

Von Gipfeln und Tälern – Differenzialrechnung für Funktionen von mehreren Variablen

2



Wie leitet man Funktionen mit mehreren Variablen ab?

Wozu kann man die Ableitung im \mathbb{R}^n gebrauchen?

Wie löst man Extremwertaufgaben mit mehreren Variablen?

Was sind implizit definierte Funktionen?

Wie leitet man vektorwertige Funktionen ab?

2.1	Partielle Ableitungen	18
2.2	Differenzierbarkeit	21
2.3	Die Taylor-Entwicklung im \mathbb{R}^n	28
2.4	Gradient und Richtungsableitung	30
2.5	Extremwertaufgaben ohne Nebenbedingung	34
2.6	Extremwertaufgaben mit Nebenbedingung	45
2.7	Die Methode der kleinsten Fehlerquadrate – eine Minimierungsaufgabe im \mathbb{R}^n	48
2.8	Implizit definierte Funktionen	53
2.9	Die Ableitung für Funktionen von $\mathbb{R}^n \rightarrow \mathbb{R}^m$	58
	Aufgaben	62

In Band 1, Kap. 6, haben wir das lokale Änderungsverhalten für Funktionen von *einer* Variablen untersucht. Das zentrale Hilfsmittel war die Ableitung, die wir uns als Tangentensteigung veranschaulichen konnten. Sie hatte sich als nützlich erwiesen, um beispielsweise Funktionen zu linearisieren oder Extremwerte zu bestimmen. In diesem Kapitel werden wir untersuchen, wie sich der Begriff der Ableitung zunächst auf reellwertige Funktionen mit mehreren Variablen verallgemeinern lässt. Wir werden dann sehen, wie sich dieses Handwerkzeug nutzen lässt, um beispielsweise in der Fehlerrechnung mithilfe des totalen Differenzials Fehler abschätzen zu können oder Extremwertaufgaben mit mehreren Variablen zu lösen. Danach werden wir uns mit implizit definierten Funktionen beschäftigen und uns zum Abschluss des Kapitels überlegen, wie wir die Ableitung auch für vektorwertige Funktionen definieren können.

2.1 Partielle Ableitungen

Wir betrachten zunächst eine einfache reellwertige Funktion f mit zwei Variablen, die in Abb. 2.1 dargestellt ist:

$$f(x, y) = x^2 + y.$$

Was geschieht, wenn wir uns eine der Variablen herausgreifen und f nach dieser Variablen ableiten? Halten wir die Variable $y = y_0$ fest, so hängt die verbleibende Funktion nur noch von x ab. Es ergibt sich die eingezeichnete Schnittparabel

$$z = f(x, y_0) = x^2 + y_0$$

als Graph. Man spricht von der **Parameterlinie** $y = y_0$. Wenn wir uns für die Steigung dieser Schnittkurve im Punkt $(x_0, y_0) \in D$ interessieren, betrachten wir den Grenzwert

$$g := \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}.$$

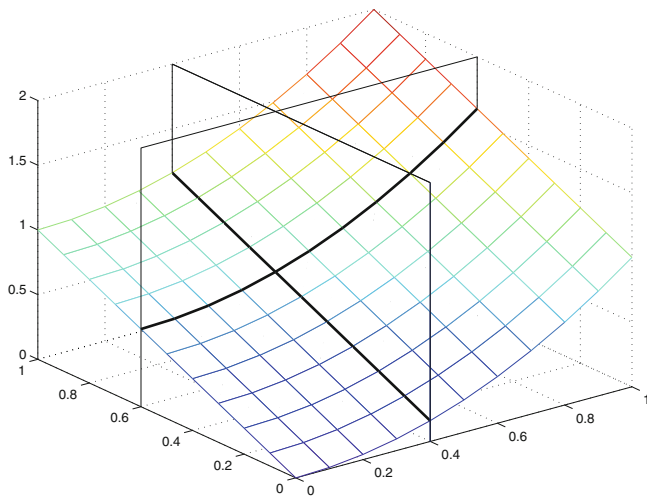


Abb. 2.1 Die Funktion f mit $f(x, y) = x^2 + y$ und die beiden Parameterlinien für $x = 0.4$ und $y = 0.6$

