MD Differentialrechnung im $\mathbb{R}^n$

Die Ableitung

Wir betrachten Funktionen

$$f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n, \quad x \mapsto f(x),$$

deren Definitionsbereich Ω grundsätzlich ein Gebiet, also eine offene und zusammenhängende Menge des $\mathbb{R}^m$ sei. Ist der Bildbereich $\mathbb{R}$, so reden wir speziell auch von "Skalarfunktionen", ist er der $\mathbb{R}^n$ mit $n > 1$, so sprechen wir auch von "vektorwertigen" Funktionen.
Wir schreiben

$$x = (x_1, ..., x_m)^T, \quad f(x) = (f_1(x), ..., f_n(x))^T$$

mit den Komponentenfunktionen $f_i(x) := f_i(x_1, ..., x_m)$.
Häufig ist es unerheblich, ob man x oder $f(x)$ als Zeile oder als Spalte denkt, und dann gehen wir auch lax mit der Notation um.
Im Falle $m = 2, 3$ schreiben wir häufig auch $f(x, y)$ bzw. $f(x, y, z)$ statt $f(x_1, x_2)$ bzw. $f(x_1, x_2, x_3)$.

Definition MD.1 *$f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n$ ist an der Stelle $x_0 \in \Omega$ differenzierbar, wenn es eine Matrix $A \in \mathbb{R}^{n\times m}$ gibt, sodaß*

$$f(x) = f(x_0) + A(x - x_0) + R(x, x_0),$$

wobei für den Rest R gilt

$$\lim_{x\to x_0} \frac{|R(x, x_0)|}{|x - x_0|} = 0.$$

Man nennt dann A die "Ableitung von f an der Stelle x_0" und notiert dafür auch

$$df(x_0), \ f'(x_0) \text{ oder } Df(x_0).$$

f ist in Ω differenzierbar, wenn es an jeder Stelle $x_0 \in \Omega$ differenzierbar ist.

Im allgemeinen Fall ist also die Ableitung an einer Stelle x_0 eine Matrix, die sog. Jacobi-Matrix. Wie man sie berechnet, werden wir etwas später sehen.
Im Falle einer Skalarfunktion in einer Variablen, d.h. $m = n = 1$ ist die Ableitung eine 1×1 Matrix, d.h. eine Zahl und unsere Definition geht in die Formulierung der Ableitung nach Satz F.50 über.
Im Falle einer Skalarfunktion $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}$ ist die Ableitung eine $1\times m$ - Matrix, d.h. eine Zeile. Man nennt sie den "Gradienten" von f und notiert dafür ∇f. Das Zeichen ∇ heißt der "Nabla"-Operator.

Die formale Übereinstimmung unserer allgemeinen Definition der Ableitung mit der Formulierung von Satz F.50 erlaubt es, auch eine Reihe von Schlüssen direkt auf den allgemeinen Fall zu übertragen.

Satz MD.2 *Ist $f : \Omega \to \mathbb{R}^n$ differenzierbar in x_0, so auch stetig in x_0.*

Satz MD.3 *Sind $f, g : \Omega \to \mathbb{R}^n$ differenzierbar in x_0 mit Ableitungen $df(x_0) = F, dg(x_0) = G$, ferner $\alpha, \beta \in \mathbb{R}$, so ist auch $h := \alpha f + \beta g$ differenzierbar in x_0 mit Ableitung $dh(x_0) = \alpha F + \beta G$.*

Ferner haben wir

Satz MD.4 (Kettenregel) *Sind $g : \Omega \to \Omega' \subset \mathbb{R}^n$, $f : \Omega' \to \mathbb{R}^k$ differenzierbar, so auch $h := f \circ g : \Omega \to \mathbb{R}^k$ und für $x_0 \in \Omega$, $y_0 := g(x_0) \in \Omega'$ gilt mit $G := dg(x_0)$, $F := df(y_0)$ dann*

$$dh(x_0) = (df(y_0))(dg(x_0)) = FG.$$

Die Beweise von Satz MD.2 und Satz MD.3 folgen trivial aus der Definition, die Kettenregel beweist man mit genau denselben Schlüssen wir in einer Dimension (Siehe Satz F.60.)

Einige

Beispiele MD.5 *Differenzierbar sind*

1. *konstante Funktionen: $f : x \mapsto f(x) = c$. Es ist $df(x) = 0$.*
2. *lineare Funktionen: Mit einer $(n \times m)$ - Matrix A sei $f(x) := Ax$. Dann ist f differenzierbar und $df(x_0) = A$.*
   (Es ist $f(x) = Ax = Ax_0 + A(x - x_0) = f(x_0) + A(x - x_0) + 0$ unabhängig von der Stelle x_0.)
   Spezialfälle sind etwa
3. *$A = e_i^T$, d.h. $p_i(x) := e_i^T x = x_i$ ist die Projektion auf die i - te Komponente mit Ableitung*
   $$dp_i(x) = e_i^T \quad \text{(Zeile)}.$$
4. *Mit einem Punkt $x_0 \in \mathbb{R}^n$ sei $k_j : \mathbb{R} \to \mathbb{R}^m : t \mapsto x_0 + te_j$ mit der Ableitung*
   $$dk_j(t) = e_j \quad \text{(Spalte)}.$$
   (Es ist $k_j(t) = x_0 + t_0 e_j + (t - t_0)e_j = k_j(t_0) + e_j(t - t_0) + 0$.)
   Die Funktion k_j bildet $\mathbb{R}$ auf die Parallele zur j-ten Koordinatenachse durch x_0 ab, wobei 0 auf x_0 geht.

