Analysis III für Studierende der Ingenieurwissenschaften

Ingenuin Gasser

Fachbereich Mathematik
Universität Hamburg

Technische Universität Hamburg–Harburg
Wintersemester 2021/22

basierend auf dem Foliensatz von Jens Struckmeier Wintersemster 2020/21

Inhalte der Vorlesung Analysis III.

1. Partielle Ableitungen, Differentialoperatoren.
2. Vektorfelder, vollständiges Differential, Richtungsableitungen.
3. Mittelwertsätze, Satz von Taylor.
4. Extrema, Satz über implizite Funktionen.
5. Implizite Darstellung von Kurven und Flächen.
6. Extrema bei Gleichungsnebenbedingungen.
7. Newton–Verfahren, nichtlineare Gleichungen und Ausgleichsrechnung.
8. Bereichsintegrale, Satz von Fubini, Transformationssatz.
9. Potentiale, Integralsatz von Green, Integralsatz von Gauß.
10. Greensche Formeln, Integralsatz von Stokes.

1.1 Partielle Ableitungen

Im Folgenden sei

$f(x_1, \ldots, x_n)$ eine skalare Funktion, die von n Variablen abhängt

Beispiel: Die Zustandsgleichung eines idealen Gases lautet $pV = RT$.

Jede der drei Größen p (Druck), V (Volumen) und T (Temperatur) läßt sich als Funktion der anderen darstellen, wobei R die universelle Gaskonstante ist.

$$\begin{aligned} p &= p(V, t) = \frac{RT}{V} \\ V &= V(p, T) = \frac{RT}{p} \\ T &= T(p, V) = \frac{pV}{R} \end{aligned}$$

1.1. Partielle Ableitungen

Definition: Sei $D \subset \mathbb{R}^n$ offen, $f : D \to \mathbb{R}$, $\mathrm{x}^0 \in D$.

- $f(\mathrm{x})$ heißt in x^0 nach x_i partiell differenzierbar, falls der Grenzwert

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\mathrm{x}^0) &:= \lim_{t\to 0} \frac{f(\mathrm{x}^0 + t\mathrm{e}_i) - f(\mathrm{x}^0)}{t} \\ &= \lim_{t\to 0} \frac{f(x_1^0, \ldots, x_i^0 + t, \ldots, x_n^0) - f(x_1^0, \ldots, x_i^0, \ldots, x_n^0)}{t} \end{aligned}$$

  existiert, wobei e_i den i–ten Einheitsvektor bezeichnet. Den Grenzwert nennt man die partielle Ableitung von $f(\mathrm{x})$ nach x_i im Punkt x^0.

- Existieren für jeden Punkt x^0 die partiellen Ableitungen nach jeder Variablen $x_i, i = 1, \ldots, n$ und sind diese **stetige Funktionen**, so nennt man $f(\mathrm{x})$ stetig partiell differenzierbar oder eine $\mathcal{C}^1$–Funktion.

Beispiele.

- Betrachte die Funktion

$$f(x_1, x_2) = x_1^2 + x_2^2$$

Für einen Punkt $\mathrm{x}^0 \in \mathbb{R}^2$ existieren beide partiellen Ableitungen und diese sind auch stetig:

$$\frac{\partial f}{\partial x_1}(\mathrm{x}^0) = 2x_1, \qquad \frac{\partial f}{\partial x_2}(\mathrm{x}^0) = 2x_2$$

Die Funktion ist also eine $\mathcal{C}^1$–Funktion.

- Die Funktion

$$f(x_1, x_2) = x_1 + |x_2|$$

ist im Punkt $\mathrm{x}^0 = (0, 0)^T$ partiell differenzierbar nach der Koordinate x_1, aber die partielle Ableitung nach x_2 existiert im Ursprung **nicht**!

Konkretes technisches Beispiel.

Der Schalldruck einer eindimensionalen Schallwelle ist gegeben durch

$$p(x,t) = A\sin(\alpha x - \omega t)$$

Die partielle Ableitung

$$\frac{\partial p}{\partial x} = \alpha A\cos(\alpha x - \omega t)$$

beschreibt zu einer festen Zeit t die örtliche Änderungsrate des Schalldrucks.

Die partielle Ableitung

$$\frac{\partial p}{\partial t} = -\omega A\cos(\alpha x - \omega t)$$

beschreibt für einen festen Ort x die zeitliche Änderung des Schalldruckes.

Differentiationsregeln

- Sind f, g partiell nach x_i differenzierbar, $\alpha, \beta \in \mathbb{R}$, so gelten die Regeln

$$\frac{\partial}{\partial x_i}\Big(\alpha f(x) + \beta g(x)\Big) = \alpha \frac{\partial f}{\partial x_i}(x) + \beta \frac{\partial g}{\partial x_i}(x)$$

$$\frac{\partial}{\partial x_i}\Big(f(x) \cdot g(x)\Big) = \frac{\partial f}{\partial x_i}(x) \cdot g(x) + f(x) \cdot \frac{\partial g}{\partial x_i}(x)$$

$$\frac{\partial}{\partial x_i}\left(\frac{f(x)}{g(x)}\right) = \frac{\frac{\partial f}{\partial x_i}(x) \cdot g(x) - f(x) \cdot \frac{\partial g}{\partial x_i}(x)}{g(x)^2} \quad \text{für } g(x) \neq 0$$

- Man verwendet alternativ die Bezeichnungen:

$$D_i f(x^0) \qquad \text{oder} \qquad f_{x_i}(x^0)$$

für die partielle Ableitung von $f(x)$ nach x_i in x^0.

Gradient und Nabla–Operator.

Definition: Sei $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, und partiell differenzierbar.

- Man bezeichnet den **Zeilenvektor**

$$\operatorname{grad} f(\mathrm{x}^0) := \left(\frac{\partial f}{\partial x_1}(\mathrm{x}^0), \ldots, \frac{\partial f}{\partial x_n}(\mathrm{x}^0) \right)$$

  als Gradient von $f(\mathrm{x})$ in x^0.

- Weiterhin bezeichnet man den symbolischen Vektor

$$\nabla := \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right)^T$$

  als Nabla–Operator.

- So bekommt man den **Spaltenvektor**

$$\nabla f(\mathrm{x}^0) := \left(\frac{\partial f}{\partial x_1}(\mathrm{x}^0), \ldots, \frac{\partial f}{\partial x_n}(\mathrm{x}^0) \right)^T$$

Weitere Differentiationsregeln.

Seien $f(\mathrm{x})$ und $g(\mathrm{x})$ partiell differenzierbar. Dann gelten die folgenden Differentiationsregeln:

$$
\begin{aligned}
\operatorname{grad}(\alpha f + \beta g) &= \alpha \cdot \operatorname{grad} f + \beta \cdot \operatorname{grad} g \\
\operatorname{grad}(f \cdot g) &= g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g \\
\operatorname{grad}\left(\frac{f}{g}\right) &= \frac{1}{g^2}(g \cdot \operatorname{grad} f - f \cdot \operatorname{grad} g), \quad g \neq 0
\end{aligned}
$$

Beispiele:

- Sei $f(x,y) = e^x \cdot \sin y$. Dann gilt:

$$\operatorname{grad} f(x,y) = (e^x \cdot \sin y, e^x \cdot \cos y) = e^x(\sin y, \cos y)$$

- Für $r(\mathrm{x}) := \|\mathrm{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$ gilt

$$\operatorname{grad} r(\mathrm{x}) = \frac{\mathrm{x}}{r(\mathrm{x})} = \frac{\mathrm{x}}{\|\mathrm{x}\|_2} \qquad \text{für } \mathrm{x} \neq 0,$$

  wobei $\mathrm{x} = (x_1, \ldots, x_n)$ Zeilenvektor.

Partiell differenzierbar impliziert nicht Stetigkeit.

Beobachtung: Eine (nach allen Koordinaten) partiell differenzierbare Funktion ist nicht notwendigerweise eine stetige Funktion.

Beispiel: Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

$$f(x,y) := \begin{cases} \dfrac{x \cdot y}{(x^2+y^2)^2} & : \quad \text{für } (x,y) \neq 0 \\[2ex] 0 & : \quad \text{für } (x,y) = 0 \end{cases}$$

Die Funktion ist auf **ganz** $\mathbb{R}^2$ partiell differenzierbar, und es gilt

$$\begin{aligned} f_x(0,0) &= f_y(0,0) = 0 \\ \frac{\partial f}{\partial x}(x,y) &= \frac{y}{(x^2+y^2)^2} - 4\frac{x^2 y}{(x^2+y^2)^3}, \quad (x,y) \neq (0,0) \\ \frac{\partial f}{\partial y}(x,y) &= \frac{x}{(x^2+y^2)^2} - 4\frac{x y^2}{(x^2+y^2)^3}, \quad (x,y) \neq (0,0) \end{aligned}$$

Beispiel (Fortsetzung).

Berechnung der partiellen Ableitungen im Ursprung $(0,0)$:

$$\frac{\partial f}{\partial x}(0,0) = \lim_{t\to 0} \frac{f(t,0)-f(0,0)}{t} = \frac{\frac{t\cdot 0}{(t^2+0^2)^2} - 0}{t} = 0$$

$$\frac{\partial f}{\partial y}(0,0) = \lim_{t\to 0} \frac{f(0,t)-f(0,0)}{t} = \frac{\frac{0\cdot t}{(0^2+t^2)^2} - 0}{t} = 0$$

Aber: Im Nullpunkt $(0,0)$ ist die Funktion **nicht** stetig, denn

$$\lim_{n\to\infty} f\left(\frac{1}{n},\frac{1}{n}\right) = \frac{\frac{1}{n}\cdot\frac{1}{n}}{\left(\frac{1}{n}\cdot\frac{1}{n}+\frac{1}{n}\cdot\frac{1}{n}\right)^2} = \frac{\frac{1}{n^2}}{\frac{4}{n^4}} = \frac{n^2}{4} \to \infty$$

und somit gilt

$$\lim_{(x,y)\to(0,0)} f(x,y) \neq f(0,0) = 0$$

Beschränkheit der Ableitungen impliziert Stetigkeit.

Um die Stetigkeit einer partiell differenzierbaren Funktion zu garantieren, benötigt man zusätzliche Voraussetzungen an f.

Satz: Ist $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, in einer Umgebung von $\mathrm{x}^0 \in D$ partiell differenzierbar, und sind die partiellen Ableitungen $\dfrac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, dort beschränkt, so ist $f(\mathrm{x})$ stetig in x^0.

Beachte: In unserem vorigem Beispiel sind die partiellen Ableitungen in einer Umgebung der Null $(0,0)$ nicht beschränkt, denn es gilt

$$\frac{\partial f}{\partial x}(x,y) = \frac{y}{(x^2+y^2)^2} - 4\frac{x^2 y}{(x^2+y^2)^3} \quad \text{für } (x,y) \neq (0,0)$$

Beweis des Satzes.

Für $\|\mathrm{x} - \mathrm{x}^0\|_\infty < \varepsilon$, $\varepsilon > 0$ hinreichend klein, schreiben wir:

$$\begin{aligned} f(\mathrm{x}) - f(\mathrm{x}^0) &= (f(x_1, \ldots, x_{n-1}, x_n) - f(x_1, \ldots, x_{n-1}, x_n^0)) \\ &+ (f(x_1, \ldots, x_{n-1}, x_n^0) - f(x_1, \ldots, x_{n-2}, x_{n-1}^0, x_n^0)) \\ &\vdots \\ &+ (f(x_1, x_2^0, \ldots, x_n^0) - f(x_1^0, \ldots, x_n^0)) \end{aligned}$$

Für jede Differenz auf der linken Seite, betrachten wir f als univariate Funktion:

$$g(x_n) - g(x_n^0) := f(x_1, \ldots, x_{n-1}, x_n) - f(x_1, \ldots, x_{n-1}, x_n^0)$$

Da f partiell differenzierbar, ist g differenzierbar und es gilt der Mittelwertsatz:

$$g(x_n) - g(x_n^0) = g'(\xi_n)(x_n - x_n^0)$$

für ein geeignetes ξ_n zwischen x_n und x_n^0.

Beweis des Satzes (Fortsetzung).