In unserem Beispiel ist

$$\begin{aligned} g &= \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x_0 + h)^2 + y_0 - x_0^2 - y_0}{h} \\ &= \lim_{h \rightarrow 0} \frac{2x_0h + h^2}{h} = 2x_0, \end{aligned}$$

also genau die Ableitung der Schnittparabel im Punkt x_0 . Allgemein sagen wir, eine Funktion f sei im Punkt (x_0, y_0) **partiell differenzierbar** nach x , wenn dieser Grenzwert existiert. Der Grenzwert heißt **partielle Ableitung** von f nach x im Punkt (x_0, y_0) , geschrieben:

$$\frac{\partial f}{\partial x}(x_0, y_0) = g, \quad \frac{\partial f(x_0, y_0)}{\partial x} = g \quad \text{oder} \quad f_x(x_0, y_0) = g.$$

Die partielle Ableitung $f_x(x_0, y_0)$ beschreibt also das Änderungsverhalten von f an der Stelle (x_0, y_0) in x -Richtung, wenn $y = y_0$ fest gelassen wird. Anders ausgedrückt gibt sie die Steigung der Tangente parallel zur Ebene $y = y_0$ an das Schaubild von f im Punkt (x_0, y_0) an.

Die partielle Ableitung nach y erhalten wir aus

$$\begin{aligned} \frac{\partial f}{\partial y}(x_0, y_0) &= \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{x_0^2 + y_0 + h - (x_0^2 + y_0)}{h} = 1. \end{aligned}$$

Partiell differenzieren heißt nach einer einzelnen Variablen ableiten

Entsprechende Überlegungen können wir natürlich auch für weitere Variablen anstellen, wenn f von mehr als zwei Variablen abhängt. Wir können also jede beliebige Variable herausgreifen und nach ihr **partiell differenzieren**.

Definition: Partielle Differenzierbarkeit

Sei $D \subset \mathbb{R}^n$ offen, $\hat{x} \in D$ ein Punkt und $f : D \rightarrow \mathbb{R}$ eine skalarwertige Funktion.

- f heißt in \hat{x} **partiell** nach x_j **differenzierbar**, $j = 1, \dots, n$, wenn der Grenzwert

$$\frac{\partial f(\hat{x})}{\partial x_j} := \lim_{h \rightarrow 0} \frac{f(\hat{x}_1, \dots, \hat{x}_j + h, \dots, \hat{x}_n) - f(\hat{x})}{h}$$

existiert.

- $\frac{\partial f(\hat{x})}{\partial x_j}$ heißt **partielle Ableitung** von f nach x_j an der

Stelle \hat{x} . Kürzer schreiben wir auch $f_{x_j}(\hat{x}) = \frac{\partial f(\hat{x})}{\partial x_j}$.

- Die Funktion f heißt in \hat{x} **partiell differenzierbar**, wenn sie dort nach jeder Variablen x_j partiell differenzierbar ist.
- Die Funktion f heißt in D **partiell differenzierbar**, wenn sie für alle $\hat{x} \in D$ partiell differenzierbar ist.

Diese Definition ist auf den ersten Blick einigermaßen abschreckend. Glücklicherweise können wir partielle Ableitungen auch ohne diese Definition ganz einfach mit den bekannten Techniken der Differenziation berechnen.

Berechnung der partiellen Ableitung

- Um f_{x_j} zu berechnen, werden alle Variablen außer x_j als fest betrachtet (also wie Parameter behandelt), und der Funktionsterm wird nach x_j abgeleitet.
- Dabei gelten die üblichen Ableitungsregeln, die aus dem Eindimensionalen bereits bekannt sind.

