmetrisch.

Übung 3.3: a) Man berechne den Gradienten $\nabla f = (\partial_i f)_{i=1}^n$ und die Hesse-Matrix $\nabla^2 f := (\partial_i \partial_j f)_{i,j=1}^n$ der folgenden Funktionen $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ für $\varepsilon \in \mathbb{R}_+$:

$$(i) \quad f_1(x) = (\|x\|_2^2 + \varepsilon)^{-1}, \quad (ii) \quad f_2(x) = e^{\|x\|_2^2}.$$

- Man diskutiere die Eigenschaften der Hesse-Matrizen dieser Funktionen im \mathbb{R}^2 .

Übung 3.4: a) Man berechne den Gradienten und die Hesse-Matrix der durch

$$f(x) = \|x\|_2^3 - 1, \quad x \in \overline{K_1(0)} = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\},$$

gegebenen Funktion $f : \overline{K_1(0)} \rightarrow \mathbb{R}$.

- Wo liegen die Maximal- und Minimalstellen der Funktion f ?

Übung 3.5: Man berechne die Jacobi-Matrix und die Jacobi-Determinante der durch

$$v(r, \theta, \varphi) := \begin{pmatrix} r \cos \theta \sin \varphi \\ r \sin \theta \sin \varphi \\ r \cos \varphi \end{pmatrix}$$

definierten Abbildung $v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. In welchen Punkten (r, θ, φ) ist $J_v(r, \theta, \varphi)$ regulär?

Übung 3.6: Punkte $(x_1, x_2) \in \mathbb{R}^2$ haben die „Polarkoordinatendarstellung“

$$x_1 = r \cos \theta, \quad x_2 = r \sin \theta.$$

mit dem Radius $r = \|x\|_2 \in \mathbb{R}_+ \cup \{0\}$ und dem Winkel $\theta \in (0, 2\pi]$ zwischen dem Ortsvektor x und der x_1 -Achse. Jede Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ lässt sich als Funktion bzgl. dieser Polarkoordinaten schreiben:

$$f(x_1, x_2) = f(r \cos \theta, r \sin \theta) =: F(r, \theta).$$

i) Man zeige, dass der Laplace-Operator $\Delta = \operatorname{div} \operatorname{grad}$ für solche Funktionen $F(r, \theta)$ die folgende Form hat:

$$\Delta f(x_1, x_2) = (\partial_1^2 + \partial_2^2)f(x_1, x_2) = (\partial_r^2 + r^{-1}\partial_r + r^{-2}\partial_\theta^2)F(r, \theta).$$

ii) Man zeige, dass für $\omega \in (0, 2\pi]$ die auf dem Sektor $S_\omega := \{(r, \theta) : r \geq 0, \theta \in [0, \omega]\}$ der (x, y) -Ebene definierte Funktion

$$s_\omega(r, \theta) := r^{\pi/\omega} \sin(\theta\pi/\omega)$$

in S_ω° harmonisch ist, d. h. $\Delta s_\omega \equiv 0$ erfüllt, und den „Randbedingungen“ $s_\omega(r, 0) = s_\omega(r, \omega) = 0$ genügt.

Übung 3.7: Bei der Untersuchung spezieller Funktionen ist es häufig nützlich, ihrer Struktur besonderes gut angepasste Koordinatensysteme zu verwenden. Dazu gehören z. B. die sog. „Kugelkoordinaten“ (auch „Polarkoordinaten“ in zwei Dimensionen):

$$n = 2 : \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r \cos(\theta) \\ r \sin(\theta) \end{bmatrix}, \quad n = 3 : \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} r \cos(\theta) \sin(\varphi) \\ r \sin(\theta) \sin(\varphi) \\ r \cos(\varphi) \end{bmatrix}.$$

mit dem Radius $r = \|x\|_2 \in \mathbb{R}_+ \cup \{0\}$ und den Winkeln $\theta \in [0, 2\pi)$ zwischen dem Ortsvektor x und der x_1 -Achse sowie, in drei Dimensionen, $\varphi \in [0, \pi]$ zwischen dem Ortsvektor x und der x_3 -Achse. Jede Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ($n = 2, 3$) lässt sich als Funktion bzgl. dieser Kugelkoordinaten schreiben:

$$n = 2 : \quad f(x_1, x_2) = f(r \cos(\theta), r \sin(\theta)) =: F(r, \theta),$$

$$n = 3 : \quad f(x_1, x_2, x_3) = f(r \cos(\theta) \sin(\varphi), r \sin(\theta) \sin(\varphi), r \cos(\varphi)) =: F(r, \theta, \varphi).$$