Anwendung des Mittelwertsatzes auf jeden Term der rechten Seite ergibt somit

$$\begin{aligned} f(x) - f(x^0) &= \frac{\partial f}{\partial x_n}(x_1, \ldots, x_{n-1}, \xi_n) \cdot (x_n - x_n^0) \\ &+ \frac{\partial f}{\partial x_{n-1}}(x_1, \ldots, x_{n-2}, \xi_{n-1}, x_n^0) \cdot (x_{n-1} - x_{n-1}^0) \\ &\vdots \\ &+ \frac{\partial f}{\partial x_1}(\xi_1, x_2^0, \ldots, x_n^0) \cdot (x_1 - x_1^0) \end{aligned}$$

Mit der Beschrankheit der partiellen Ableitungen gilt

$$|f(x) - f(x^0)| \leq C_1|x_1 - x_1^0| + \cdots + C_n|x_n - x_n^0|$$

für $\|x - x^0\|_\infty < \varepsilon$, und damit ist $f(x)$ stetig in x^0, denn es gilt

$$f(x) \to f(x^0) \qquad \text{für } \|x - x^0\|_\infty \to 0$$

Höhere Ableitungen.

Definition: Eine skalare Funktion $f(x)$ sei auf einer offenen Menge $D \subset \mathbb{R}^n$ partiell differenzierbar. Sind die partiellen Ableitungen erneut partiell differenzierbar, so erhält man sämtliche partiellen Ableitungen zweiter Ordnung von f mit

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

Beispiel: Partielle Ableitungen zweiter Ordnung einer Funktion $f(x, y)$:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y^2}$$

Seien nun $i_1, \ldots, i_k \in \{1, \ldots, n\}$. Dann definiert man rekursiv

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \ldots \partial x_{i_1}} := \frac{\partial}{\partial x_{i_k}}\left(\frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \partial x_{i_{k-2}} \ldots \partial x_{i_1}}\right)$$

Ableitungen höherer Ordnung.

Definition: Die Funktion $f(\mathrm{x})$ heißt k–fach partiell differenzierbar, falls alle Ableitungen der Ordnung k,

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \ldots \partial x_{i_1}} \qquad \text{für alle } i_1, \ldots, i_k \in \{1, \ldots, n\},$$

auf D existieren.

Alternative Notationen:

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \ldots \partial x_{i_1}} = D_{i_k} D_{i_{k-1}} \ldots D_{i_1} f = f_{x_{i_1} \ldots x_{i_k}}$$

Sind alle Ableitungen k–ter Ordnung stetig, so heißt die Funktion $f(\mathrm{x})$ k–fach stetig partiell differenzierbar oder auch $\mathcal{C}^k$–Funktion auf D. Stetige Funktionen $f(\mathrm{x})$ nennt man auch $\mathcal{C}^0$–Funktionen.

Beispiel: Für die Funktion $f(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i^i$ gilt $\frac{\partial^n f}{\partial x_n \ldots \partial x_1} = ?$

Partielle Ableitungen sind nicht beliebig vertauschbar.

ACHTUNG: Die Reihenfolge, in der die partiellen Ableitungen durchzuführen sind, ist im Allgemeinen nicht beliebig vertauschbar!

Beispiel: Für die Funktion

$$f(x,y) := \begin{cases} xy\dfrac{x^2-y^2}{x^2+y^2} & : \quad \text{für } (x,y) \neq (0,0) \\[2ex] 0 & : \quad \text{für } (x,y) = (0,0) \end{cases}$$

berechnet man direkt

$$\begin{aligned} f_{xy}(0,0) &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}(0,0)\right) = -1 \\ f_{yx}(0,0) &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}(0,0)\right) = +1 \end{aligned}$$

d.h. $f_{xy}(0,0) \neq f_{yx}(0,0)$.

Vertauschbarkeitssatz von Schwarz.

Satz: Ist $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, eine $\mathcal{C}^2$–Funktion, so gilt

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x_1, \ldots, x_n) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \ldots, x_n)$$

für alle $i, j \in \{1, \ldots, n\}$.

Beweisidee:

Zweifache Anwendung des Mittelwertsatzes.

Folgerung:

Ist $f(\mathrm{x})$ eine C^k–Funktion, so kann man die Reihenfolge der Differentiationen zur Berechnung der partiellen Ableitungen bis zur k–ten Ordnung beliebig vertauschen!

Beispiel zur Vertauschbarkeit partieller Ableitungen.

Berechne für die Funktion

$$f(x,y,z) = y^2 z \sin(x^3) + (\cosh y + 17e^{x^2})z^2$$

die partielle Ableitung dritter Ordnung f_{xyz}.

Die Reihenfolge der partiellen Ableitungen ist vertauschbar, da $f \in \mathcal{C}^3$.

- Differenziere zunächst nach z:

$$\frac{\partial f}{\partial z} = y^2 \sin(x^3) + 2z(\cosh y + 17e^{x^2})$$

- Differenziere dann f_z nach x (damit fällt $\cosh y$ raus):

$$\begin{aligned} f_{zx} &= \frac{\partial}{\partial x}\left(y^2 \sin(x^3) + 2z(\cosh y + 17e^{x^2})\right) \\ &= 3x^2y^2\cos(x^3) + 68xze^{x^2} \end{aligned}$$

- Für die partielle Ableitung von f_{zx} nach y erhalten wir schließlich

$$f_{xyz} = 6x^2 y \cos(x^3)$$

Der Laplace–Operator.

Der Laplace–Operator ist definiert durch

$$\Delta := \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$$

Für eine skalare Funktion $u(x) = u(x_1, \ldots, x_n)$ gilt somit

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = u_{x_1 x_1} + \cdots + u_{x_n x_n}$$

Beispiele für wichtige partielle Differentialgleichungen zweiter Ordnung:

$$\Delta u - \frac{1}{c^2} u_{tt} = 0 \qquad \text{(Wellengleichung)}$$

$$\Delta u - \frac{1}{k} u_t = 0 \qquad \text{(Wärmeleitungsgleichung)}$$

$$\Delta u = 0 \qquad \text{(Laplace–Gleichung oder Potentialgleichung)}$$

Vektorwertige Funktionen.

Definition: Sei $\mathrm{f} : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$ offen, $\mathrm{f} = (f_1, \ldots, f_m)^T$, eine vektorwertige Funktion.

Die Funktion f heißt partiell differenzierbar in $\mathrm{x}^0 \in D$, falls für alle $i = 1, \ldots, n$ die Grenzwerte

$$\frac{\partial \mathrm{f}}{\partial x_i}(\mathrm{x}^0) = \lim_{t \to 0} \frac{\mathrm{f}(\mathrm{x}^0 + t\mathrm{e}_i) - \mathrm{f}(\mathrm{x}^0)}{t}$$

existieren. Die Berechnung erfolgt komponentenweise

$$\frac{\partial \mathrm{f}}{\partial x_i}(\mathrm{x}^0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_i} \\ \vdots \\ \dfrac{\partial f_m}{\partial x_i} \end{pmatrix} \qquad \text{für } i = 1, \ldots, n$$

Vektorfelder.

Definition: Für $m = n$ nennt man die Funktion $\mathrm{f} : D \to \mathbb{R}^n$ ein Vektorfeld auf D. Ist jede Koordinatenfunktion $f_i(\mathrm{x})$ von $\mathrm{f} = (f_1, \ldots, f_n)^T$ eine $\mathcal{C}^k$–Funktion, so nennt man f ein $\mathcal{C}^k$–Vektorfeld.

Beispiele für Vektorfelder:

- Geschwindigkeitsfelder von strömenden Flüssigkeiten oder Gasen;
- elektromagnetische Felder;
- Temperaturgradienten in Festkörpern.

Definition: Für ein partiell differenzierbares Vektorfeld $\mathrm{f} : D \to \mathbb{R}^n$ definiert man die Divergenz in $\mathrm{x} \in D$ durch

$$\operatorname{div} \mathrm{f}(\mathrm{x}^0) := \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathrm{x}^0)$$

oder

$$\operatorname{div} \mathrm{f}(\mathrm{x}) = \nabla^T \mathrm{f}(\mathrm{x}) = (\nabla, \mathrm{f}(\mathrm{x}))$$

Rechenregeln und Rotation.

Es gelten die folgenden Rechenregeln:

$$\begin{aligned} \operatorname{div}(\alpha \mathrm{f} + \beta \mathrm{g}) &= \alpha \operatorname{div} \mathrm{f} + \beta \operatorname{div} \mathrm{g} \quad \text{für } \mathrm{f}, \mathrm{g} : D \to \mathbb{R}^n \\ \operatorname{div}(\varphi \cdot \mathrm{f}) &= (\nabla \varphi, \mathrm{f}) + \varphi \operatorname{div} \mathrm{f} \quad \text{für } \varphi : D \to \mathbb{R}, \mathrm{f} : D \to \mathbb{R}^n \end{aligned}$$

Bemerkung: Ist $f : D \to \mathbb{R}$ eine $\mathcal{C}^2$–Funktion, so gilt für den Laplace–Operator

$$\Delta f = \operatorname{div}(\nabla f)$$

Definition: Für ein partiell differenzierbares Vektorfeld im $\mathbb{R}^3$, $\mathrm{f} : D \to \mathbb{R}^3$, $D \subset \mathbb{R}^3$ offen, definiert man die Rotation durch

$$\operatorname{rot} \mathrm{f}(\mathrm{x}^0) := \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right)^T \Bigg|_{\mathrm{x}^0}$$

Alternativ Notationen und weitere Rechenregeln.

$$\operatorname{rot} \mathrm{f}(\mathrm{x}) = \nabla \times \mathrm{f}(\mathrm{x}) = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ f_1 & f_2 & f_3 \end{vmatrix}$$

Bemerkung: Es gelten die folgenden Rechenregeln:

$$\begin{aligned} \operatorname{rot}(\alpha\,\mathrm{f} + \beta\,\mathrm{g}) &= \alpha \operatorname{rot} \mathrm{f} + \beta \operatorname{rot} \mathrm{g} \\ \operatorname{rot}(\varphi \cdot \mathrm{f}) &= (\nabla\varphi) \times \mathrm{f} + \varphi \operatorname{rot} \mathrm{f} \end{aligned}$$

Bemerkung: Ist $\varphi : D \to \mathbb{R}$, $D \subset \mathbb{R}^3$, eine $\mathcal{C}^2$–Funktion, so folgt

$$\operatorname{rot}(\nabla\varphi) = 0\,,$$

mit dem Vertauschbarkeitssatz von Schwarz, d.h. Gradientenfelder sind stets rotationsfrei.

1.2 Das vollständige Differential

Definition: Sei $D \subset \mathbb{R}^n$ offen, $x^0 \in D$ und $f : D \to \mathbb{R}^m$. Die Funktion $f(x)$ heißt differenzierbar in x^0 (oder vollständig differenzierbar bzw. total differenzierbar in x_0), falls es eine lineare Abbildung

$$l(x, x^0) := A \cdot (x - x^0)$$

mit einer Matrix $A \in \mathbb{R}^{m \times n}$ gibt, für die die Approximationseigenschaft

$$f(x) = f(x^0) + A \cdot (x - x^0) + o(\|x - x^0\|)$$

gilt, d.h.

$$\lim_{x \to x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|} = 0.$$

Das vollständige Differential und die Jacobi–Matrix.

Bezeichnungen: Man nennt die lineare Abbildung l das vollständige Differential oder das totale Differential von $\mathrm{f(x)}$ im Punkt x^0, und man bezeichnet l mit $\mathrm{df(x^0)}$.

Die zugehörige Matrix A heißt Jacobi–Matrix oder Funktionalmatrix von $\mathrm{f(x)}$ im Punkt x^0 und wird mit $\mathrm{J\,f(x^0)}$ (manchmal auch mit $\mathrm{Df(x^0)}$ oder $\mathrm{f'(x^0)}$) bezeichnet.

Bemerkung: Für $m = n = 1$ erhalten wir die bekannte Beziehung

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|)$$

für die Ableitung $f'(x_0)$ im Punkt x_0.

Bemerkung: Im Fall einer skalaren Funktion ($m = 1$) ist $\mathrm{A = a}$ ein Zeilenvektor und $\mathrm{a(x - x^0)}$ ein Skalarprodukt $\langle \mathrm{a}^T, \mathrm{x - x^0} \rangle$.

Vollständige und partielle Differenzierbarkeit.

Satz: Sei $\mathrm{f} : D \to \mathbb{R}^m$, $\mathrm{x}^0 \in D \subset \mathbb{R}^n$, D offen.

a) Ist $\mathrm{f}(\mathrm{x})$ in x^0 differenzierbar, so ist $\mathrm{f}(\mathrm{x})$ auch stetig in x^0.

b) Ist $\mathrm{f}(\mathrm{x})$ in x^0 differenzierbar, so ist das (vollständige) Differential und damit auch die Jacobi–Matrix eindeutig bestimmt und es gilt

$$\mathrm{J\,f}(\mathrm{x}^0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{x}^0) & \dots & \frac{\partial f_1}{\partial x_n}(\mathrm{x}^0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{x}^0) & \dots & \frac{\partial f_m}{\partial x_n}(\mathrm{x}^0) \end{pmatrix} = \begin{pmatrix} Df_1(\mathrm{x}^0) \\ \vdots \\ Df_m(\mathrm{x}^0) \end{pmatrix}$$

c) Ist $\mathrm{f}(\mathrm{x})$ eine $\mathcal{C}^1$–Funktion auf D, so ist $\mathrm{f}(\mathrm{x})$ auf D differenzierbar.