Wir demonstrieren dieses Vorgehen an einigen Beispielen.

Beispiel

1. Für $f(x, y) = x^2 + y$ aus dem obigen Beispiel ist

$$f_x = 2x \quad \text{und} \quad f_y = 1.$$

2. Für $f(x, y) = x \cdot e^y + \cos(xy)$ ist

$$f_x = e^y - y \sin(xy) \quad \text{und} \quad f_y = x \cdot e^y - x \sin(xy).$$

3. Für $f(x, y, z) = x \cdot y^2 + z^3$ ist

$$f_x = y^2, \quad f_y = 2xy \quad \text{und} \quad f_z = 3z^2. \quad \blacktriangleleft$$

Beispiel

Wir haben in Abschn. 1.1 erwähnt, dass sich die Temperatur T eines Punktes auf einer Herdplatte mit den Koordinaten (x, y) zum Zeitpunkt t als Funktion von drei Variablen auffassen lässt: $T = T(x, y, t)$. Für eine kreisrunde Herdplatte, die in der Mitte am wärmsten ist, ist der Temperaturverlauf in Abb. 2.2 dargestellt. Auf der t -Achse (d. h. nach hinten) ist am Rand zu erkennen, wie sich die Platte im Lauf der Zeit abkühlt.

Physikalisch gesehen bedeuten nun die partiellen Ableitungen T_x bzw. T_y die räumliche und T_t die zeitliche Änderungsrate der Temperatur. Befinden wir uns etwa zum Zeitpunkt t_0 im Punkt (x_0, y_0) und bewegen uns um ein sehr kleines Stück Δy nach unten, so ändert sich die Temperatur ungefähr um den Betrag $T_y(x_0, y_0, t_0) \cdot \Delta y$. Verbleiben wir hingegen am Ort (x_0, y_0) , so ändert sich

dort die Temperatur in der sehr kurzen Zeitspanne Δt näherungsweise um $T_t(x_0, y_0, t_0) \cdot \Delta t$. Die partiellen Ableitungen erlauben uns also genau wie im Eindimensionalen, das Änderungsverhalten des Funktionswertes bei Änderung einer einzelnen Variablen zu beschreiben.

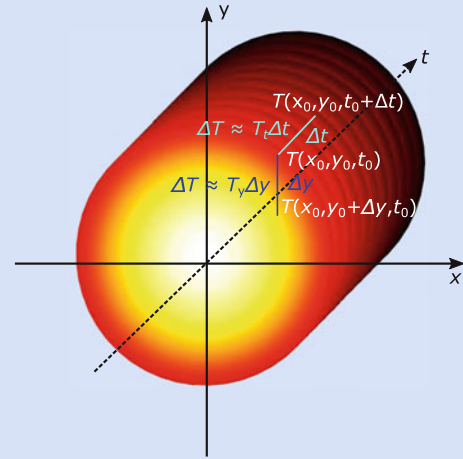


Abb. 2.2 Die örtliche Temperaturverteilung zum Zeitpunkt $t = 0$ sowie der zeitliche Temperaturverlauf am Rand der Herdplatte ▶

Beispiel

Betrachten wir das schwingende Seil mit der vom Ort $x \in [0, L]$ und der Zeit $t \geq 0$ abhängenden Auslenkung $u(x, t)$, so können wir uns $u_x(x_0, t_0)$ als die örtliche Änderung der Auslenkung (also die Steigung des Seiles) an einer bestimmten Stelle x_0 zu einem festen Zeitpunkt t_0 vorstellen. Die partielle Ableitung $u_t(x_0, t_0)$ gibt die zeitliche Änderung der Auslenkung an einer festen Stelle x_0 zu einem bestimmten Zeitpunkt t_0 an. Sie sagt uns also, mit welcher Geschwindigkeit (der Physiker spricht hier von der **Schnelle**) sich das Seil an einer bestimmten Stelle auf und ab bewegt.