Ist nun $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n, x \mapsto (f_1(x), ..., f_n(x))^T$ differenzierbar mit Ableitung Df, so ist $p_i \circ f : \Omega \to \mathbb{R}$ die Funktion $x \mapsto f_i(x)$, also einfach die i-te Komponentenfunktion. Ferner ist für $x_0 \in \Omega$ und kleine $t \in \mathbb{R}$ dann

$$\varphi_{ij} := p_i \circ f \circ k_j = f_i \circ k_j : \mathbb{R} \supset (-\delta, +\delta) \to \mathbb{R}$$

die Einschränkung von f_i auf die Parallele durch x_0 zur j-ten Koordinatenachse, d.h.

$$\varphi_{ij}(t) := \big(p_i \circ f \circ k_j\big)(t) = f_i(x_0 + te_j).$$

Nach der Kettenregel ist diese zusammengesetzte Funktion an der Stelle $t = 0$ differenzierbar und die Ableitung ist

$$\varphi_{ij}(0) = D\big(p_i \circ f \circ k_j\big)(0) = Dp_i(f(x_0))Df(x_0)Dk_j(0) = e_i^T Df(x_0)e_j = \big(Df(x_0)\big)_{ij},$$

d.h. dies ist das Element an der Position (i, j) in der Ableitungsmatrix $Df(x_0)$. Natürlich ist dies gleichzeitig die gewöhnliche Ableitung der Funktion $f_i(x_0 + te_j)$ nach t und die hat einen besonderen Namen:

Definition MD.6 *Für $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}, x_0 \in \Omega$ heißt die Ableitung der Funktion*

$$\varphi_j(t) := f(x_0 + te_j)$$

an der Stelle $t = 0$ die "partielle Ableitung von f nach der Variablen x_j". Es sind dafür verschiedene Bezeichnungen gebräuchlich wie

$$\partial_j f(x_0) := \partial_{x_j} f(x_0) := \frac{\partial f}{\partial x_j}(x_0) := f_{x_j}(x_0) := \frac{d}{dt}\big(f(x_0 + te_j)\big)\big|_{t=0}.$$

Zusammen mit den obigen Überlegungen bekommen wir hieraus

Satz MD.7 *Ist $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n, x \mapsto (f_1(x), ..., f_n(x))^T$ in $x_0 \in \Omega$ differenzierbar, so sind sämtliche Komponentenfunktionen f_i in x_0 nach allen Variablen partiell differenzierbar und für die* JACOBI*-Matrix $Df(x_0)$ gilt die Darstellung*

$$Df(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0) \right)_{n,m}.$$

Die JACOBI*-Matrix enthält also in Zeile i stets nur die Ableitungen der Funktion f_i und in Spalte j stets nur Ableitungen nach der Variablen x_j.*

Unter einer wichtigen Zusatzbedingung kann man dieses Ergebnis umkehren. Wir erhalten als

Satz MD.8 (Das Hauptkriterium für Differenzierbarkeit) *Existieren zu einer Funktion $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n$ an jeder Stelle $x_0 \in \Omega$ die partiellen Ableitungen der Komponentenfunktionen und sind diese auf ganz Ω* ***stetig****, so ist f auf Ω differenzierbar und die Ableitung ist die* JACOBI*-Matrix aus den partiellen Ableitungen*

$$Df(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0) \right)_{n,m}.$$

Warnung MD.9 *Ohne die Voraussetzung der Stetigkeit ist die obige Aussage* ***falsch****.*

Wir werden daher im Weiteren stets die Stetigkeit der Ableitungen fordern. Wir verabreden als

Bezeichnung MD.10 *Eine Funktion $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}^n$ heiße r - mal stetig differenzierbar, wenn sämtliche partiellen Ableitungen bis zur Ordnung r auf ganz Ω existieren und stetig sind. Wir notieren dafür*

$$C^r(\Omega, \mathbb{R}^n) := \{ f : \Omega \to \mathbb{R}^n \,\big|\, f \text{ ist } r \text{ - mal stetig differenzierbar} \}.$$

Im Falle $n = 1$ schreibt man dafür auch kurz $C^r(\Omega)$

Offensichtlich gilt

Satz MD.11 *$C^r(\Omega, \mathbb{R}^n)$ ist ein reeller Vektorraum.*

Ferner erhält man mit dem Hauptkriterium leicht

Satz MD.12 *Eine Funktion $f \in C^r(\Omega, \mathbb{R}^n)$ $(r \geq 1)$ ist an jeder Stelle $x_0 \in \Omega$ differenzierbar im Sinne von Definition MD.1 und die Funktion*

$$Df : \Omega \to \mathbb{R}^{n \times m} : \quad x_0 \to Df(x_0)$$

ist selbst stetig.

Im Falle einer Skalarfunktion $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}$ hatten wir die Ableitung "Gradient" genannt und mit $\nabla f(x_0)$ bezeichnet. Nach Satz MD.8 ist dann

$$\nabla f(x_0) = \big(\partial_1 f(x_0), ..., \partial_m f(x_0) \big)$$

also der (Zeilen-) Vektor aus den partiellen Ableitungen.
Von Skalarfunktionen über demselben Gebiet kann man Linearkombinationen bilden, ferner kann man sie multiplizieren und - wo der Nenner $\neq 0$ ist - dividieren. Dafür gilt

Satz MD.13 *Sind $f, g \in C^1(\Omega), \alpha, \beta \in \mathbb{R}$, so sind auch $\alpha f + \beta g$, $f \cdot g$ und $\frac{f}{g}$ (abgesehen von den Nennernullstellen) in $C^1(\Omega)$ und es gelten*

$$\begin{array}{lcl} \nabla(\alpha f + \beta g) & = & \alpha \nabla f + \beta \nabla g, \\ \nabla(f \cdot g) & = & f \nabla g + g \nabla f, \\ \nabla(\frac{f}{g}) & = & \frac{g \nabla f - f \nabla g}{g^2}. \end{array}$$

Zum **Beweis** nutzen wir die Kettenregel und die Hilfsfunktionen

$$\begin{array}{lll} H_1(u,v): & R^2 \to R: & (u,v) \mapsto \alpha u + \beta v, \\ H_2(u,v): & R^2 \to R: & (u,v) \mapsto u \cdot v, \\ H_3(u,v): & R^2 \supset \{(u,v) \big| v \neq 0\} \to R: & (u,v) \mapsto \frac{u}{v}. \end{array}$$