a) Man zeige, dass der Laplace-Operator $\Delta = \operatorname{div} \operatorname{grad} = \sum_{i=1}^n \partial_i^2$ für solche Funktionen $F(r, \theta)$ bzw. $F(r, \theta, \varphi)$ die folgende Form hat:

$$n = 2 : \quad \Delta f(x_1, x_2) = (\partial_r^2 + r^{-1}\partial_r + r^{-2}\partial_\theta^2)F(r, \theta),$$

$$n = 3 : \quad \Delta f(x_1, x_2, x_3) = \left(\partial_r^2 + \frac{2}{r}\partial_r + \frac{1}{r^2 \sin^2(\varphi)}\partial_\theta^2 + r^{-2}\partial_\varphi^2 + \frac{\cos(\varphi)}{r^2 \sin(\varphi)}\partial_\varphi \right) F(r, \theta, \varphi)$$

b) Man bestimme für den Laplace-Operator $\Delta := \operatorname{div} \operatorname{grad}$ im $\mathbb{R}^n \setminus \{0\}$:

$$n = 2 : \quad \Delta \log(\|x\|_2) = ?, \quad n = 3 : \quad \Delta(\|x\|_2^{-1}) = ?$$

c) Man berechne die Jacobi-Matrix und die Jacobi-Determinante für die durch

$$i) \quad v(r, \theta, \varphi) := \begin{pmatrix} r \cos(\theta) \sin(\varphi) \\ r \sin(\theta) \sin(\varphi) \\ r \cos(\varphi) \end{pmatrix}, \quad ii) \quad v(r, \theta, z) := \begin{pmatrix} r \cos(\theta) \\ r \sin(\theta) \\ z \end{pmatrix},$$

definierten Abbildungen $v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Zwischen welchen Teilmengen Urbild- und Bildbereich sind diese Abbildungen jeweils bijektiv? In welchen Punkten ist J_v jeweils irregulär? (Bemerkung: Die Abbildung (i) entspricht gerade den o. a. Kugelkoordinaten im \mathbb{R}^3 , während die Abbildung (ii) zu den sog. „Zylinderkoordinaten“ im \mathbb{R}^3 gehört.)

Übung 3.8: Wo ist die durch

$$f(x, y) = |xy|$$

definierte Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ partiell bzw. total differenzierbar?

Übung 3.9: Für welche Argumente $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ sind die durch

$$i) \quad f_1(x) = |x_1 x_2| + |x_3|, \quad ii) \quad f_2(x) = |x_1 x_2 x_3|^{1/3}$$

definierten Funktionen $f_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ partiell bzw. total differenzierbar?

Übung 3.10: Sei $M \subset \mathbb{R}^n$ eine offene Menge und $f : M \rightarrow \mathbb{R}^n$ eine Lipschitz-stetige Abbildung mit Lipschitz-Konstante L . Man zeige:

i) Die Abbildung f besitzt eine Lipschitz-stetige Fortsetzung \bar{f} auf \bar{M} , d. h. eine Abbildung $\bar{f} : \bar{M} \rightarrow \mathbb{R}^n$ mit $\bar{f}(x) = f(x)$ für $x \in M$ und

$$\|\bar{f}(x) - \bar{f}(y)\| \leq L\|x - y\|, \quad x, y \in \bar{M}.$$

ii) Ist M beschränkt, so ist der Bildbereich $\bar{f}(\bar{M}) \subset \mathbb{R}^n$ abgeschlossen.

Übung 3.11: a) Sei $D \subset \mathbb{R}^n$ eine offene und konvexe Menge und $f : D \rightarrow \mathbb{R}^n$ eine differenzierbare Abbildung mit gleichmäßig beschränkter Jacobi-Matrix

$$\sup_{x \in D} \|J_f(x)\|_2 \leq K_2 < \infty.$$

Man zeige, dass f dann in D Lipschitz-stetig ist, d. h.:

$$\|f(y) - f(x)\|_2 \leq K_2 \|y - x\|_2, \quad x, y \in D.$$

b) Gilt eine analoge Aussage auch, wenn man die Spektralnorm $\|\cdot\|_2$ durch eine beliebige andere (natürliche) Matrixnorm $\|\cdot\|$ ersetzt?