Beweis von a).

Ist f in x^0 differenzierbar, so gilt nach Definition

$$\lim_{x \to x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|} = 0$$

Daraus folgt aber

$$\lim_{x \to x^0} \|f(x) - f(x^0) - A \cdot (x - x^0)\| = 0$$

und wir erhalten

$$\begin{aligned} \|f(x) - f(x^0)\| &\leq \|f(x) - f(x^0) - A \cdot (x - x^0)\| + \|A \cdot (x - x^0)\| \\ &\to 0 \qquad \text{für } x \to x^0 \end{aligned}$$

Damit ist die Funktion f stetig im Punkt x^0.

Beweis von b).

Sei $x = x^0 + te_i$, $|t| < \varepsilon$, $i \in \{1, \ldots, n\}$. Da f im Punkt x^0 differenzierbar ist, folgt

$$\lim_{x \to x^0} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|_\infty} = 0$$

Wir schreiben nun

$$\begin{aligned} \frac{f(x) - f(x^0) - A \cdot (x - x^0)}{\|x - x^0\|_\infty} &= \frac{f(x^0 + te_i) - f(x^0)}{|t|} - \frac{tAe_i}{|t|} \\ &= \frac{t}{|t|} \cdot \left(\frac{f(x^0 + te_i) - f(x^0)}{t} - Ae_i \right) \\ &\to 0 \qquad \text{für } t \to 0 \end{aligned}$$

Daraus folgt

$$\lim_{t \to 0} \frac{f(x^0 + te_i) - f(x^0)}{t} = Ae_i \qquad i = 1, \ldots, n$$

Beispiele.

- Betrachte die skalare Funktion $f(x_1, x_2) = x_1 e^{2x_2}$. Dann lautet die Jacobi–Matrix:

$$Jf(x_1, x_2) = Df(x_1, x_2) = e^{2x_2}(1, 2x_1)$$

- Betrachte die Funktion $\mathrm{f} : \mathbb{R}^3 \to \mathbb{R}^2$ definiert durch

$$\mathrm{f}(x_1, x_2, x_3) = \begin{pmatrix} x_1 x_2 x_3 \\ \sin(x_1 + 2x_2 + 3x_3) \end{pmatrix}$$

Die Jacobi–Matrix ergibt sich in der Form

$$\mathrm{Jf}(x_1, x_2, x_3) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_1}{\partial x_3} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \dfrac{\partial f_2}{\partial x_3} \end{pmatrix} = \begin{pmatrix} x_2 x_3 & x_1 x_3 & x_1 x_2 \\ \cos(s) & 2\cos(s) & 3\cos(s) \end{pmatrix}$$

wobei $s = x_1 + 2x_2 + 3x_3$.

Weitere Beispiele.

- Sei $f(x) = Ax$, $A \in \mathbb{R}^{m\times n}$ und $x \in \mathbb{R}^n$. Dann gilt

$$Jf(x) = A \qquad \text{für alle } x \in \mathbb{R}^n$$

- Sei $f(x) = x^T Ax = \langle x, Ax\rangle$, $A \in \mathbb{R}^{n\times n}$ und $x \in \mathbb{R}^n$. Dann gilt

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= \langle e_i, Ax\rangle + \langle x, Ae_i\rangle \\ &= e_i^T Ax + x^T Ae_i \\ &= x^T(A^T + A)e_i \end{aligned}$$

Daraus folgt

$$Jf(x) = \operatorname{grad} f(x) = x^T(A^T + A)$$

Differentiationsregeln.

Satz:

a) **Linearität:** Sind $\mathrm{f}, \mathrm{g} : D \to \mathbb{R}^m$ differenzierbar in $\mathrm{x}^0 \in D$, D offen, so ist auch $\alpha\, \mathrm{f}(\mathrm{x}^0) + \beta\, \mathrm{g}(\mathrm{x}^0)$, $\alpha, \beta \in \mathbb{R}$, differenzierbar in x^0 und es gilt

$$\begin{aligned} \mathrm{d}(\alpha \mathrm{f} + \beta \mathrm{g})(\mathrm{x}^0) &= \alpha\, \mathrm{df}(\mathrm{x}^0) + \beta\, \mathrm{dg}(\mathrm{x}^0) \\ \mathrm{J}(\alpha \mathrm{f} + \beta \mathrm{g})(\mathrm{x}^0) &= \alpha\, \mathrm{Jf}(\mathrm{x}^0) + \beta\, \mathrm{Jg}(\mathrm{x}^0) \end{aligned}$$

b) **Kettenregel:** Ist $\mathrm{f} : D \to \mathbb{R}^m$ differenzierbar in $\mathrm{x}^0 \in D$, D offen, und ist $\mathrm{g} : E \to \mathbb{R}^k$ differenzierbar in $\mathrm{y}^0 = f(\mathrm{x}^0) \in E \subset \mathbb{R}^m$, E offen, so ist $g \circ f$ ebenfalls in x^0 differenzierbar.

Für die Differentiale gilt

$$\mathrm{d}(\mathrm{g} \circ \mathrm{f})(\mathrm{x}^0) = \mathrm{dg}(\mathrm{y}^0) \circ \mathrm{df}(\mathrm{x}^0)$$

und analog für die Jacobi–Matrizen

$$\mathrm{J}(\mathrm{g} \circ \mathrm{f})(\mathrm{x}^0) = \mathrm{Jg}(\mathrm{y}^0) \cdot \mathrm{Jf}(\mathrm{x}^0)$$

Beispiel zur Kettenregel.

Sei $\mathbf{h} : I \to \mathbb{R}^n$, $I \subset \mathbb{R}$ Intervall, eine in $t_0 \in I$ differenzierbare Kurve mit Werten in $D \subset \mathbb{R}^n$, D offen, und $f : D \to \mathbb{R}$ eine in $\mathbf{x}^0 = \mathbf{h}(t_0)$ differenzierbare skalare Funktion.

Dann ist auch die Hintereinanderausführung

$$(f \circ \mathbf{h})(t) = f(h_1(t), \ldots, h_n(t))$$

in t_0 differenzierbar, und für die Ableitung gilt:

$$\begin{aligned}
(f \circ \mathbf{h})'(t_0) &= \mathrm{J}f(\mathbf{h}(t_0)) \cdot \mathrm{J}\mathbf{h}(t_0) \\
&= \operatorname{grad} f(\mathbf{h}(t_0)) \cdot \mathbf{h}'(t_0) \\
&= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{h}(t_0)) \cdot h_k'(t_0)
\end{aligned}$$

Richtungsableitungen.

Definition: Sei $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, $\mathrm{x}^0 \in D$, und $\mathrm{v} \in \mathbb{R} \setminus \{0\}$ ein Vektor. Dann heißt

$$D_{\mathrm{v}} f(\mathrm{x}^0) := \lim_{t \to 0} \frac{f(\mathrm{x}^0 + t\mathrm{v}) - f(\mathrm{x}^0)}{t}$$

die Richtungsableitung (Gateaux–Ableitung) von $f(\mathrm{x})$ in Richtung v.

Beispiel: Sei $f(x, y) = x^2 + y^2$ und $\mathrm{v} = (1, 1)^T$. Dann gilt für die Richtungsableitung in Richtung v:

$$\begin{aligned} D_{\mathrm{v}} f(x, y) &= \lim_{t \to 0} \frac{(x + t)^2 + (y + t)^2 - x^2 - y^2}{t} \\ &= \lim_{t \to 0} \frac{2xt + t^2 + 2yt + t^2}{t} \\ &= 2(x + y) \end{aligned}$$

Bemerkungen.

- Für $\mathbf{v} = \mathbf{e}_i$ ist die Richtungsableitung in Richtung $\mathbf{v}$ gegeben durch die partielle Ableitung nach der Koordinatenrichtung x_i:

$$D_{\mathbf{v}} f(\mathbf{x}^0) = \frac{\partial f}{\partial x_i}(\mathbf{x}^0)$$

- Ist $\mathbf{v}$ ein Einheitsvektor, also $\|\mathbf{v}\| = 1$, so beschreibt die Richtungsableitung $D_{\mathbf{v}} f(\mathbf{x}^0)$ den Anstieg (bzw. die Steigung) von $f(\mathbf{x})$ in Richtung $\mathbf{v}$.

- Ist $f(\mathbf{x})$ in $\mathbf{x}^0$ differenzierbar, so existieren sämtliche Richtungsableitungen von $f(\mathbf{x})$ in $\mathbf{x}^0$ und mit $\mathbf{h}(t) = \mathbf{x}^0 + t\mathbf{v}$ gilt

$$D_{\mathbf{v}} f(\mathbf{x}^0) = \frac{d}{dt}(f \circ \mathbf{h})|_{t=0} = \operatorname{grad} f(\mathbf{x}^0) \cdot \mathbf{v}$$

  Dies folgt unmittelbar aus der Anwendung der Kettenregel.

Eigenschaften des Gradienten.

Satz: Sei $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ offen, in $\mathrm{x}^0 \in D$ differenzierbar. Dann gilt

a) Der Gradientenvektor $\operatorname{grad} f(\mathrm{x}^0) \in \mathbb{R}^n$ steht senkrecht auf der Niveaumenge

$$N_{\mathrm{x}^0} := \{\mathrm{x} \in D \mid f(\mathrm{x}) = f(\mathrm{x}^0)\}$$

Im Fall $n = 2$ nennt man die Niveaumengen auch Höhenlinien, im Fall $n = 3$ heißen die Niveaumengen auch Äquipotentialflächen.

2) Der Gradient $\operatorname{grad} f(\mathrm{x}^0)$ gibt die Richtung des steilsten Anstiegs von $f(\mathrm{x})$ in x^0 an.

Beweisidee:

a) Anwendung der Kettenregel.

b) Für beliebige Richtung v gilt mit der Cauchy–Schwarzschen Ungleichung

$$|D_{\mathrm{v}} f(\mathrm{x}^0)| = |(\operatorname{grad} f(\mathrm{x}^0), \mathrm{v})| \ \leq \ \|\operatorname{grad} f(\mathrm{x}^0)\|_2$$

Gleichheit wird für $\mathrm{v} = \operatorname{grad} f(\mathrm{x}^0)/\|\operatorname{grad} f(\mathrm{x}^0)\|_2$ angenommen.

Krummlinige Koordinaten.

Definition: Sei $\Phi : U \to V$, $U, V \subset \mathbb{R}^n$ offen, eine $\mathcal{C}^1$–Abbildung, für die die Jacobimatrix $\mathrm{J}\Phi(\mathrm{u}^0)$ an jeder Stelle $\mathrm{u}^0 \in U$ regulär ist.

Weiterhin existiere die Umkehrabbildung $\Phi^{-1} : V \to U$ und diese sei ebenfalls eine $\mathcal{C}^1$–Abbildung.

Dann definiert $\mathrm{x} = \Phi(\mathrm{u})$ eine Koordinatentransformation von den Koordinaten u auf x.

Beispiel: Betrachte für $n = 2$ die Polarkoordinaten $\mathrm{u} = (r, \varphi)$ mit $r > 0$ und $-\pi < \varphi < \pi$ und setze

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

mit den kartesischen Koordinaten $\mathrm{x} = (x, y)$.

Umrechnung der partiellen Ableitungen.

Für alle $u \in U$ mit $x = \Phi(u)$ gelten die Relationen

$$\begin{aligned} \Phi^{-1}(\Phi(u)) &= u \\ J\,\Phi^{-1}(x) \cdot J\,\Phi(u) &= I_n \qquad \text{(Kettenregel)} \\ J\,\Phi^{-1}(x) &= (J\,\Phi(u))^{-1} \end{aligned}$$

Sei nun $\tilde{f} : V \to \mathbb{R}$ eine gegebene Funktion und setze

$$f(u) := \tilde{f}(\Phi(u))$$

Dann folgt aus der Kettenregel:

$$\frac{\partial f}{\partial u_i} = \sum_{j=1}^{n} \frac{\partial \tilde{f}}{\partial x_j} \frac{\partial \Phi_j}{\partial u_i} =: \sum_{j=1}^{n} g^{ij} \frac{\partial \tilde{f}}{\partial x_j}$$

mit

$$g^{ij} := \frac{\partial \Phi_j}{\partial u_i}, \qquad G(u) := (g^{ij}) = (J\,\Phi(u))^T$$

Notationen.