Diese Funktionen sind differenzierbar und somit auch die zusammengesetzten Funktionen $H_i(f,g)$, die offenbar gerade die zu betrachtenden Funktionen sind. Nach der Kettenregel ist

$$\nabla\big(H_i(f,g)\big) = \frac{\partial H_i}{\partial u}(f,g)\nabla f + \frac{\partial H_i}{\partial v}(f,g)\nabla g,$$

womit man sofort die behaupteten Darstellungen verifiziert. □

Der Gradient einer Funktion $f : \mathbb{R}^m \supset \Omega \to \mathbb{R}$ hat eine wichtige geometrische Bedeutung. Es sei $x_0 \in \Omega \subset \mathbb{R}^m$, ferner $v \in \mathbb{R}^m$ ein Einheitsvektor, also $|v|_2 = 1$. Dann ist für hinreichend kleine $t \in \mathbb{R}$ die Funktion

$$\varphi(t) := f(x_0 + tv) \qquad (-\delta, +\delta) \to \mathbb{R}$$

definiert und stetig differenzierbar. Nach der Kettenregel ist

$$\frac{d}{dt}\varphi\big|_{t=0} = \frac{d}{dt} f(x_0 + tv)\big|_{t=0} = \nabla f(x_0) \frac{d}{dt}(x_0 + tv)\big|_{t=0} = \nabla f(x_0) v,$$

wobei rechts Zeile × Spalte gerechnet wird, sodaß also letztlich das Skalarprodukt $\langle \nabla f(x_0), v \rangle$ zu bilden ist.

Definition MD.14 *Wir nennen*

$$\frac{d}{dt}\varphi\big|_{t=0} = \frac{d}{dt} f(x_0 + tv)\big|_{t=0}$$

die Richtungsableitung von f in Richtung des Einheitsvektors v, d.h. mit $|v|_2 = 1$. Ist f in x_0 differenzierbar, so ist die Richtungsableitung in Richtung v gerade gleich dem eben erhaltenen Ausdruck $\langle \nabla f(x_0), v \rangle$.

Die Richtungsableitung in Richtung eines kanonischen Einheitsvektors ist offensichtlich gerade die entsprechende partielle Ableitung.

Nach der CAUCHY-SCHWARZschen Ungleichung gilt bezüglich der EUKLID-Norm für jeden Einheitsvektor v :

$$|\langle \nabla f(x_0), v \rangle| \leq |\nabla f(x_0)|,$$

wobei Gleichheit genau dann eintritt, wenn $v = \frac{\pm 1}{|\nabla f(x_0)|} \nabla f(x_0)$ ist, d.h. ein Einheitsvektor in Richtung des Gradienten. Dies bedeutet

Satz MD.15 *Der Gradient zeigt in die Richtung des stärksten Anstiegs von f.*

Wie hatten die Richtungsableitung erhalten, indem wir die Funktion längs eines Geradenstückes betrachtet hatten. Dies kann man allgemeiner machen.

Eine C^1-Funktion

$$\alpha : I \to \mathbb{R}^m : t \mapsto \alpha(t) = (\alpha_1(t), ..., \alpha_m(t))^T$$

auf einem reellen Intervall I nennen wir einen C^1 - Weg. Für jedes $t_0 \in I$ liefert die Ableitung

$$\alpha'(t) := \begin{pmatrix} \alpha_1'(t) \\ \vdots \\ \alpha_m'(t) \end{pmatrix}$$

den Tangentenvektor an den Weg an der Stelle $\alpha(t_0)$. (Geometrisch sinnvoll ist dies nur, wenn $\alpha'(t_0) \neq 0$ ist.)
Bei der Deutung von t als Zeit und von $\alpha(t)$ als Ort ist dann $\alpha'(t)$ die momentane Geschwindigkeit. Die Kettenregel besagt dann einfach

Satz MD.16 *Ist $f \in C^1(\Omega)$ und $\alpha : I \to \Omega$ ein C^1-Weg, so ist*

$$\frac{d}{dt}(f(\alpha(t))) = \langle (\nabla f)(a(t)), \alpha'(t) \rangle.$$

Sorgt man dafür, daß die Tangente $\alpha'(t)$ jeweils die Norm 1 hat, so erhält man hier einfach die Richtungsableitung in Richtung der jeweiligen Tangente.
Dieser Satz hat eine Reihe nützlicher Folgerungen.

Satz MD.17 *Es sei $f \in C^1(\Omega)$, $\alpha : [0,1] \to \Omega$ ein C^1-Weg und $a := \alpha(0), b := \alpha(1)$. Dann ist*

$$f(b) - f(a) = \int_0^1 \langle (\nabla f)(a(t)), \alpha'(t) \rangle dt.$$

Man nennt dies den Mittelwertsatz in Integral - Form.

Beweis: Nach dem Hauptsatz ist

$$f(b) - f(a) = f(\alpha(1)) - f(\alpha(0)) = \int_0^1 \frac{d}{dt}\big(f(\alpha(t))\big) dt$$

und durch Einsetzen folgt mit Satz MD.16 die Behauptung. □

Damit bekommen wir eine vielseitig einsetzbare Abschätzung.

Satz MD.18 *Ist $f \in C^1(\Omega)$ und sind $a, b \in \Omega$ so, daß auch die Verbindungsstrecke*

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

so ist

$$|f(b) - f(a)| \leq M \cdot |b - a|,$$

wobei $M := \max\{|\nabla f(x)| \,\big|\, x \in S\}$.

Beweis: Mit $\alpha(t) := a + t(b-a)$ ist $\alpha'(t) = b - a$ und mittels der Ungleichung von CAUCHY-SCHWARZ folgt aus dem vorigen Satz

$$\begin{aligned} |f(b) - f(a)| = | \int_0^1 \langle (\nabla f)(a(t)), \alpha'(t) \rangle dt \,| &\leq \int_0^1 |\langle (\nabla f)(a(t)), \alpha'(t) \rangle| dt \\ &\leq \int_0^1 |(\nabla f)(a(t))| \cdot |b-a| \, dt \leq M \cdot |b-a|. \end{aligned}$$

□

Ferner bekommen wir

Satz MD.19 *Verschwindet der Gradient einer Funktion f auf einem (zusammenhängenden!) Gebiet Ω, so ist dort f konstant.*

Beweis: In einem Gebiet lassen sich je zwei Punkte a, b durch einen Weg α verbinden, sodaß man nach Satz MD.17 die Darstellung

$$f(b) - f(a) = \int_0^1 \langle (\nabla f)(a(t)), \alpha'(t) \rangle dt$$

hat. Verschwindet der Gradient auf Ω, so ist das Integral stets 0, somit f konstant. □