c) Zusatzaufgabe für Anspruchsvolle: Mit den von der l_∞ - und der l_1 -Norm erzeugten natürlichen Matrixnormen $\|\cdot\|_\infty$ („maximale Zeilensumme“) bzw. $\|\cdot\|_1$ („maximale Spaltensumme“) gelte

$$\sup_{x \in D} \|J_f(x)\|_\infty \leq K_\infty, \quad \sup_{x \in D} \|J_f(x)\|_1 \leq K_1.$$

Man zeige, dass dann

$$\|f(y) - f(x)\|_2 \leq \sqrt{K_\infty K_1} \|y - x\|_2, \quad x, y \in D.$$

Übung 3.12: Im Band Analysis 1 (Abschnitt 5.3) wird gezeigt, dass für eine $r + 1$ -mal stetig differenzierbare Funktion $f : (a, b) \rightarrow \mathbb{R}$ um jeden Punkt $x_0 \in (a, b)$ die Taylor-Approximation

$$f(x) = \sum_{k=0}^r \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + R_{r+1}^f(x_0; x - x_0),$$

besteht, mit dem Restglied $R_{r+1}^f(x_0; x - x_0)$ in integraler und differentieller Form

$$R_{r+1}^f(x_0; x - x_0) = \frac{1}{r!} \int_{x_0}^x (x-t)^r f^{(r+1)}(t) dt = \frac{f^{(r+1)}(\xi)}{(r+1)!} (x-x_0)^{r+1}, \quad \xi \in (x_0, x).$$

Man leite durch eine Variablentransformation hieraus die folgenden Darstellungen ab:

$$f(x_0 + h) = \sum_{k=0}^r \frac{f^{(k)}(x_0)}{k!} h^k + R_{r+1}^f(x_0; h),$$

mit dem Restglied $R_{r+1}^f(x_0; x_0 + h)$ in integraler und differentieller Form

$$R_{r+1}^f(x_0; h) = \frac{h^{r+1}}{r!} \int_0^1 (1-s)^r f^{(r+1)}(x_0 + sh) ds = \frac{f^{(r+1)}(x_0 + \theta h)}{(r+1)!} h^{r+1}, \quad \theta \in (0, 1),$$

(Hinweis: Man versuche es mit der Transformation $s := (t - x_0)/h$.)

Übung 3.13: Man gebe die Taylor-Entwicklung der durch

$$f(x) = \frac{x_1 - x_2}{x_1 + x_2}$$

definierten Funktion $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ um den Punkt $x = (1, 1)$ bis zum Restglied $R_3^f(x; h)$ 3-ter Ordnung an (Das Restglied selbst braucht nicht berechnet zu werden.).

Übung 3.14: Man zeige, dass die durch

$$f(x) := \ln(1 + x_2 - x_1), \quad x \in D = \{y \in \mathbb{R}^2 : y_2 - y_1 > -1\},$$

definierte Funktion $f : D \rightarrow \mathbb{R}$ die folgende Taylor-Reihe um den Punkt $x = 0$ hat:

$$T_\infty^f(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (x_2 - x_1)^k$$

Wie sieht der Konvergenzbereich dieser Reihe aus, und stellt sie die Funktion dort dar?

Übung 3.15: Man betrachte die durch

$$f(x_1, x_2) = 2x_1^2 - 3x_1x_2^2 + x_2^4$$

definierte Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$.

a) Man zeige, dass f entlang jeder Geraden durch den Nullpunkt in $(0, 0)$ ein (relatives) Minimum besitzt.

b) Man zeige, dass f im Nullpunkt $(0, 0)$ aber kein lokales Minimum in \mathbb{R}^2 besitzt.

(Bem.: Dieses warnende Beispiel sollte man sich ganz klar machen.)

Übung 3.16: Seien $a^{(j)} \in \mathbb{R}^n$, $j = 1, \dots, m$, gegebene Punkte. Man bestimme die lokalen und globalen Minima der durch

$$f(x) := \sum_{j=1}^m \|x - a^{(j)}\|_2^2$$

definierten Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$.

Übung 3.17: Man bestimme die Extrema der folgenden Funktionen auf \mathbb{R}^2 :

$$a) \quad f(x) = x_1^3 + x_2^3 - 9x_1x_2 + 27; \quad b) \quad f(x) = x_1^2 + x_2^2 - 2x_1x_2 + 1.$$

Übung 3.18: Man zeige, dass durch die Gleichung

$$F(x, y, z) = z^3 + (x^2 + y^2)z + 1 = 0$$

eindeutig eine Funktion $z = f(x, y)$ auf ganz \mathbb{R}^2 bestimmt ist, welche stetig differenzierbar ist.