Wir verwenden die abkürzende Schreibweise

$$\frac{\partial}{\partial u_i} = \sum_{j=1}^{n} g^{ij} \frac{\partial}{\partial x_j}$$

Analog lassen sich die partiellen Ableitungen nach x_i durch die partiellen Ableitungen nach u_j ausdrücken mit

$$\frac{\partial}{\partial x_i} = \sum_{j=1}^{n} g_{ij} \frac{\partial}{\partial u_j}$$

wobei

$$(g_{ij}) := (g^{ij})^{-1} = (\mathrm{J}\,\Phi)^{-T} = (\mathrm{J}\,\Phi^{-1})^{T}$$

Man erhält diese Beziehungen durch Anwendung der Kettenregel auf Φ^{-1}.

Beispiel: Polarkoordinaten.

Wir betrachten die Polarkoordinaten

$$\mathrm{x} = \Phi(\mathrm{u}) = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}$$

Dann berechnet man

$$\mathrm{J}\,\Phi(\mathrm{u}) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

und damit

$$(g^{ij}) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -r\sin\varphi & r\cos\varphi \end{pmatrix} \qquad (g_{ij}) = \begin{pmatrix} \cos\varphi & -\frac{1}{r}\sin\varphi \\ \sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix}$$

Partielle Ableitungen für die Polarkoordinaten.

Für die Umrechnung der partiellen Ableitungen bekommt man nun

$$\begin{aligned}\frac{\partial}{\partial x} &= \cos\varphi\,\frac{\partial}{\partial r} - \frac{1}{r}\sin\varphi\,\frac{\partial}{\partial\varphi}\\ \frac{\partial}{\partial y} &= \sin\varphi\,\frac{\partial}{\partial r} + \frac{1}{r}\cos\varphi\,\frac{\partial}{\partial\varphi}\end{aligned}$$

Beispiel: Umrechnung des Laplace–Operator auf Polarkoordinaten

$$\begin{aligned}\frac{\partial^2}{\partial x^2} &= \cos^2\varphi\,\frac{\partial^2}{\partial r^2} - \frac{\sin(2\varphi)}{r}\,\frac{\partial^2}{\partial r\partial\varphi} + \frac{\sin^2\varphi}{r^2}\,\frac{\partial^2}{\partial\varphi^2} + \frac{\sin(2\varphi)}{r^2}\,\frac{\partial}{\partial\varphi} + \frac{\sin^2\varphi}{r}\,\frac{\partial}{\partial r}\\ \frac{\partial^2}{\partial y^2} &= \sin^2\varphi\,\frac{\partial^2}{\partial r^2} + \frac{\sin(2\varphi)}{r}\,\frac{\partial^2}{\partial r\partial\varphi} + \frac{\cos^2\varphi}{r^2}\,\frac{\partial^2}{\partial\varphi^2} - \frac{\sin(2\varphi)}{r^2}\,\frac{\partial}{\partial\varphi} + \frac{\cos^2\varphi}{r}\,\frac{\partial}{\partial r}\\ \Delta &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\,\frac{\partial^2}{\partial\varphi^2} + \frac{1}{r}\,\frac{\partial}{\partial r}\end{aligned}$$

Beispiel: Kugelkoordinaten.

Wir betrachten die Kugelkoordinaten

$$x = \Phi(u) = \begin{pmatrix} r\cos\varphi\cos\theta \\ r\sin\varphi\cos\theta \\ r\sin\theta \end{pmatrix}$$

Die Jacobi–Matrix ist dann gegeben durch:

$$J\,\Phi(u) = \begin{pmatrix} \cos\varphi\cos\theta & -r\sin\varphi\cos\theta & -r\cos\varphi\sin\theta \\ \sin\varphi\cos\theta & r\cos\varphi\cos\theta & -r\sin\varphi\sin\theta \\ \sin\theta & 0 & r\cos\theta \end{pmatrix}$$

Partielle Ableitungen für die Kugelkoordinaten.

Für die Umrechnung der partiellen Ableitungen bekommt man nun

$$\frac{\partial}{\partial x} \quad = \quad \cos\varphi\,\cos\theta\,\frac{\partial}{\partial r} - \frac{\sin\varphi}{r\cos\theta}\,\frac{\partial}{\partial\varphi} - \frac{1}{r}\,\cos\varphi\,\sin\theta\,\frac{\partial}{\partial\theta}$$

$$\frac{\partial}{\partial y} \quad = \quad \sin\varphi\,\cos\theta\,\frac{\partial}{\partial r} + \frac{\cos\varphi}{r\cos\theta}\,\frac{\partial}{\partial\varphi} - \frac{1}{r}\,\sin\varphi\,\sin\theta\,\frac{\partial}{\partial\theta}$$

$$\frac{\partial}{\partial z} \quad = \quad \sin\theta\,\frac{\partial}{\partial r} + \frac{1}{r}\,\cos\theta\,\frac{\partial}{\partial\theta}$$

Beispiel: Umrechnung des Laplace–Operators auf Kugelkoordinaten

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2\cos^2\theta}\,\frac{\partial^2}{\partial\varphi^2} + \frac{1}{r^2}\,\frac{\partial^2}{\partial\theta^2} + \frac{2}{r}\,\frac{\partial}{\partial r} - \frac{\tan\theta}{r^2}\,\frac{\partial}{\partial\theta}$$

1.3 Mittelwertsätze und Taylor–Entwicklungen

Satz (Mittelwertsatz): Sei $f : D \to \mathbb{R}$ eine auf einer offenen Menge $D \subset \mathbb{R}^n$ differenzierbare, skalare Funktion. Weiterhin seien $\mathrm{a}, \mathrm{b} \in D$ Punkte in D, so dass die Verbindungsstrecke

$$[\mathrm{a}, \mathrm{b}] := \{\mathrm{a} + t(\mathrm{b} - \mathrm{a}) \,|\, t \in [0, 1]\}$$

ganz in D liegt. Dann gibt es eine Zahl $\theta \in (0, 1)$ mit

$$f(\mathrm{b}) - f(\mathrm{a}) = \operatorname{grad} f(\mathrm{a} + \theta(\mathrm{b} - \mathrm{a})) \cdot (\mathrm{b} - \mathrm{a})$$

Beweis: Wir setzen

$$h(t) := f(\mathrm{a} + t(\mathrm{b} - \mathrm{a}))$$

Aus dem Mittelwertsatz für eine Veränderliche folgt dann mit der Kettenregel

$$\begin{aligned} f(\mathrm{b}) - f(\mathrm{a}) &= h(1) - h(0) = h'(\theta) \cdot (1 - 0) \\ &= \operatorname{grad} f(\mathrm{a} + \theta(\mathrm{b} - \mathrm{a})) \cdot (\mathrm{b} - \mathrm{a}) \end{aligned}$$

Definition und Beispiel.

Definition: Gilt die Bedingung $[a, b] \subset D$ für **alle** Punkte $a, b \in D$, so heißt die Menge D konvex.

Beispiel zum Mittelwertsatz: Gegeben sei die skalare Funktion

$$f(x, y) := \cos x + \sin y$$

Offensichtlich gilt

$$f(0, 0) = f(\pi/2, \pi/2) = 1 \quad \Rightarrow \quad f(\pi/2, \pi/2) - f(0, 0) = 0$$

Nach dem Mittelwertsatz existiert ein $\theta \in (0, 1)$ mit

$$\text{grad}\, f\left(\theta\begin{pmatrix}\pi/2\\ \pi/2\end{pmatrix}\right) \cdot \begin{pmatrix}\pi/2\\ \pi/2\end{pmatrix} = 0$$

In der Tat gilt diese Beziehung für $\theta = \frac{1}{2}$.

Mittelwertsatz gilt nur für skalare Funktionen.

Beachte: Der Mittelwertsatz für mehrere Variablen gilt nur für skalare Funktionen, aber i.A. nicht für vektorwertige Funktionen!

Beispiel: Betrachte die **vektorwertige** Funktion

$$\mathrm{f}(t) := \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \qquad t \in [0, \pi/2]$$

Nun gilt

$$\mathrm{f}(\pi/2) - \mathrm{f}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

und

$$\mathrm{f}'\left(\theta \, \frac{\pi}{2}\right) \cdot \left(\frac{\pi}{2} - 0\right) = \frac{\pi}{2} \begin{pmatrix} -\sin(\theta\pi/2) \\ \cos(\theta\pi/2) \end{pmatrix}$$

ABER: Die Vektoren auf der rechten Seite haben die Längen $\sqrt{2}$ bzw. $\pi/2$!

Der Mittelwert–Abschätzungssatz für vektorwertige Funktionen.

Satz: Die Funktion $\mathrm{f} : D \to \mathbb{R}^m$ sei differenzierbar auf der offenen Menge $D \subset \mathbb{R}^n$. Weiterhin seien a, b Punkte in D mit $[\mathrm{a}, \mathrm{b}] \subset D$. Dann gibt es ein $\theta \in (0, 1)$ mit

$$\|\mathrm{f}(\mathrm{b}) - \mathrm{f}(\mathrm{a})\|_2 \leq \|\mathrm{J}\,\mathrm{f}(\mathrm{a} + \theta(\mathrm{b} - \mathrm{a})) \cdot (\mathrm{b} - \mathrm{a})\|_2$$

Beweisidee: Anwendung des Mittelwertsatzes auf die skalare Funktion $g(\mathrm{x})$ definiert durch

$$g(\mathrm{x}) := (\mathrm{f}(\mathrm{b}) - \mathrm{f}(\mathrm{a}))^T \mathrm{f}(\mathrm{x}) \qquad \text{(Skalarprodukt!)}$$

Bemerkung: Eine andere (abgeschwächte) Form der Mittelwert–Abschätzung ist

$$\|\mathrm{f}(\mathrm{b}) - \mathrm{f}(\mathrm{a})\| \leq \sup_{\xi \in [\mathrm{a},\mathrm{b}]} \|\mathrm{J}\,\mathrm{f}(\xi))\| \cdot \|(\mathrm{b} - \mathrm{a})\|$$

wobei $\|\cdot\|$ eine beliebige Vektor– bzw. zugehörige Matrixnorm ist.

Taylor–Entwicklungen: Notationen.

Zunächst definieren wir einen Multiindex $\alpha \in \mathbb{N}_0^n$ als

$$\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$$

Weiterhin sei

$$|\alpha| := \alpha_1 + \cdots + \alpha_n \qquad \alpha! := \alpha_1! \cdot \cdots \cdot \alpha_n!$$

Ist $f : D \to \mathbb{R}$ $|\alpha|$–mal stetig differenzierbar, so setzen wir

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} \ldots D_n^{\alpha_n} = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}},$$

wobei $D_i^{\alpha_i} = \underbrace{D_i \ldots D_i}_{\alpha_i\text{–mal}}$ und wir schreiben

$$\mathrm{x}^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} \qquad \text{für } \mathrm{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Der Satz von Taylor.

Satz: (Satz von Taylor)
Sei $D \subset \mathbb{R}^n$ offen und konvex, $f : D \to \mathbb{R}$ eine $\mathcal{C}^{m+1}$–Funktion und sei $x_0 \in D$. Dann gilt für $x \in D$ die Taylor–Entwicklung

$$
\begin{aligned}
f(x) &= T_m(x; x_0) + R_m(x; x_0) \\
T_m(x; x_0) &= \sum_{|\alpha| \leq m} \frac{D^\alpha f(x_0)}{\alpha!}(x - x_0)^\alpha \\
R_m(x; x_0) &= \sum_{|\alpha| = m+1} \frac{D^\alpha f(x_0 + \theta(x - x_0))}{\alpha!}(x - x_0)^\alpha
\end{aligned}
$$

mit einem geeigneten $\theta \in (0, 1)$.

Bezeichnung: In der obigen Taylor–Entwicklung heißt $T_m(x; x_0)$ Taylor–Polynom m–ten Grades und $R_m(x; x_0)$ wird als Lagrange–Restglied bezeichnet.

Herleitung der Taylorschen Formel.

Wir definieren eine skalare Funktion einer Variablen $t \in [0,1]$ als

$$g(t) := f(\mathrm{x}_0 + t(\mathrm{x} - \mathrm{x}_0))$$

und berechnen die Taylor–Entwicklung um $t = 0$. Es gilt:

$$g(1) = g(0) + g'(0) \cdot (1-0) + \frac{1}{2} g''(\xi) \cdot (1-0)^2 \quad \text{für ein } \xi \in (0,1).$$

Die Berechnung von $g'(0)$ liefert mit der Kettenregel

$$\begin{aligned} g'(0) &= \frac{d}{dt} f(x_1^0 + t(x_1 - x_1^0), x_2^0 + t(x_2 - x_2^0), \ldots, x_n^0 + t(x_n - x_n^0))\Big|_{t=0} \\ &= D_1 f(\mathrm{x}_0) \cdot (x_1 - x_1^0) + \ldots + D_n f(\mathrm{x}_0) \cdot (x_n - x_n^0) \\ &= \sum_{|\alpha|=1} \frac{D^\alpha f(\mathrm{x}_0)}{\alpha!} \cdot (\mathrm{x} - \mathrm{x}_0)^\alpha \end{aligned}$$

Fortsetzung der Herleitung.