Höhere Ableitungen

Satz MD.20 (von H.A.Schwarz) *Für $f \in C^2(\Omega)$ ist für alle i, j :*

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f,$$

d.h. bei den "gemischten" zweiten Ableitungen kommt es nicht auf die Reihenfolge der Differentiation an.
Man beachte, daß dazu auch die zweiten Ableitungen ***stetig*** *sein müssen.*

Für besonders interessierte Leser sei der **Beweis** hier dargestellt:
Sei zunächst Ω das Einheitsquadrat im $\mathbb{R}^2$ und $f : (x, y) \mapsto f(x, y)$ zweimal stetig differenzierbar.Wir bilden zu festem x mit $|x| < 1$ die Funktion

$$F_x(y) := f(x, y) - f(0, y), \quad (|y| < 1).$$

Dann ist

$$\begin{aligned} F_x(y) - F_x(0) &= \big(f(x, y) - f(0, y)\big) - \big(f(x, 0) - f(0, 0)\big) \\ &= f(x, y) + f(0, 0) - \big(f(0, y) + f(x, 0)\big). \end{aligned}$$

Nach dem Mittelwertsatz ist dies auch darstellbar als

$$F_x(y) - F_x(0) = \big(\frac{d}{dy} F_x(\theta y)\big) \cdot y.$$

Es ist

$$\frac{d}{dy} F_x(\theta y) = \frac{\partial}{\partial y} f(x, \theta y) - \frac{\partial}{\partial y} f(0, \theta y)$$

und wieder nach dem Mittelwertsatz formt man dies um zu

$$\frac{d}{dy} F_x(\theta y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} f(\theta' x, \theta y) \cdot x.$$

Somit ist für $(x, y) \in \Omega$ mit gewissen $|\theta|, |\theta'| < 1$

$$F_x(y) - F_x(0) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} f(\theta' x, \theta y) \cdot x \cdot y.$$

Völlig analog erhält man für

$$G_y(x) := f(x, y) - f(x, 0), \quad (|x| < 1).$$

die Darstellung

$$G_y(x) - G_y(0) = \frac{\partial}{\partial y}\frac{\partial}{\partial x} f(\delta' x, \delta y) \cdot x \cdot y,$$

wobei $|\delta|, |\delta'| < 1$. Nun ist aber

$$F_x(y) - F_x(0) = G_y(x) - G_y(0)$$

und somit
$$\frac{\partial}{\partial x}\frac{\partial}{\partial y}f(\theta' x, \theta y)\cdot x\cdot y = \frac{\partial}{\partial y}\frac{\partial}{\partial x}f(\delta' x, \delta y)\cdot x\cdot y$$
und dies für alle x, y und passende $\delta, \delta', \theta, \theta'$. Wegen der Stetigkeit der zweiten Ableitungen ist dann auch
$$\frac{\partial}{\partial x}\frac{\partial}{\partial y}f(0,0) = \frac{\partial}{\partial y}\frac{\partial}{\partial x}f(0,0).$$
Den allgemeinen Fall führt man darauf zurück, indem man setzt
$$\varphi(u,v) := f(x_0 + ue_i + ve_j).$$
Dann ist
$$\frac{\partial\varphi}{\partial u}(u,v) = \langle\nabla f, e_i\rangle = \frac{\partial f}{\partial x_i}\Big|_{x_0+ue_i+ve_j}$$
und entsprechend
$$\frac{\partial}{\partial v}\frac{\partial}{\partial u}\varphi(u,v) = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i}f\Big|_{x_0+ue_i+ve_j},$$
woraus man die Behauptung abliest. □

Bezeichnung MD.21 *Zu einer $C^2(\Omega)$-Funktion f und einer Stelle $x_0 \in \Omega$ heißt die Matrix*
$$f''(x_0) := \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}f(x_0)\right)$$
die HESSE*-Matrix zu f. Sie ist nach dem Satz von* SCHWARZ *symmetrisch.*

Um etwa eine TAYLOR-Entwicklung für C^r-Funktionen f untersuchen zu können, verwenden wir die schon mehrfach benutzte Technik, die Funktion f nur auf einer Geraden durch den Aufpunkt x_0 zu betrachten.
Zu $h \in \mathbb{R}^m, \neq 0$ betrachten wir also
$$\psi(t) := f(x_0 + t\cdot h) \qquad \text{für } |t| < 1.$$
– Ω enthält ja mit x_0 eine ganze Kugel und wir nehmen an, daß $x_0 + h$ in dieser Kugel liegt.– Dann ist nach der Kettenregel diese Funktion stetig differenzierbar und es ist
$$\frac{d}{dt}\psi(t) = \langle\nabla f(x_0 + th), h\rangle = \sum_{i=1}^m \frac{\partial}{\partial x_i}f(x_0+th)\cdot h_i\,.$$
Nun kann man nochmal die Kettenregel anwenden und erhält
$$\begin{aligned}\frac{d^2}{dt^2}\psi(t) &= \frac{d}{dt}\Big(\sum_{i=1}^m \frac{\partial}{\partial x_i}f(x_0+th)\cdot h_i\Big)\\ &= \sum_{j=1}^m \frac{\partial}{\partial x_j}\Big(\sum_{i=1}^m \frac{\partial}{\partial x_i}f(x_0+th)\cdot h_i\Big)\cdot h_j\\ &= \sum_{i,j}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}f(x_0+th)\cdot h_i h_j\\ &= \langle h, f''(x_0+th)h\rangle,\end{aligned}$$
wobei f'' die HESSE-Matrix ist.
Entsprechend – mit vielen Indizes oder mit dem Tensor-Kalkül – kann man höhere Ableitungen von ψ darstellen, sofern die entsprechenden partiellen Ableitungen existieren und stetig sind.

Nun kann man mit ψ den Mittelwertsatz oder die einfachste weiterführende TAYLOR-Entwicklung ausführen und erhält

– beim Mittelwertsatz:

$$\psi(1) = \psi(0) + \psi'(\tau) \cdot 1 \qquad \text{mit } 0 < \tau < 1,$$

– bei TAYLOR:

$$\psi(1) = \psi(0) + \psi'(0) \cdot 1 + \frac{1}{2}\psi''(\tau) \cdot 1 \qquad \text{mit } 0 < \tau < 1.$$

Setzt man darin die oben berechneten Werte ein, so erhält man

Satz MD.22 (Mittelwertsatz, TAYLOR-Formel) *Für $f : \Omega \to \mathbb{R}^1$, $f \in C^2(\Omega)$, $x_0 \in \Omega$ und jedes $h \in R^m$, für das die Verbindungsstrecke von x_0 und $x_0 + h$ ganz in Ω liegt, gelten mit geeigneten $0 < \tau < 1$*

$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0 + \tau h), h\rangle,$$
$$f(x_0 + h) = f(x_0) + \langle \nabla f(x_0), h\rangle + \frac{1}{2}\langle h, f''(x_0 + \tau h)h\rangle,$$

wobei f'' wieder die HESSE-Matrix ist.