Übung 3.19: Man untersuche mit Hilfe des Satzes über implizite Funktionen, ob die Gleichung

$$x^y = y^x$$

in der Nähe der Punkte $(2, 4)$ und (e, e) nach einer der beiden Variablen auflösbar ist.

Übung 3.20: Man untersuche die Abbildung

$$u = e^x \cos(y), \quad v = e^x \sin(y),$$

der Menge $M := \{(x, y) \in \mathbb{R}^2 : x \in [0, 1], y \in [-\frac{1}{2}\pi, \frac{1}{2}\pi]\}$ nach \mathbb{R}^2 , bestimme ihren Bildbereich und gegebenenfalls die inverse Abbildung. Ist die Abbildung im Großen, d. h. als Abbildung von \mathbb{R}^2 nach \mathbb{R}^2 umkehrbar?

Übung 3.21: Die reelle Zahl $a > 0$ ist so in drei positive Summanden zu zerlegen, dass deren Produkt maximal ist. (Hinweis: Man formuliere dies als eine Optimierungsaufgabe mit Gleichungsrestriktion und benutze zu deren Lösung den Lagrange-Ansatz.)

Übung 3.22: Man bestimme die Maxima und Minima des Polynoms

$$f(x, y) = 4x^2 - 3xy$$

auf der Kreisscheibe $K := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

(Hinweis: Man berechne zunächst die lokalen Extrema von f im Innern von K und dann auf dem Rand von K , d. h. unter der Gleichungsnebenbedingung $x^2 + y^2 = 1$.)

Übung 3.23: Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : D \rightarrow \mathbb{R}$ eine unendlich oft differenzierbare Funktion mit gleichmäßig beschränkten partiellen Ableitungen, d. h.: In der üblichen Multi-Indexschreibweise gilt:

$$\sup_{|\alpha| \geq 0} \left(\sup_{x \in D} |\partial^\alpha f(x)| \right) < \infty.$$

a) Man zeige, dass dann die Taylor-Reihe von f ,

$$T_\infty^f(x+h) := \sum_{|\alpha|=0}^{\infty} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha, \quad x \in D,$$

für Inkremente $h \in \mathbb{R}^n$ mit $x+th \in D$, $t \in [0, 1]$, absolut konvergiert und die Funktion f dort darstellt:

$$f(x+h) = T_\infty^f(x+h), \quad x \in D.$$

b) Man bestimme die formale Taylor-Reihe um den Punkt $x = (1, 0, 0)$ der durch

$$f(x) := e^{1+x_1+x_2+x_3}$$

definierten Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}$. Was ist der Konvergenzbereich dieser Reihe und wo stellt sie die Funktion f dar?

Übung 3.24: Man bestimme die Extrema der folgenden Funktionen auf \mathbb{R}^2 :

$$\begin{aligned} a) \quad & f(x) = x_1^3 + x_2^3 - 16x_1x_2 + 1; \\ b) \quad & f(x) = (x_1 - x_2)^4 + (x_2 - 1)^4. \end{aligned}$$

Man diskutiere, ob die gefundenen Extrema auch „global“ sind.

Übung 3.25: Man betrachte die durch

$$u(x, y) := \frac{1}{2} \ln(x^2 + y^2), \quad v(x, y) := \begin{cases} \arctan(y/x), & x \neq 0, \\ -\frac{\pi}{2}, & x = 0, \end{cases}$$

definierte Abbildung der Menge $M := \{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq e^2, x \leq 0\}$ nach \mathbb{R}^2 .

a) Man bestimme die Jacobi-Matrix und -Determinante dieser Abbildung; ist die Abbildung regulär?

b) Man bestimme den Bildbereich und gegebenenfalls die zugehörige inverse Abbildung.

Übung 3.26: Man beschreibe das Prinzip des Lagrange-Formalismus und wende diesen an zur Bestimmung der Maxima und Minima des Polynoms

$$f(x, y) = x - y$$

auf der Kreislinie $K := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

Übung 3.27: Man bestimme mit Hilfe des Lagrange-Formalismus den euklidischen Abstand des Punktes $x^* := (1, -1, 0) \in \mathbb{R}^3$ zum Rotationshyperboloid

$$M := \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1^2 + x_2^2 - x_3^2 = 1\}$$

(Hinweis: Der euklidische Abstand zweier Punkte $x, y \in \mathbb{R}^3$ ist definiert durch $d(x, y) := \|x - y\|_2$.)