Berechnung von $g''(0)$ liefert

$$g''(0) = \frac{d^2}{dt^2} f(\mathrm{x}_0 + t(\mathrm{x} - \mathrm{x}_0))\Big|_{t=0} = \frac{d}{dt} \sum_{k=1}^{n} D_k f(\mathrm{x}^0 + t(\mathrm{x} - \mathrm{x}^0))(x_k - x_k^0)\Big|_{t=0}$$

$$= D_{11} f(\mathrm{x}_0)(x_1 - x_1^0)^2 + D_{21} f(\mathrm{x}_0)(x_1 - x_1^0)(x_2 - x_2^0)$$

$$+ \ldots + D_{ij} f(\mathrm{x}_0)(x_i - x_i^0)(x_j - x_j^0) + \ldots +$$

$$+ D_{n-1,n} f(\mathrm{x}_0)(x_{n-1} - x_{n-1}^0)(x_n - x_n^0) + D_{nn} f(\mathrm{x}_0)(x_n - x_n^0)^2)$$

$$= \sum_{|\alpha|=2} \frac{D^\alpha f(\mathrm{x}_0)}{\alpha!}(\mathrm{x} - \mathrm{x}_0)^\alpha \qquad \text{(Vertauschungssatz von Schwarz!)}$$

Nun: Beweis der Taylor–Formel mittels vollständiger Induktion!

Beweis des Satzes von Taylor.

Die Funktion

$$g(t) := f(\mathrm{x}^0 + t(\mathrm{x} - \mathrm{x}^0))$$

ist $(m+1)$–mal stetig differenzierbar, und es gilt

$$g(1) = \sum_{k=0}^{m} \frac{g^{(k)}(0)}{k!} + \frac{g^{(m+1)}(\theta)}{(m+1)!} \quad \text{für ein } \theta \in [0,1].$$

Weiterhin gilt (per Induktion über k)

$$\frac{g^{(k)}(0)}{k!} = \sum_{|\alpha|=k} \frac{D^\alpha f(\mathrm{x}^0)}{\alpha!} (\mathrm{x} - \mathrm{x}^0)^\alpha$$

und

$$\frac{g^{(m+1)}(\theta)}{(m+1)!} = \sum_{|\alpha|=m+1} \frac{D^\alpha f(\mathrm{x}^0 + \theta(\mathrm{x} - \mathrm{x}^0))}{\alpha!} (\mathrm{x} - \mathrm{x}^0)^\alpha$$

Beispiel zur Taylor–Entwicklung.

1. Berechne das Taylor–Polynom $T_2(\mathrm{x}; \mathrm{x}_0)$ zweiten Grades der Funktion

$$f(x, y, z) = x\,y^2 \sin z$$

zum Entwicklungspunkt $(x, y, z) = (1, 2, 0)^T$.

2. Die Berechnung von $T_2(\mathrm{x}; \mathrm{x}_0)$ benötigt die partiellen Ableitungen bis zur zweiten Ordnung.

3. Diese Ableitungen müssen am Punkt $(x, y, z) = (1, 2, 0)^T$ ausgewertet werden.

4. Als Ergebnis erhält man $T_2(\mathrm{x}; \mathrm{x}_0)$ in der Form

$$T_2(\mathrm{x}; \mathrm{x}_0) = 4z(x + y - 2)$$

5. Berechnung auf Folie.

Bemerkung zum Restglied eines Taylor–Polynoms.

Bemerkung: Das Restglied eine Taylor–Polynoms enthält **alle** partiellen Ableitungen der Ordnung $(m+1)$:

$$R_m(x; x_0) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(x_0 + \theta(x - x_0))}{\alpha!}(x - x_0)^\alpha$$

Sind all diese Ableitungen in der Nähe von x_0 durch eine Konstante C beschränkt, so gilt die Restgliedabschätzung

$$|R_m(x; x_0)| \leq \frac{n^{m+1}}{(m+1)!} C \, \|x - x_0\|_\infty^{m+1}$$

Für die Approximationsgüte des Taylor–Polynoms einer $\mathcal{C}^{m+1}$–Funktion folgt daher

$$f(x) = T_m(x; x_0) + O\left(\|x - x_0\|^{m+1}\right)$$

Spezialfall $m = 1$: Für eine $\mathcal{C}^2$–Funktion $f(x)$ bekommt man

$$f(x) = f(x^0) + \operatorname{grad} f(x^0) \cdot (x - x^0) + O(\|x - x^0\|^2).$$

Die Hesse–Matrix.

Man nennt die Matrix

$$\mathrm{H}f(\mathrm{x}_0) := \begin{pmatrix} f_{x_1x_1}(\mathrm{x}_0) & \dots & f_{x_1x_n}(\mathrm{x}_0) \\ \vdots & & \vdots \\ f_{x_nx_1}(\mathrm{x}_0) & \dots & f_{x_nx_n}(\mathrm{x}_0) \end{pmatrix}$$

die Hesse–Matrix von $f(\mathrm{x})$ im Punkt x_0.

$$\text{Hesse–Matrix} = \text{Jacobi–Matrix des Gradienten } \nabla f$$

Die Taylor–Entwicklung einer $\mathcal{C}^3$–Funktion lautet daher

$$f(\mathrm{x}) = f(\mathrm{x}_0) + \operatorname{grad} f(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0) + \frac{1}{2}(\mathrm{x}-\mathrm{x}_0)^T \mathrm{H}f(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0) + O(\|\mathrm{x}-\mathrm{x}_0\|^3)$$

Die Hesse–Matrix einer $\mathcal{C}^2$–Funktion ist symmetrisch.

2.1 Extrema von Funktionen mehrerer Veränderlichen

Definition: Sei $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$, und $\mathrm{x}^0 \in D$. Dann hat $f(\mathrm{x})$ in x^0

- ein globales Maximum, falls $f(\mathrm{x}) \le f(\mathrm{x}^0)$ für alle $\mathrm{x} \in D$.
- ein strenges globales Maximum, falls $f(\mathrm{x}) < f(\mathrm{x}^0)$ für alle $\mathrm{x} \in D$.
- ein lokales Maximum, falls es ein $\varepsilon > 0$ gibt mit

$$f(\mathrm{x}) \le f(\mathrm{x}^0) \qquad \text{für alle } \mathrm{x} \in D \text{ mit } \|\mathrm{x} - \mathrm{x}^0\| < \varepsilon.$$

- ein strenges lokales Maximum, falls es ein $\varepsilon > 0$ gibt mit

$$f(\mathrm{x}) < f(\mathrm{x}^0) \qquad \text{für alle } \mathrm{x} \in D \text{ mit } \|\mathrm{x} - \mathrm{x}^0\| < \varepsilon.$$

Analoge Definitionen für Minima.

Notwendige Bedingung für lokale Extrema.

Satz: Besitzt eine $\mathcal{C}^1$–Funktion $f(\mathrm{x})$ in einem Punkt $\mathrm{x}^0 \in D^0$ ein lokales Extremum (Minimum oder Maximum), so gilt

$$\operatorname{grad} f(\mathrm{x}^0) = 0 \in \mathbb{R}^n$$

Beweis: Für ein beliebiges $\mathrm{v} \in \mathbb{R}^n$, $\mathrm{v} \neq 0$ ist die Funktion

$$\varphi(t) := f(\mathrm{x}^0 + t\mathrm{v})$$

in einer Umgebung von $t^0 = 0$ stetig differenzierbar.

Weiterhin hat $\varphi(t)$ bei $t^0 = 0$ ein lokales Extremum. Damit folgt:

$$\varphi'(0) = \operatorname{grad} f(\mathrm{x}^0)\,\mathrm{v} = 0$$

Da dies für alle $\mathrm{v} \neq 0$ gilt, folgt die Bedingung:

$$\operatorname{grad} f(\mathrm{x}^0) = (0, \ldots, 0)^T$$

Bemerkungen zu lokalen Extremwerten.

Bemerkungen:

- Die Bedingung $\operatorname{grad} f(\mathrm{x}^0) = 0$ liefert gewöhnlich ein nichtlineares Gleichungssystem zur Berechnung von $\mathrm{x} = \mathrm{x}^0$ mit n Gleichungen und n Unbekannten.

- Die Punkte $\mathrm{x}^0 \in D^0$ mit $\operatorname{grad} f(\mathrm{x}^0) = 0$ nennt man stationäre Punkte von $f(x)$. Stationäre Punkte sind **nicht** notwendigerweise lokale Extremwerte. Zum Beispiel besitzt die Funktion

$$f(x, y) := x^2 - y^2$$

den Gradienten

$$\operatorname{grad} f(x, y) = 2(x, -y)$$

und hat daher nur einen stationären Punkt $\mathrm{x}^0 = (0, 0)^T$. Der Punkt x^0 ist jedoch ein Sattelpunkt von f, d.h. in jeder Umgebung von x^0 gibt es zwei Punkte x^1 und x^2 mit

$$f(\mathrm{x}^1) < f(\mathrm{x}^0) < f(\mathrm{x}^2).$$

Klassifikation stationärer Punkte.

Satz: Sei $f(\mathrm{x})$ eine $\mathcal{C}^2$–Funktion auf D^0 und $\mathrm{x}^0 \in D^0$ ein stationärer Punkt von $f(\mathrm{x})$, d.h. $\operatorname{grad} f(\mathrm{x}^0) = 0$.

a) Notwendige Bedingung

Ist x^0 ein lokales Extremum von $f(\mathrm{x})$, so gilt:

$$\mathrm{x}^0 \text{ lokales Minimum } \Rightarrow \mathrm{H}\, f(\mathrm{x}^0) \text{ positiv semidefinit}$$
$$\mathrm{x}^0 \text{ lokales Maximum } \Rightarrow \mathrm{H}\, f(\mathrm{x}^0) \text{ negativ semidefinit}$$

b) Hinreichende Bedingung

Ist $\mathrm{H}\, f(\mathrm{x}^0)$ positiv definit (bzw. negativ definit), so ist x^0 ein strenges lokales Minimum (bzw. Maximum) von $f(\mathrm{x})$.

Ist $\mathrm{H}\, f(\mathrm{x}^0)$ indefinit, so ist x^0 ein Sattelpunkt, d.h. es gibt in jeder Umgebung von x^0 Punkte x^1 und x^2 mit $f(\mathrm{x}^1) < f(\mathrm{x}^0) < f(\mathrm{x}^2)$.

Beweis des Satzes, Teil a).

Sei x^0 ein lokales Minimum. Für $v \neq 0$ und $\varepsilon > 0$ hinreichend klein folgt aus der Taylor–Formel

$$f(x^0 + \varepsilon v) - f(x^0) = \frac{1}{2}(\varepsilon v)^T \mathrm{H}\, f(x^0 + \theta\varepsilon v)(\varepsilon v) \geq 0 \qquad (1)$$

mit $\theta = \theta(\varepsilon, v) \in (0, 1)$.

Der Gradient in der Taylorentwicklung verschwindet, $\operatorname{grad} f(x^0) = 0$, denn x^0 ist stationär.

Aus (1) folgt

$$v^T \mathrm{H}\, f(x^0 + \theta\varepsilon v)v \geq 0 \qquad (2)$$

Da $f(x)$ eine $\mathcal{C}^2$–Funktion ist, ist die Hesse–Matrix eine stetige Abbildung. Im Grenzwert $\varepsilon \to 0$ folgt daher aus (2),

$$v^T \mathrm{H}\, f(x^0)v \geq 0$$

d.h. $\mathrm{H}\, f(x^0)$ ist positiv semidefinit.

Beweis des Satzes, Teil b).

Ist $\mathrm{H}\,f(\mathrm{x}^0)$ positiv definit, so ist $\mathrm{H}\,f(\mathrm{x})$ ebenfalls in einer hinreichend kleinen Umgebung $\mathrm{x} \in K_\varepsilon(\mathrm{x}^0) \subset D$ um x^0 positiv definit. Dies folgt aus der Stetigkeit der zweiten partiellen Ableitungen.