Mit dieser Darstellung können wir auch wieder lokale Extrema charakterisieren (vergl. Satz F.74).

Satz MD.23 *Es sei $f \in C^2(\Omega), x_0 \in \Omega$.*

1. *Hat f in x_0 ein lokales Minimum (Maximum), so ist $\nabla f(x_0) = 0$ und die HESSE-Matrix $f''(x_0)$ ist positiv (bzw. negativ) semidefinit.*
2. *Diese Bedingungen sind hinreichend, sofern die HESSE-Matrix sogar entsprechend definit (und nicht nur semidefinit) ist.*

Beweis: Für einen festen Vektor h betrachte wieder

$$\psi(t) := f(x_0 + th).$$

Hat nun f an x_0 ein Minimum bzw. Maximum, so hat entsprechend ψ ein solches Extremum an der Stelle $t = 0$ und nach Satz F.74 gilt

$$\psi'(0) = 0 \text{ und } \psi''(0) \geq 0 \text{ bzw. } \leq 0$$

Dies bedeutet aber nach dem eben Dargestellten, daß

$$\psi'(0) = \langle \nabla f(x_0), h\rangle = 0,$$
$$\psi''(0) = \langle h, f''(x_0)h\rangle \geq 0 \text{ bzw. } \leq 0$$

ist. Da hier h beliebig gewählt werden kann, heißt dies gerade, daß $\nabla f(x_0) = 0$ und daß die HESSE-Matrix semidefinit ist.

Ist die HESSE-Matrix $f''(x_0)$ sogar definit, so ist für alle $h \neq 0$ dann $\psi''(0) > 0$ bzw. stets $\psi''(0) < 0$ und mit Satz F.74 folgt, daß bei ψ und damit auch bei f ein Minimum bzw. Maximum vorliegt. $\square$

Implizite Funktionen

Bei der Behandlung von Funktionen in einer Variablen hatten wir festgestellt, daß eine differenzierbare Funktion f (lokal) eine differenzierbare Umkehrfunktion f^{-1} besitzt, sofern ihre Ableitung nicht verschwindet. Diese Situation sei nun im mehrdimensionalen Fall betrachtet, wo sie in unterschiedlicher Form auftritt.
Zunächst erweitern wir die Bezeichnung MD.10 zu

Definition MD.24 *Eine Abbildung $f : U \to V$ zwischen Gebieten $U, V \subset \mathbb{R}^n$ heißt ein C^r-Diffeomorphismus (zwischen U und V), wenn die Umkehrabbildung $f^{-1} : V \to U$ existiert und beide Abbildungen f und f^{-1} jeweils C^r-Abbildungen sind. Dabei ist stets $r \geq 1$ vorausgesetzt.*

Die üblicherweise in der Physik vorgenommenen Koordinaten-Transformationen sind – bei etwas Vorsicht, d.h. bei Beschränkung auf geeignete Teilbereiche – sämtlich Diffeomorphismen meist sogar mit $r = \infty$.

Beispiel MD.25 (Zylinder-Koordinaten)

$$f : \begin{pmatrix} r \\ \varphi \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z \end{pmatrix} := \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \\ z \end{pmatrix}.$$

Wählt man hier etwa $U = \{(r, \varphi, z)^T \,|\, r > 0,\ 0 < \varphi < 2\pi\}$, so bekommt man einen C^∞-Diffeomorphismus zwischen U und dem $\mathbb{R}^3$ ohne die Punkte der “Halbebene” $\{(x, y, z)^T \,|\, x \geq 0, y = 0\}$.

Aus der Kettenregel (Satz MD.4) folgt sofort

Satz MD.26 *Ist $f : U \to V$ ein C^r-Diffeomorphismus, so gilt für jede Stelle $u \in U$ und $v = f(u) \in V$:*

$$d\big(f^{-1}\big)(v) = \big(df(u)\big)^{-1},$$

d.h. an zusammengehörigen Stellen sind die Ableitungen d.h. die Jacobi*-Matrizen der Abbildung f und ihrer Umkehrabbildung f^{-1} zueinander inverse Matrizen.*

Für die Zylinder-Koordinaten haben wir

$$df = \begin{pmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Deren Inverse ist

$$\big(df\big)^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{r^2}\begin{pmatrix} r^2\cos\varphi & r^2\sin\varphi & 0 \\ -r\sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & r^2 \end{pmatrix}.$$

Nun ist $r = \sqrt{x^2 + y^2}$, $\cos\varphi = \frac{x}{r}$, $\sin\varphi = \frac{y}{r}$, somit

$$\big(df\big)^{-1} = \frac{1}{x^2 + y^2}\begin{pmatrix} x\sqrt{x^2 + y^2} & y\sqrt{x^2 + y^2} & 0 \\ -y & x & 0 \\ 0 & 0 & x^2 + y^2 \end{pmatrix}.$$

Die Ableitung an einer Stelle u_0 ist ja eine lineare Abbildung, die in der Nähe von u_0 die Funktion f sehr gut approximiert. Bis auf einen “geringen” Fehler kann man also in der Nähe von u_0 die Funktion

$$\varphi : u \mapsto f(u) - f(u_0)$$

ersetzen durch die lineare Abbildung

$$u \mapsto \big(df(u_0)\big)(u - u_0)$$

und diese Abbildung ist umkehrbar, sofern die Matrix $df(u_0)$ invertierbar ist. Diese durch die inverse Matrix gegebene Abbildung macht damit die Wirkung von f beinahe wieder rückgängig. Untersucht man dies genauer, so erhält man den zentralen

Satz MD.27 (von der Umkehrabbildung) *Es sei $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^n$ eine C^r-Abbildung ($r \geq 1$), deren Ableitung $df(u_0)$ an einer Stelle u_0 invertierbar sei. Dann gibt es Umgebungen U und V : $u_0 \in U \subset \Omega$, $f(u_0) \in V \subset \mathbb{R}^n$, sodaß f einen C^r-Diffeomorphismus $U \to V$ bewirkt.*
Ist $f : \Omega \to \Omega'$ bijektiv und stets $df(u)$ invertierbar, so ist auch die Umkehrfunktion $f^{-1} : \Omega' \to \Omega$ ein C^r-Diffeomorphismus.