Für $\mathrm{x} \in K_\varepsilon(\mathrm{x}^0)$, $\mathrm{x} \neq \mathrm{x}^0$ gilt damit

$$\begin{aligned} f(\mathrm{x}) - f(\mathrm{x}^0) &= \frac{1}{2}(\mathrm{x}-\mathrm{x}^0)^T \mathrm{H}\,f(\mathrm{x}^0 + \theta(\mathrm{x}-\mathrm{x}^0))(\mathrm{x}-\mathrm{x}^0) \\ &> 0 \end{aligned}$$

mit $\theta \in (0,1)$, d.h. $f(\mathrm{x})$ hat in x^0 ein strenges lokales Minimum.

Ist $\mathrm{H}\,f(\mathrm{x}^0)$ indefinit, so existieren zu Eigenwerten von $\mathrm{H}\,f(\mathrm{x}^0)$ mit verschiedenen Vorzeichen gewisse Eigenvektoren v, w mit

$$\mathrm{v}^T \mathrm{H} f(\mathrm{x}^0)\mathrm{v} > 0 \qquad \mathrm{w}^T \mathrm{H} f(\mathrm{x}^0)\mathrm{w} < 0$$

und somit ist x^0 ein Sattelpunkt.

Bemerkungen.

- Ein stationärer Punkt x^0 mit $\det Hf(x^0) = 0$ heißt ausgeartet. Die Hesse–Matrix besitzt dann den Eigenwert $\lambda = 0$.
- Ist x^0 **nicht** ausgeartet, so gibt es 3 Fälle für die Eigenwerte von $Hf(x^0)$:

$$\begin{aligned}
\text{alle EW sind strikt positiv} \quad &\Rightarrow \quad x^0 \text{ ist strenges lokales Minimum} \\
\text{alle EW sind strikt negativ} \quad &\Rightarrow \quad x^0 \text{ ist strenges lokales Maximum} \\
\text{es gibt strikt pos. und neg. EW} \quad &\Rightarrow \quad x^0 \text{ Sattelpunkt}
\end{aligned}$$

- Die folgenden Implikationen gelten (aber für keine die Umkehrung)

$$\begin{array}{ccc}
x^0 \text{ lokales Minimum} & \Leftarrow & x^0 \text{ strenges lokales Minimum} \\
\Downarrow & & \Uparrow \\
Hf(x^0) \text{ positiv semidefinit} & \Leftarrow & Hf(x^0) \text{ positiv definit}
\end{array}$$

Weitere Bemerkung.

- Ist $f(\mathrm{x})$ eine $\mathcal{C}^3$–Funktion, x^0 ein stationärer Punkt von $f(\mathrm{x})$ und $\mathrm{H}f(\mathrm{x}^0)$ positiv definit, so gilt die Abschätzung:

$$(\mathrm{x}-\mathrm{x}^0)^T\,\mathrm{H}f(\mathrm{x}^0)\,(\mathrm{x}-\mathrm{x}^0) \geq \lambda_{min}\cdot\|\mathrm{x}-\mathrm{x}^0\|^2$$

wobei λ_{min} den kleinsten Eigenwert der Hesse–Matrix bezeichnet.

Nach dem Satz von Taylor gilt dann:

$$\begin{aligned} f(\mathrm{x})-f(\mathrm{x}^0) &\geq \frac{1}{2}\lambda_{min}\|\mathrm{x}-\mathrm{x}^0\|^2 + R_3(\mathrm{x};\mathrm{x}^0) \\ &\geq \|\mathrm{x}-\mathrm{x}^0\|^2\left(\frac{\lambda_{min}}{2} - C\|\mathrm{x}-\mathrm{x}^0\|\right) \end{aligned}$$

mit einer geeigneten Konstanten $C>0$.

Um x^0 wächst $f(\mathrm{x})$ somit mindestens quadratisch mit dem Abstand von x^0.

Beispiel.

Wir betrachten die Funktion

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

und suchen die stationären Punkte:

$$\operatorname{grad} f(x,y) = (y^2 + x(3x+2), 2y(x-1))^T$$

Die Bedingung $\operatorname{grad} f(x,y) = 0$ liefert die beiden stationären Punkte

$$\mathrm{x}^0 = (0,0)^T \qquad \text{und} \qquad \mathrm{x}^1 = (-2/3, 0)^T.$$

Die jeweiligen Hesse–Matrizen von f an den Stellen x^0 und x^1 lauten

$$\mathrm{H}f(\mathrm{x}^0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \qquad \text{und} \qquad \mathrm{H}f(\mathrm{x}^1) = \begin{pmatrix} -2 & 0 \\ 0 & -10/3 \end{pmatrix}$$

Die Matrix $\mathrm{H}f(\mathrm{x}^0)$ ist indefinit, also ist x^0 ein Sattelpunkt, $\mathrm{H}f(\mathrm{x}^1)$ ist negativ definit, somit ist x^1 ein strenges lokales Maximum von $f(\mathrm{x})$.

2.2 Implizit definierte Funktionen

Ziel: Untersuche die Lösungsmengen von *nichtlinearen* Gleichungssystemen der Form

$$\mathrm{g}(\mathrm{x}) = 0$$

mit $\mathrm{g} : D \to \mathbb{R}^m$, $D \subset \mathbb{R}^n$, d.h. wir betrachten m Gleichungen für n Unbekannte mit

$$m < n.$$

Also: Es gibt weniger Gleichungen als Unbekannte.

Man nennt dann das Gleichungssystem unterbestimmt und die Lösungsmenge $G \subset \mathbb{R}^n$ enthält gewöhnlich *unendlich* viele Punkte.

Auflösbarkeit von (nichtlinearen) Gleichungen.

Frage: Kann man das System $g(x) = 0$ nach bestimmten Unbekannten, zum Beispiel den letzten m Variablen $x_{n-m+1}, \dots, x_n$ **auflösen**?

Mit anderen Worten: Existiert eine Funktion $f(x_1, \dots, x_{n-m})$ mit

$$g(x) = 0 \quad \Longleftrightarrow \quad (x_{n-m+1}, \dots, x_n)^T = f(x_1, ..., x_{n-m})$$

Terminologie: "Auflösen" bedeutet also die letzten m Variablen durch die ersten $n - m$ Variablen zu beschreiben.

Weitere Frage: Nach welchen m Variablen lässt sich das Gleichungssystem auflösen? Ist die Auflösung *global* auf dem Definitionsbereich D möglich oder nur *lokal* auf einer Teilmenge $\tilde{D} \subset D$?

Geometrische Interpretation: Die Lösungsmenge G von $g(x) = 0$ lässt sich (zumindest lokal) als Graph einer Funktion $f : \mathbb{R}^{n-m} \to \mathbb{R}^m$ darstellen.

Beispiel.

Die Kreisgleichung

$$g(x, y) = x^2 + y^2 - r^2 = 0 \qquad \text{mit } r > 0$$

definiert ein unterbestimmtes nichtlineares Gleichungssystem, denn wir haben **zwei** Unbekannte (x, y), aber nur **eine** Gleichung.

Die Kreisgleichung lässt sich lokal auflösen und definiert dabei die folgenden vier Funktionen:

$$\begin{aligned}
y &= \sqrt{r^2 - x^2}, \quad -r \leq x \leq r \\
y &= -\sqrt{r^2 - x^2}, \quad -r \leq x \leq r \\
x &= \sqrt{r^2 - y^2}, \quad -r \leq y \leq r \\
x &= -\sqrt{r^2 - y^2}, \quad -r \leq y \leq r
\end{aligned}$$

Beispiel.

Sei g(x) eine affin–lineare Funktion, d.h. g hat die Form

$$\mathrm{g}(\mathrm{x}) = \mathrm{Cx} + \mathrm{b} \quad \text{für } \mathrm{C} \in \mathbb{R}^{m\times n},\ \mathrm{b} \in \mathbb{R}^m$$

Wir spalten die Variablen x in zwei Vektoren auf

$$\mathrm{x}^{(1)} = (x_1, \ldots, x_{n-m})^T \in \mathbb{R}^{n-m} \quad \text{und} \quad \mathrm{x}^{(2)} = (x_{n-m+1}, \ldots, x_n)^T \in \mathbb{R}^n$$

Aufspaltung der Matrix $\mathrm{C} = [\mathrm{B}, \mathrm{A}]$ ergibt die Darstellung

$$\mathrm{g}(\mathrm{x}) = \mathrm{Bx}^{(1)} + \mathrm{Ax}^{(2)} + \mathrm{b}$$

mit $\mathrm{B} \in \mathbb{R}^{m\times(n-m)}$, $\mathrm{A} \in \mathbb{R}^{m\times m}$.

Das Gleichungssystem $\mathrm{g}(\mathrm{x}) = 0$ ist genau dann nach den Variablen $\mathrm{x}^{(2)}$ (eindeutig) auflösbar, falls A regulär ist:

$$\mathrm{g}(\mathrm{x}) = 0 \quad \Longleftrightarrow \quad \mathrm{x}^{(2)} = -\mathrm{A}^{-1}(\mathrm{Bx}^{(1)} + \mathrm{b}) = \mathrm{f}(\mathrm{x}^{(1)})$$

Fortsetzung des Beispiels.

Frage: Wie kann man die Matrix A in Abhängigkeit von g schreiben?

Aus der Darstellung

$$g(x) = Bx^{(1)} + Ax^{(2)} + b$$

erkennt man direkt, dass

$$A = \frac{\partial g}{\partial x^{(2)}}(x^{(1)}, x^{(2)})$$

gilt, d.h. A ist die Jacobi–Matrix der Abbildung

$$x^{(2)} \to g(x^{(1)}, x^{(2)})$$

für festes $x^{(1)}$!

Fazit: Auflösbarkeit ist somit gegeben, falls die Jacobi–Matrix regulär ist.

Satz über implizite Funktionen.

Satz: Sei $g : D \to \mathbb{R}^m$ eine $\mathcal{C}^1$–Funktion, $D \subset \mathbb{R}^n$ offen. Die Variablen in D seien (x, y) mit $\mathrm{x} \in \mathbb{R}^{n-m}$ und $\mathrm{y} \in \mathbb{R}^m$. Der Punkt $(\mathrm{x}^0, \mathrm{y}^0) \in D$ sei eine Lösung von $\mathrm{g}(\mathrm{x}^0, \mathrm{y}^0) = 0$.

Falls die Jacobi–Matrix

$$\frac{\partial \mathrm{g}}{\partial \mathrm{y}}(\mathrm{x}^0, \mathrm{y}^0) := \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(\mathrm{x}^0, \mathrm{y}^0) & \dots & \frac{\partial g_1}{\partial y_m}(\mathrm{x}^0, \mathrm{y}^0) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial y_1}(\mathrm{x}^0, \mathrm{y}^0) & \dots & \frac{\partial g_m}{\partial y_m}(\mathrm{x}^0, \mathrm{y}^0) \end{pmatrix}$$

regulär ist, so gibt es Umgebungen U von x^0 und V von y^0, $U \times V \subset D$ und eine eindeutig bestimmte stetig differenzierbare Funktion $\mathrm{f} : U \to V$ mit

$$\mathrm{f}(\mathrm{x}^0) = \mathrm{y}^0 \quad \text{und} \quad \mathrm{g}(\mathrm{x}, \mathrm{f}(\mathrm{x})) = 0 \quad \text{für alle } \mathrm{x} \in U$$

und

$$J\,\mathrm{f}(\mathrm{x}) = -\left(\frac{\partial \mathrm{g}}{\partial \mathrm{y}}(\mathrm{x}, \mathrm{f}(\mathrm{x}))\right)^{-1} \left(\frac{\partial \mathrm{g}}{\partial \mathrm{x}}(\mathrm{x}, \mathrm{f}(\mathrm{x}))\right)$$

Beispiel.

Für die Kreisgleichung $g(x, y) = x^2 + y^2 - r^2 = 0, r > 0$ findet man im Punkt $(x^0, y^0) = (0, r)$

$$\frac{\partial g}{\partial x}(0, r) = 0, \quad \frac{\partial g}{\partial y}(0, r) = 2r \neq 0$$

Man kann also in einer Umgebung von $(0, r)$ die Kreisgleichung nach y auflösen:

$$f(x) = \sqrt{r^2 - x^2}$$

Die Ableitung $f'(x)$ kann man durch implizite Diffentiation berechnen:

$$g(x, y(x)) = 0 \quad \Longrightarrow \quad g_x(x, y(x)) + g_y(x, y(x))y'(x) = 0$$

Also

$$2x + 2y(x)y'(x) = 0 \quad \Rightarrow \quad y'(x) = f'(x) = -\frac{x}{y(x)}$$

Ein weiteres Beispiel.

Betrachte die Gleichung $g(x,y) = e^{y-x} + 3y + x^2 - 1 = 0$.