Dieser Satz liefert einmal ein Kriterium, wann durch eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$ wenigstens lokal ein krummliniges Koordinatensystem gegeben ist. Eine viel wichtigere Anwendung ist aber sein Einsatz beim Auflösen von nichtlinearen Gleichungen.

Beispiel MD.28 *Es sei eine C^r-Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben mit $f(0,0) = 0$, etwa $f(x,y) = \sin(xe^{x+y} + \pi y)$. Frage: Kann man in der Nähe von $(0,0)$ nach der zweiten Variablen auflösen, d.h. gibt es eine "Lösungsfunktion" $y = \varphi(x)$, d.h. also eine Funktion $\varphi : \mathbb{R} \supset U(0) \to \mathbb{R}$, mit der wenigstens lokal (also in einer Umgebung $U(0)$) gilt*

$$f(x, \varphi(x)) = 0 \qquad (\text{und } \varphi(0) = 0.)$$

Dazu bilden wir die C^r-Abbildung

$$F : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ f(x,y) \end{pmatrix}.$$

Hier ist $F(0,0) = (0,0)^T$ und, sofern es die Umkehrabbildung

$$G : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} \psi_1(u,v) \\ \psi_2(u,v) \end{pmatrix}$$

gibt, ist

$$\begin{pmatrix} u \\ v \end{pmatrix} = F(G(u,v)) = \begin{pmatrix} \psi_1(u,v) \\ f(\psi_1(u,v), \psi_2(u,v)) \end{pmatrix}.$$

Damit ist dann $\psi_1(u,v) = u$ und $v = f(u, \psi_2(u,v))$. Setzen wir hierin $v := 0$ und dann $\varphi(u) := \psi_2(u,0)$, so ist offenbar $0 = f(u, \varphi(u))$, d.h. die gesuchte Lösung gefunden.
Um hier den Satz über die Umkehrabbildung anwenden zu können, müssen wir zeigen, daß die JACOBI-Matrix von F invertierbar ist. Nun ist

$$dF = \begin{pmatrix} 1 & 0 \\ \partial_x f & \partial_y f \end{pmatrix},$$

d.h. dF ist überall dort invertierbar, wo die partielle Ableitung $\partial_y f \neq 0$ ist. Da können wir den Satz über die Umkehrabbildung anwenden und den obigen Schluß durchführen.

Diese Überlegung läßt sich sofort auf den allgemeinen Fall verallgemeinern, wobei dann x und y jeweils zwei Pakete von p bzw. m vielen Variablen sind, also $x := (x_1, ..., x_p)$, $y := (y_1, ..., y_m)$ und entsprechend

$$f : \mathbb{R}^{p+m} \supset \Omega \to \mathbb{R}^m$$

eine C^r-Abbildung. Statt $\partial_y f \neq 0$ muß dann die entsprechende Teilmatrix der JACOBI-Matrix von f invertierbar sein. Dann geht alles völlig analog und wir erhalten den

Satz MD.29 (über implizite Funktionen) *Es sei*

$$f : \mathbb{R}^{p+m} \supset \Omega \to \mathbb{R}^m,\ (x,y) \mapsto f(x,y)$$

eine C^r-Abbildung mit $r \geq 1$. An einer Stelle $(x_0, y_0) \in \Omega$ sei $f(x_0, y_0) = 0$ und es gelte die "Auflöse-Bedingung"

$$\det\left(\frac{\partial f_i}{\partial y_k}\right)_{i,k=1,..,m} \neq 0.$$

Dann gibt es Umgebungen $x_0 \in U \subset \mathbb{R}^p$, $y_0 \in V \subset \mathbb{R}^m$, sodaß

$$U \times V := \{(x,y) \,|\, x \in U, y \in V\} \subset \Omega$$

und die Gleichung $f(x,y) = 0$ in $U \times V$ eindeutig nach y auflösbar ist, d.h. es gibt genau eine C^r-Abbildung $\varphi : U \to V$, sodaß

$$f(x,y) = 0 \Leftrightarrow y = \varphi(x) \quad \text{für } (x,y) \in U \times V.$$

Ferner ergibt sich durch Differentiation der Identität

$$0 = f(x, \varphi(x))$$

mit den Teilmatrizen der Jacobimatrix von f :

$$D_x f := \left(\frac{\partial f_i}{\partial x_j}\right)_{i=1,..,m;j=1,...,p} \quad \text{und} \quad D_y f := \left(\frac{\partial f_i}{\partial y_k}\right)_{i=1,..,m;j=1,...,m}$$

$$0 = D_x f(x,y)\big|_{y=\varphi(x)} + D_y f(x,y)\big|_{y=\varphi(x)} \cdot D\varphi\big|_{\varphi(x)}$$

woraus man mit der "Auflösebedingung" die Darstellung

$$D\varphi\big|_{\varphi(x)} = -\left(D_y f(x,y)\big|_{y=\varphi(x)}\right)^{-1} \cdot D_x f(x,y)\big|_{y=\varphi(x)}$$

erhält.

Auf den formalen Beweis verzichten wir. Man beachte, daß man zwar mit diesem Satz auf die Existenz einer entsprechend differenzierbaren "Lösungs-Abbildung" schließen, diese aber nur in Ausnahmefällen explizit angegeben werden kann.

Zur Anwendung dieses Satzes seien noch ein paar Hinweise gegeben.

– Häufig soll man ein Problem der Form $g(x,y) = c$ nach dem Variablensatz y auflösen. Hier setzt man einfach

$$f(x,y) := g(x,y) - c.$$

Für die Auflösebedingung kann man direkt mit g arbeiten, da ja für alle i, k

$$\frac{\partial f_i}{\partial y_k} = \frac{\partial g_i}{\partial y_k}.$$

- Setzen wir auf dem $\mathbb{R}^2$ einmal $g(x,y) := x^2 + 2y^2$, so ist einfach die partielle Ableitung nach y zu überprüfen, d.h. die Auflösebedingung lautet hier

  $$\frac{\partial g}{\partial y} = 4y \overset{!}{\neq} 0.$$

  Dies ist für $y \neq 0$ erfüllt.

  Trotzdem ist die Gleichung $g(x,y) = -1$ nirgends nach y auflösbar!

  Im Satz wird gesagt, daß wir in der Umgebung einer Stelle (x_0, y_0) auflösen können, wo $f(x_0, y_0) = 0$ ist, d.h. hier also $x_0^2 + 2y_0^2 = -1$ ist. So eine Stelle gibt es überhaupt nicht, also nützt die ganze Bedingung über die Ableitung nichts.

- Nehmen wir $g(x,y) = +1$, so finden wir (viele) passende Stellen (x_0, y_0). Aber in der Umgebung von $y_0 = 0$ ist die Auflösebedingung nicht erfüllt, sodaß wir nicht nach y auflösen können. Überlegen Sie selbst, daß **hier** eine Auflösung nach der Variablen x möglich ist.

Wir geben noch zwei spezielle Anwendungen.

Niveau-und Gradientenlinien MD.30 *$f : \mathbb{R}^2 \supset \Omega \to \mathbb{R}$ sei eine C^1-Funktion. Für $c \in \mathbb{R}$ heiße*

$$N_c := \{(x,y) \in \Omega \,\big|\, f(x,y) = c\}$$

eine Niveaumenge von f. N_c kann für gewisse c leer sein.
Ist $N_c \neq \emptyset$ und $\nabla f(x,y) \neq (0,0)$ auf N_c, so kann man lokal die Niveaugleichung $f(x,y) = c$ nach x oder nach y auflösen, d.h. man erhält eine Beschreibung der Niveaumengen als C^1-Wege

$$t \mapsto (t, y(t)) \quad \text{oder} \quad t \mapsto (x(t), t).$$

Man redet daher auch von Niveaulinien.
Ein C^1-Weg $\psi : I \to \Omega$, dessen Tangentenvektor $\psi'(t)$ stets $\neq 0$ ist und an jeder Stelle in Richtung des Gradienten von f zeigt

$$\psi'(t) = \lambda(t) \cdot \big(\nabla f\big)(\psi(t)) \neq 0 \quad \text{mit } \lambda(t) > 0$$

heißt Gradientenlinie.

Für diese Linien gilt

Satz MD.31 *Gradienten- und Niveaulinien schneiden sich stets senkrecht.*
Sie können damit lokal als rechtwinkeliges krummliniges Koordinatensystem verwendet werden.

Beweis: Ist $s \mapsto \varphi(s) = (s, y(s))$ eine Niveaulinie, so ist $f(\varphi(s)) = c$, somit

$$0 = \frac{d}{ds}\big(f(\varphi(s))\big) = \frac{\partial f}{\partial x}\frac{ds}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} = \langle \big(\nabla f\big)(\varphi(s)), \varphi'(s) \rangle.$$

Am Schnittpunkt (x_0, y_0) von Gradienten- und Niveaulinie ist dann

$$\big(\nabla f\big)(\varphi(s)) = \big(\nabla f\big)(x_0, y_0) = \big(\nabla f\big)(\psi(t)) = \frac{1}{\lambda(t)}\psi'(t),$$

woraus die Behauptung unmittelbar folgt. $\square$

Extrema unter Nebenbedingungen

Extremstellen einer Funktion $f : \mathbb{R}^m \supset \Omega \to R$, die im Innern des Definitionsgebietes Ω liegen, hatten wir schon im Zusammenhang mit der mehrdimensionalen TAYLOR-Entwicklung behandelt (Satz MD.23). Damit lassen sich aber Aufgaben wie die folgende nur schwer bearbeiten:
Bestimme das Maximum (oder Minimum) der Funktion $g(x, y, z) := x^2 + y^2 + z^2$ *unter den Nebenbedingungen*

$$f_1(x, y, z) := x^2 + 2y^2 + 3z^2 - 1 = 0,$$
$$f_2(x, y, z) := x + y - z = 0.$$

Die Abbildung

$$f : \mathbb{R}^3 \to \mathbb{R}^2 : \ (x, y, z) \mapsto \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \end{pmatrix}$$

beschreibt dabei über die Forderung $f = 0$ sogenannte "Nebenbedingungen". Hier wäre also das Maximum oder Minimum der Funktion g zu suchen unter der Nebenbedingung, daß sie nur auf dem Durchschnitt des Ellipsoids $f_1 = 0$ mit der Ebene $f_2 = 0$ betrachtet werden soll.
Die JACOBI-Matrix von f lautet

$$df = \begin{pmatrix} 2x & 4y & 6z \\ 1 & 1 & -1 \end{pmatrix}.$$

Sie ist auf der Ebene $f_2 = 0$ abgesehen von dem Punkt $(0, 0, 0)$, der aber nicht infrage kommt, stets vom Rang 2, d.h. $f = 0$ kann stets lokal aufgelöst werden. Diese Auflösung kann man in g einsetzen und die entstehende Extremwertaufgabe in einer Variablen lösen.
In unserem Beispiel ist dieser Weg gerade noch möglich, im allgemeinen Fall wird man das nicht praktisch ausführen können. Es gibt aber einen viel einfacheren Weg. Betrachten wir gleich die allgemeine Situation.

Gegeben seien C^1*-Funktionen* $g, f_1, ..., f_m : \mathbb{R}^n \supset \Omega \to \mathbb{R}^1$, *wobei* $m < n$. *Es sei*

$$N := \{x \in \Omega \,|\, f_1(x) = ... = f_m(x) = 0\}.$$

Wir suchen lokale Extrema von g *auf* N, *d.h. Punkte* $x_0 \in N$, *sodaß*

$$g(x_0) \le g(x) \quad \text{bzw. } g(x_0) \ge g(x)$$

für alle $x \in U(x_0) \cap N$.
Solche Punkte heißen lokales Minimum bzw. lokales Maximum unter den Nebenbedingungen $f_1 = ... = f_m = 0$.