Es gilt

$$\frac{\partial g}{\partial y}(x,y) = e^{y-x} + 3 > 0 \qquad \text{für alle } x \in \mathbb{R}.$$

Die Gleichung ist also für jedes $x \in \mathbb{R}$ nach $y =: f(x)$ auflösbar und $f(x)$ ist eine stetig differenzierbare Funktion. Implizite Differentiation liefert

$$e^{y-x}(y'-1) + 3y' + 2x = 0 \quad \Longrightarrow \quad y' = \frac{e^{y-x} - 2x}{e^{y-x} + 3}$$

Erneute Differentiation liefert

$$e^{y-x}y'' + e^{y-x}(y'-1)^2 + 3y'' + 2 = 0 \quad \Longrightarrow \quad y' = -\frac{2 + e^{y-x}(y'-1)^2}{e^{y-x} + 3}$$

Aber: Explizites Auflösen nach y (mit Hilfe elementarer Funktionen) ist in diesem Fall nicht möglich!

Allgemeine Bemerkung.

Implizites Differenzieren einer durch

$$g(x,y) = 0, \quad \frac{\partial g}{\partial y} \neq 0$$

implizit definierten Funktion $y = f(x)$, mit $x, y \in \mathbb{R}$, ergibt

$$\begin{aligned} f'(x) &= -\frac{g_x}{g_y} \\ f''(x) &= -\frac{g_{xx}g_y^2 - 2g_{xy}g_xg_y + g_{yy}g_x^2}{g_y^3} \end{aligned}$$

Daher ist der Punkt x^0 ein stationärer Punkt von $f(x)$, falls gilt

$$g(x^0, y^0) = g_x(x^0, y^0) = 0 \quad \text{und} \quad g_y(x^0, y^0) \neq 0$$

Weiter ist x^0 ein lokales Maximum (bzw. Minimum), falls

$$\frac{g_{xx}(x^0, y^0)}{g_y(x^0, y^0)} > 0 \qquad \left(\text{bzw. } \frac{g_{xx}(x^0, y^0)}{g_y(x^0, y^0)} < 0 \right)$$

Implizite Darstellung ebener Kurven.

Betrachte die Lösungsmenge einer skalaren Gleichungen

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

Falls gilt

$$\operatorname{grad} g = (g_x, g_y) \neq 0$$

so definiert $g(x, y)$ lokal eine Funktion $y = f(x)$ oder $x = \bar{f}(y)$.

Definition: Ein Lösungspunkt (x^0, y^0) der Gleichung $g(x, y) = 0$ mit

- $\operatorname{grad} g(x^0, y^0) \neq 0$ heißt regulärer Punkt,
- $\operatorname{grad} g(x^0, y^0) = 0$ heißt singulärer Punkt.

Beispiel: Betrachte wieder die Kreisgleichung

$$g(x, y) = x^2 + y^2 - r = 0 \quad \text{mit } r > 0.$$

Auf der Kreislinie liegen keine singulären Punkte!

Horizontale und vertikale Tangenten.

Bemerkung:

a) Gilt für einen regulären Punkt (x^0, y^0)

$$g_x(\mathrm{x}^0) = 0 \quad \text{und} \quad g_y(\mathrm{x}^0) \neq 0$$

so besitzt die Lösungskurve eine horizontale Tangente in x^0.

b) Gilt für einen regulären Punkt (x^0, y^0)

$$g_x(\mathrm{x}^0) \neq 0 \quad \text{und} \quad g_y(\mathrm{x}^0) = 0$$

so besitzt die Lösungskurve eine vertikale Tangente in x^0.

c) Ist x^0 ein singulärer Punkt, so wird die Lösungsmenge bei x^0 "in zweiter Näherung" durch folgende quadratische Gleichung approximiert.

$$g_{xx}(\mathrm{x}^0)(x - x^0)^2 + 2g_{xy}(\mathrm{x}^0)(x - x^0)(y - y^0) + g_{yy}(\mathrm{x}^0)(y - y^0)^2 = 0$$

Bemerkungen.

Wegen c) erhält man für $g_{xx}, g_{xy}, g_{yy} \neq 0$:

$$\det Hg(x^0) > 0 \quad : \quad x^0 \text{ ist ein isolierter Punkt der Lösungsmenge}$$

$$\det Hg(x^0) < 0 \quad : \quad x^0 \text{ ist ein Doppelpunkt}$$

$$\det Hg(x^0) = 0 \quad : \quad x^0 \text{ ist ein Rückkehrpunkt bzw. eine Spitze}$$

Geometrische Interpretation:

a) Gilt $\det Hg(x^0) > 0$, so sind beide Eigenwerte von $Hg(x^0)$ entweder strikt positiv oder strikt negativ, d.h. x^0 ist ein strenges lokales Minimum oder Maximum von $g(x)$.

b) Gilt $\det Hg(x^0) < 0$, so haben die beiden Eigenwerte von $Hg(x^0)$ ein unterschiedliches Vorzeichen, d.h. x^0 ist ein Sattelpunkt von $g(x)$.

c) Gilt $\det Hg(x^0) = 0$, so ist der stationäre Punkt x^0 von $g(x)$ ausgeartet.

Beispiel 1.

Betrachte den singulären Punkt $x^0 = 0$ der impliziten Gleichung

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

Berechnung der partiellen Ableitungen bis zur Ordnung 2:

$$\begin{aligned}
g_x &= y^2 + 3x^2 - 4x \\
g_y &= 2y(x-1) \\
g_{xx} &= 6x - 4 \\
g_{xy} &= 2y \\
g_{yy} &= 2(x-1) \\
\mathrm{H}g(0) &= \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix}
\end{aligned}$$

Also ist $x^0 = 0$ ein isolierter Punkt.

Beispiel 2.

Betrachte den singulären Punkt $\mathrm{x}^0 = 0$ der impliziten Gleichung

$$g(x,y) = y^2(x-1) + x^2(x+q^2) = 0$$

Berechnung der partiellen Ableitungen bis zur Ordnung 2:

$$\begin{aligned}
g_x &= y^2 + 3x^2 + 2xq^2 \\
g_y &= 2y(x-1) \\
g_{xx} &= 6x + 2q^2 \\
g_{xy} &= 2y \\
g_{yy} &= 2(x-1) \\
\mathrm{H}g(0) &= \begin{pmatrix} 2q^2 & 0 \\ 0 & -2 \end{pmatrix}
\end{aligned}$$

Also ist $\mathrm{x}^0 = 0$ für $q \neq 0$ ein Doppelpunkt.

Beispiel 3.

Betrachte den singulären Punkt $\mathrm{x}^0 = 0$ der impliziten Gleichung

$$g(x, y) = y^2(x-1) + x^3 = 0$$

Berechnung der partiellen Ableitungen bis zur Ordnung 2:

$$\begin{aligned} g_x &= y^2 + 3x^2 \\ g_y &= 2y(x-1) \\ g_{xx} &= 6x \\ g_{xy} &= 2y \\ g_{yy} &= 2(x-1) \\ \mathrm{H}g(0) &= \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \end{aligned}$$

Also ist $\mathrm{x}^0 = 0$ eine Spitze (bzw. ein Rückkehrpunkt).

Implizite Darstellung von Flächen.

- Die Lösungsmenge einer skalaren Gleichung $g(x,y,z)=0$ ist für $\operatorname{grad} g \neq 0$ *lokal* eine Fläche im $\mathbb{R}^3$.
- Für die Tangentialebene in $\mathrm{x}^0=(x^0,y^0,z^0)^T$ mit $g(\mathrm{x}^0)=0$ und $\operatorname{grad} g(\mathrm{x}^0) \neq 0^T$ bekommen wir für $\Delta\mathrm{x}^0=\mathrm{x}-\mathrm{x}^0$ mit Taylor–Entwicklung

$$\operatorname{grad} g \cdot \Delta\mathrm{x}^0 = g_x(\mathrm{x}^0)(x-x^0)+g_y(\mathrm{x}^0)(y-y^0)+g_z(\mathrm{x}^0)(z-z_0)=0$$

d.h. der Gradient steht senkrecht auf der Fläche $g(x,y,z)=0$.

- Ist zum Beispiel $g_z(\mathrm{x}^0)\neq 0$, so gibt es lokal bei x^0 eine Darstellung der Form

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

und für die partielle Ableitungen von $f(x,y)$ bekommt man

$$\operatorname{grad} f(x,y)=(f_x,f_y)=-\frac{1}{g_z}(g_x,g_y)=\left(-\frac{g_x}{g_z},\frac{g_y}{g_z}\right)$$

mit dem Satz über implizite Funktionen.

Das Umkehrproblem.

Frage: Lässt sich ein vorgegebenes Gleichungssystem

$$\mathrm{y} = \mathrm{f}(\mathrm{x})$$

mit $\mathrm{f} : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$ offen, nach x auflösen, also **invertieren**?

Satz: (Umkehrsatz)

Sei $\mathrm{f} : D \to \mathbb{R}^n$, $D \subset \mathbb{R}^n$ offen, eine $\mathcal{C}^1$–Funktion. Ist für ein $\mathrm{x}^0 \in D$ die Jacobi–Matrix $\mathrm{J}\,\mathrm{f}(\mathrm{x}^0)$ regulär, so gibt es Umgebungen U und V von x^0 und $\mathrm{y}^0 = \mathrm{f}(\mathrm{x}^0)$, so dass f den Bereich U **bijektiv** auf V abbildet.

Die Umkehrfunktion $\mathrm{f}^{-1} : V \to U$ ist ebenfalls eine $\mathcal{C}^1$–Funktion und es gilt für alle $\mathrm{x} \in U$:

$$\mathrm{J}\,\mathrm{f}^{-1}(\mathrm{y}) = (\mathrm{J}\,\mathrm{f}(\mathrm{x}))^{-1}, \quad \mathrm{y} = \mathrm{f}(\mathrm{x})$$

Bemerkung: Man nennt dann f lokal einen $\mathcal{C}^1$–Diffeomorphismus.

2.3 Extremalprobleme unter Nebenbedingungen

Frage: Welche Abmessungen sollte eine Metalldose haben, damit bei vorgegebenem Volumen der Materialverbrauch am geringsten ist?

Lösungsansatz: Sei $r > 0$ der Radius und $h > 0$ die Höhe der Dose. Dann gilt

$$\begin{aligned} V &= \pi r^2 h \\ O &= 2\pi r^2 + 2\pi r h \end{aligned}$$

Setze bei vorgebenem Volumen $c \in \mathbb{R}_+$, und mit $x := r, y := h$,

$$\begin{aligned} f(x,y) &= 2\pi x^2 + 2\pi x y \\ g(x,y) &= \pi x^2 y - c = 0 \end{aligned}$$

Bestimme das Minimum der Funktion $f(x,y)$ auf der Menge

$$G := \{(x,y) \in \mathbb{R}_+^2 \,|\, g(x,y) = 0\}$$

Lösung des restringierten Minimierungsproblems.

Aus $g(x, y) = \pi x^2 y - c = 0$ folgt

$$y = \frac{c}{\pi x^2}$$

Einsetzen in $f(x, y)$ ergibt

$$h(x) := 2\pi x^2 + 2\pi x \frac{c}{\pi x^2} = 2\pi x^2 + \frac{2c}{x}$$

Bestimme das Minimum der Funktion $h(x)$:

$$h'(x) = 4\pi x - \frac{2c}{x^2} = 0 \quad \Rightarrow \quad 4\pi x = \frac{2c}{x^2} \quad \Rightarrow \quad x = \left(\frac{c}{2\pi}\right)^{1/3}$$

Hinreichende Bedingung

$$h''(x) = 4\pi + \frac{4c}{x^3} \quad \Rightarrow \quad h''\left(\left(\frac{c}{\pi}\right)^{1/3}\right) = 12\pi > 0$$

Allgemeine Formulierung des Problems.

Bestimme die Extremwerte der Funktion $f : \mathbb{R}^n \to \mathbb{R}$ unter den Nebenbedingungen

$$\mathrm{g}(\mathrm{x}) = 0$$

wobei $\mathrm{g} : \mathbb{R}^n \to \mathbb{R}^m$.

Die Nebenbedingungen lauten also

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \ldots, x_n) &= 0 \end{aligned}$$

Alternativ: Bestimme die Extremwerte der Funktion $f(\mathrm{x})$ auf der Menge

$$G := \{\mathrm{x} \in \mathbb{R}^n \,|\, \mathrm{g}(\mathrm{x}) = 0\}$$

Die Lagrange–Funktion und das Lagrange–Lemma.

Wir definieren die Lagrange–Funktion

$$F(x) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

und suchen die Extremwerte von $F(x)$ für festes $\lambda = (\lambda_1, \ldots, \lambda_m)^T$.