Dafür gilt die folgende notwendige Bedingung, die häufig die Menge der möglichen Extrema so eingrenzt, daß man die tatsächlich vorhandenen Minima und Maxima direkt erkennen kann.

Satz MD.32 *Es sei* N *die Lösungsmenge von* $f_1 = ... = f_m = 0$ *und* g *habe in* $x_0 \in N$ *ein lokales Extremum. Ferner seien an* x_0 *die Gradienten* $\nabla f_1(x_0), ..., \nabla f_m(x_0)$ *linear unabhängig.*
Dann gibt es eindeutig bestimmte Zahlen $\lambda_1, ..., \lambda_m \in \mathbb{R}$, *die sog.* LAGRANGE-*Multiplikatoren, sodaß*

$$\nabla g(x_0) = \sum_{i=1}^{m} \lambda_i \nabla f_i(x_0).$$

Beweis: Für die Abbildung $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^m : x \mapsto (f_1(x), ..., f_m(x))$ erlauben die Voraussetzungen, die Gleichung $f(x) = 0$ in der Umgebung jeder Extremalstelle x_0 nach m Variablen aufzulösen. Bei passender Nummerierung bekommen wir also mit $p := n-m$ dann C^1-Funktionen $\varphi_1(u), ..., \varphi_m(u)$, wobei $u := (u_1, \ldots, u_p) \in \mathbb{R}^p$, die lokal in einer Umgebung Ω' von $u_0 := (x_{01}, ..., x_{0p}) \in \mathbb{R}^p$ definiert sind und dort

$$0 = f(u_1, ..., u_p, \varphi_1(u), ..., \varphi_m(u))$$

erfüllen. Mit der Abbildung Φ :

$$\Phi(u) := (u_1, ..., u_p, \varphi_1(u), ..., \varphi_m(u))$$

gilt dann also in einer Umgebung von $u_0 \in \mathbb{R}^p$

$$0 = f(\Phi(u))$$

und damit auch

$$0 = \big(df\big)(x_0)\big(d\Phi\big)(u_0).$$

Letzteres ist gleichbedeutend damit, daß

$$0 = \langle \nabla f_i, \partial_k \Phi \rangle \quad \text{für } i = 1, ..., m;\ k = 1, ..., p$$

ist, d.h. daß der Gradient ∇f_i jeder "Bedingungsfunktion" senkrecht auf dem Unterraum U steht, der von den Spalten von $d\Phi$ aufgespannt wird. Dieser Unterraum hat wegen der speziellen Gestalt der Spalten von $d\Phi$ die Dimension p.
An einer Extremstelle x_0 $(\in N)$ hat die Funktion $u \mapsto \big(g \circ \Phi\big)(u)$ ein lokales Extremum, somit verschwindet der Gradient, d.h. es ist

$$0 = \nabla(g \circ \Phi) = \nabla g \cdot d\Phi,$$

was wieder gleichbedeutend damit ist, daß

$$0 = \langle \nabla g, \partial_k \Phi \rangle \quad \text{für alle } k = 1, ..., p,$$

sodaß auch $\nabla g \in U^\perp$.
Also sind die $m+1$ vielen Vektoren $\nabla g, \nabla f_1, ..., \nabla f_m$ (jeweils an der Extremstelle x_0) sämtlich im orthogonalen Komplement des p-dimensionalen Unterraumes $U \subset \mathbb{R}^n$, was wegen $p + m = n$ bedeutet, daß sie linear abhängig sind. Da die Gradienten ∇f_i nach Voraussetzung unabhängig sind, folgt die Behauptung. □

Bemerkung MD.33 *Zum Bestimmen solcher Extremalpunkte* $(x_{01}, ..., x_{0n})$ *hat man also für die* $n+m$ *vielen Unbekannten* $x_1, ..., x_n, \lambda_1, ..., \lambda_m$ *die Gleichungen*

$$f_i(x_1, ..., x_n) = 0 \qquad (i = 1, ..., m),$$
$$\partial_k g(x_1, ..., x_n) = \sum_{i=1}^{m} \lambda_i \partial_k f_i(x_1, .., x_n)$$

zu lösen.

Das folgende Beispiel enthält die Problemstellung in etwas verwischter Form.

Beispiel MD.34 *Es sei*

$$g(x, y, z) := x^2 + 2xy - \frac{2}{3}z^3,$$
$$f(x, y, z) := x^2 + \frac{1}{2}y^2 + \frac{1}{2}z^2.$$

Man bestimme ein Maximum von g *unter der Nebenbedingung* $f \le 1$.

Der Bereich, auf dem $f \leq 1$ ist, ist offenbar kompakt, sodaß ein Maximum (und ein Minimum) existiert. Die Nebenbedingung hat aber zunächst nicht die gewünschte Form. Wir zerlegen deshalb die Aufgabe in zwei Teile:

- Bestimme Maxima im Innern, d.h. für $f < 1$ und
- Maxima auf dem Rand, d.h. für $f = 1$.

Zum ersten:
Es ist $\nabla g(x,y,z) = (2x+2y, 2x, -2z^2)$. Die einzige Stelle, wo dies verschwindet, ist $(0,0,0)$, was zwar $f(0,0,0) < 1$ erfüllt, aber sicher kein Maximum von g ist. In diesem Innen-Bereich liegt also kein Maximum.

Auf dem Rand:
Es ist $\nabla f(x,y,z) = (2x, y, z)$. Somit sind nach Bemerkung MD.33 Werte für x, y, z, λ so zu bestimmen, daß gilt

$$0 = f(x,y,z) - 1 = x^2 + \frac{1}{2}y^2 + \frac{1}{2}z^2 - 1$$

und

$$(2x+2y, 2x, -2z^2) = \lambda(2x, y, z)$$

Es ist also das folgende nichtlineare Gleichungssystem zu lösen:

$$\begin{aligned} x^2 + \frac{1}{2}y^2 + \frac{1}{2}z^2 - 1 &= 0, \\ 2x + 2y &= \lambda 2x, \\ 2x &= \lambda y, \\ -2z^2 &= \lambda z. \end{aligned}$$

Dies liefert endlich viele Stellen (x,y,z) und nur an denen kann ein Maximum von g unter der Nebenbedingung $f = 1$ vorliegen. An welchen tatsächlich ein Maximum vorliegt, hat man einzeln durchzuprüfen.
Führen Sie das aus!