Die Zahlen λ_i, $i = 1, \ldots, m$ nennt man Lagrange–Multiplikatoren.

Satz: (Lagrange–Lemma) Minimiert (bzw. maximiert) x^0 die Lagrange–Funktion $F(x)$ (für ein festes λ) über D und gilt $g(x^0) = 0$, so liefert x^0 das Minimum (bzw. Maximum) von $f(x)$ über $G := \{x \in D \,|\, g(x) = 0\}$.

Beweis: Für ein beliebiges $x \in D$ gilt nach Vorausetzung

$$f(x^0) + \lambda^T g(x^0) \leq f(x) + \lambda^T g(x)$$

Wählt man speziell $x \in G$, so ist $g(x) = g(x^0) = 0$, also auch $f(x^0) \leq f(x)$.

Eine notwendige Bedingung für lokale Extrema.

Sind f und g_i, $i = 1, \ldots, m$, $\mathcal{C}^1$–Funktionen, so ist eine notwendige Bedingung für eine Extremstelle x^0 von $F(\mathrm{x})$ gegeben durch

$$\operatorname{grad} F(\mathrm{x}) = \operatorname{grad} f(\mathrm{x}) + \sum_{i=1}^{m} \lambda_i \operatorname{grad} g_i(\mathrm{x}) = 0$$

Zusammen mit den Nebenbedinungen $\mathrm{g}(\mathrm{x}) = 0$ ergibt sich ein (nichtlineares) Gleichungssystem mit $(n+m)$ Gleichungen und $(n+m)$ Unbekannten x und λ.

Die Lösungen $(\mathrm{x}^0, \lambda^0)$ sind die Kandidaten für die gesuchten Extremstellen, denn diese erfüllen die o.g. notwendige Bedingung.

Alternativ: Definiere eine Langrange–Funktion

$$G(\mathrm{x}, \lambda) := f(\mathrm{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathrm{x})$$

und suche die Extremstellen von $G(\mathrm{x}, \lambda)$ bezüglich x **und** λ.

Einige Bemerkungen zu hinreichenden Bedingungen.

1. Man kann auch eine **hinreichende** Bedingung aufstellen:
   Sind die Funktionen f und g $\mathcal{C}^2$–Funktionen und ist die Hesse–Matrix $\mathrm{H}F(\mathrm{x}^0)$ der Lagrange–Funktion positiv (bzw. negativ) definit, so ist x^0 tatsächlich ein strenges lokales Minimum (bzw. Maximum) von $f(\mathrm{x})$ auf G.

2. In den meisten Anwendungen ist die hinreichende Bedingung allerdings **nicht** erfüllt, obwohl x^0 ein strenges lokales Extremum ist.

3. Insbesondere kann man aus der Indefinitheit der Hesse–Matrix $\mathrm{H}F(\mathrm{x}^0)$ **nicht** schließen, dass x^0 kein Extremwert ist.

4. Ähnlich problematisch ist die hinreichende Bedingung, die man aus der Hesse–Matrix für die Lagrange–Funktion $G(\mathrm{x}, \lambda)$ bezüglich x **und** λ erhält.

Ein Beispiel zu restringierten Minimierungsproblems.

Gesucht seien die Extrema von $f(x,y) := xy$ auf der Kreisscheibe

$$K := \{(x,y)^T \,|\, x^2 + y^2 \leq 1\}$$

Da die betrachte Funktion f stetig und $K \subset \mathbb{R}^2$ kompakt ist, folgt aus der Min–Max–Eigenschaft die Existenz von globalen Maxima und Minima auf K.

Wir betrachten zunächst das Innere K^0 von K, also die offene Menge

$$K^0 := \{(x,y)^T \,|\, x^2 + y^2 < 1\}$$

Die notwendige Bedingung für einen Extremwert lautet nun

$$\operatorname{grad} f = (y,x) = 0$$

Somit ist der Ursprung $\mathrm{x}^0 = 0$ Kandidat für ein (lokales) Extremum.

Fortsetzung des Beispiels.

Die Hesse–Matrix im Ursprung, gegeben durch

$$\mathrm{H}f(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

und ist indefinit. Daher ist x^0 ein Sattelpunkt.

Die Extrema der Funktion müssen also auf dem Rand liegen, der eine Gleichungsnebenbedingung darstellt:

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

Wir suchen also die Extremwerte von $f(x,y) = xy$ unter der Nebenbedingung $g(x,y) = 0$.

Die Lagrange–Funktion lautet

$$F(x,y) = xy + \lambda(x^2 + y^2 - 1)$$

Komplettierung des Beispiels.

Damit ergibt sich das (nichtlineare) Gleichungssystem

$$\begin{aligned} y + 2\lambda x &= 0 \\ x + 2\lambda y &= 0 \\ x^2 + y^2 &= 1 \end{aligned}$$

mit den vier Lösungen

$$\begin{aligned} \lambda = \frac{1}{2} \quad &: \quad \mathrm{x}^{(1)} = (\sqrt{1/2}, -\sqrt{1/2})^T \quad \mathrm{x}^{(2)} = (-\sqrt{1/2}, \sqrt{1/2})^T \\ \lambda = -\frac{1}{2} \quad &: \quad \mathrm{x}^{(3)} = (\sqrt{1/2}, \sqrt{1/2})^T \quad \mathrm{x}^{(4)} = (-\sqrt{1/2}, -\sqrt{1/2})^T \end{aligned}$$

Minima und Maxima lassen sich nun einfach aus den Funktionswerten ablesen

$$f(\mathrm{x}^{(1)}) = f(\mathrm{x}^{(2)}) = -1/2 \qquad f(\mathrm{x}^{(3)}) = f(\mathrm{x}^{(4)}) = 1/2$$

d.h. Minima sind $\mathrm{x}^{(1)}$ und $\mathrm{x}^{(2)}$, Maxima sind $\mathrm{x}^{(3)}$ und $\mathrm{x}^{(4)}$.

Lagrange–Multiplikatoren–Regel.

Satz: Seien $f, g_1, \ldots, g_m : D \to \mathbb{R}$ jeweils $\mathcal{C}^1$–Funktionen, und sei $\mathrm{x}^0 \in D$ ein lokales Extremum von $f(\mathrm{x})$ unter der Nebenbedingung $\mathrm{g}(\mathrm{x}) = 0$. Weiterhin gelte die Regularitätsbedingung

$$\mathrm{rang}\left(\mathrm{J}\,\mathrm{g}(\mathrm{x}^0)\right) = m$$

Dann existieren Lagrange–Multiplikatoren $\lambda_1, \ldots, \lambda_m$, so dass für die Lagrange Funktion

$$F(\mathrm{x}) := f(\mathrm{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathrm{x})$$

die folgende **notwendige Bedingung erster Ordnung** gilt:

$$\mathrm{grad}\, F(\mathrm{x}^0) = 0$$

Notwendige Bedingung zweiter Ordnung und hinreichende Bedingung.

Satz: 1) Ist $x^0 \in D$ ein lokales Minimum von $f(x)$ unter der Nebenbedingung $g(x) = 0$, ist die Regularitätsbedingung erfüllt und sind $\lambda_1, \ldots, \lambda_m$ zugehörige Lagrange–Multiplikatoren, so ist die Hesse–Matrix $HF(x^0)$ der Lagrange–Funktion positiv semidefinit auf dem Tangentialraum

$$TG(x^0) := \{y \in \mathbb{R}^n \,|\, \operatorname{grad} g_i(x^0) \cdot y = 0 \text{ für } i = 1, \ldots, m\}$$

d.h., es gilt $y^T\, HF(x^0)\, y \geq 0$ für alle $y \in TG(x^0)$.

2) Ist für einen Punkt $x^0 \in G$ die Regularitätsbedingung erfüllt, existieren Lagrange–Multiplikatoren $\lambda_1, \ldots, \lambda_m$, so dass x^0 ein stationärer Punkt der zugehörigen Lagrange–Funktion ist, und ist die Hesse–Matrix $HF(x^0)$ positiv definit auf dem Tangentialraum $TG(x^0)$, d.h., gilt

$$y^T\, HF(x^0)\, y > 0 \quad \forall\, y \in TG(x^0) \setminus \{0\},$$

so ist x^0 ein strenges lokales Minimum von $f(x)$ unter der Nebenbedingung $g(x) = 0$.

Beispiel.

Man bestimme das globale Maximum der Funktion

$$f(x, y) = -x^2 + 8x - y^2 + 9$$

unter der Nebenbedingung

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

Die Lagrange–Funktion ist

$$F(x) = -x^2 + 8x - y^2 + 9 + \lambda(x^2 + y^2 - 1)$$

Aus der notwendigen Bedingung ergibt sich das nichtlineare System

$$\begin{aligned} -2x + 8 &= -2\lambda x \\ -2y &= -2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

Fortsetzung des Beispiels.

Aus der notwendigen Bedingung ergibt sich das nichtlineare System

$$\begin{aligned} -2x + 8 &= -2\lambda x \\ -2y &= -2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

Aus der ersten Gleichung folgt $\lambda \neq 1$. Verwendet man dies in der zweiten Gleichung, so gilt $y = 0$. Aus der dritten Gleichung erkennt man sofort $x = \pm 1$.

Demnach sind die beiden Punkte $(x, y) = (1, 0)$ und $(x, y) = (-1, 0)$ Kandidaten für das globale Maximum. Wegen

$$f(1, 0) = 16 \qquad f(-1, 0) = 0$$

wird das globale Maximum von $f(x, y)$ unter der Nebenbedingung $g(x, y) = 0$ im Punkt $(x, y) = (1, 0)$ angenommen.

Ein weiteres Beispiel.

Man bestimme die lokalen Extremwerte der Funktion

$$f(x, y, z) = 2x + 3y + 2z$$

auf dem Durchschnitt des Zylindersmantels

$$M_Z := \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x^2 + y^2 = 2\}$$

mit der Ebene

$$E := \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x + z = 1\}$$

Umformulierung: Bestimme die Extremwerte der Funktion $f(x, y, z)$ unter den Nebenbedingungen

$$\begin{aligned} g_1(x, y, z) &:= x^2 + y^2 - 2 = 0 \\ g_2(x, y, z) &:= x + z - 1 = 0 \end{aligned}$$

Fortsetzung des Beispiel.

Die Jacobi–Matrix

$$\mathrm{Jg(x)} = \begin{pmatrix} 2x & 2y & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

hat den Rang 2, d.h. wir können über die Lagrange–Funktion Extremwerte bestimmen:

$$F(x,y,z) = 2x + 3y + 2z + \lambda_1(x^2 + y^2 - 2) + \lambda_2(x + z - 1)$$

Die notwendige Bedingung ergibt das nichtlineare Gleichungssystem

$$\begin{aligned} 2 + 2\lambda_1 x + \lambda_2 &= 0 \\ 3 + 2\lambda_1 y &= 0 \\ 2 + \lambda_2 &= 0 \\ x^2 + y^2 &= 2 \\ x + z &= 1 \end{aligned}$$

Fortsetzung des Beispiel.

Die notwendige Bedingung ergibt das nichtlineare Gleichungssystem

$$\begin{aligned} 2 + 2\lambda_1 x + \lambda_2 &= 0 \\ 3 + 2\lambda_1 y &= 0 \\ 2 + \lambda_2 &= 0 \\ x^2 + y^2 &= 2 \\ x + z &= 1 \end{aligned}$$

Aus der ersten und dritten Gleichung folgt

$$2\lambda_1 x = 0$$

Aus der zweiten Gleichung folgt $\lambda_1 \neq 0$, also $x = 0$.
Damit ergeben sich die möglichen Extremwerte als

$$(x, y, z) = (0, \sqrt{2}, 1) \quad (x, y, z) = (0, -\sqrt{2}, 1)$$

Komplettierung des Beispiel.

Die möglichen Extremwerte sind also

$$(x, y, z) = (0, \sqrt{2}, 1) \quad (x, y, z) = (0, -\sqrt{2}, 1)$$

und liegen offensichtlich auf der Mantelfläche M_Z des Zylinders Z mit

$$\begin{aligned} Z &= \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x^2 + y^2 \leq 2\} \\ M_Z &= \{(x, y, z)^T \in \mathbb{R}^3 \,|\, x^2 + y^2 = 2\} \end{aligned}$$

Man berechnet nun die zugehörigen Funktionswerte

$$\begin{aligned} f(0, \sqrt{2}, 1) &= 3\sqrt{2} + 2 \\ f(0, -\sqrt{2}, 1) &= -3\sqrt{2} + 2 \end{aligned}$$

Daher liegt im Punkt $(x, y, z) = (0, \sqrt{2}, 1)$ ein Maximum und im Punkt $(x, y, z) = (0, -\sqrt{2}, 1)$ ein Minimum vor.