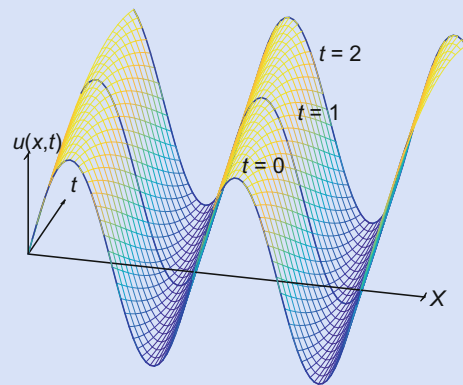


Abb. 2.3 Die Auslenkung $u(x, t)$ eines sinusförmig schwingenden Seiles in Abhängigkeit vom Ort x und der Zeit t ▶

Mehrfaches Differenzieren ergibt höhere partielle Ableitungen

Wenn wir eine Funktion nach einer beliebigen Variablen partiell abgeleitet haben, ist das Ergebnis wieder eine Funktion. Diese können wir selbstverständlich wieder nach jeder beliebigen Variablen ableiten und erhalten so eine **partielle Ableitung zweiter Ordnung**. Formal erhält man für eine Funktion von zwei Variablen die folgenden vier partiellen Ableitungen:

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

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

Es ist zu beachten, dass die Reihenfolge der Ableitung bei der *∂*-Schreibweise von *rechts nach links* zu lesen ist, bei der „Index“-Schreibweise hingegen von *links nach rechts*. So bedeutet die Notation

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

dass auf f zunächst die Rechenoperation „Ableiten nach y “ und auf das Ergebnis dann die Operation „Ableiten nach x “ angewandt wird. Andererseits ist die Schreibweise

$$f_{yx} = (f_y)_x$$

so zu verstehen, dass zunächst f_y gebildet wird und das Ergebnis dann nach x differenziert wird. Wie wir bald sehen werden, ist die Reihenfolge in der Praxis aber meist ohnehin nicht von Bedeutung.

Auf die gleiche Weise können wir partielle Ableitungen noch höherer Ordnung definieren, z. B.

$$\frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial y \partial x} \right) = \frac{\partial^3 f}{\partial y \partial y \partial x} = f_{xyy}, \quad \frac{\partial}{\partial x} \left(\frac{\partial^2 f}{\partial y^2} \right) = f_{yyx}, \dots$$

Beispiel

1. Für die Funktion f mit $f(x, y) = x^2 + y$ haben wir

$$f_x = 2x, \quad \text{also} \quad f_{xx} = 2 \quad \text{und} \quad f_{xy} = 0, \\ f_y = 1, \quad \text{also} \quad f_{yx} = 0 \quad \text{und} \quad f_{yy} = 0.$$

2. Für die Funktion f mit $f(x) = y^2 \sin x$ ist

$$f_x = y^2 \cos x, \\ \text{also} \quad f_{xx} = -y^2 \sin x \quad \text{und} \quad f_{xy} = 2y \cos x, \\ f_y = 2y \sin x, \\ \text{also} \quad f_{yx} = 2y \cos x \quad \text{und} \quad f_{yy} = 2 \sin x.$$

In Abb. 2.4 sind die Zusammenhänge zwischen der Funktion und ihren partiellen Ableitungen dargestellt.

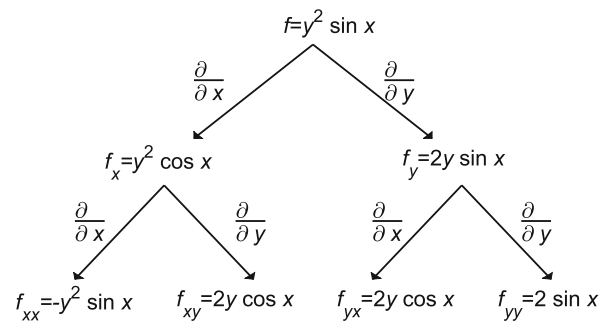


Abb. 2.4 Der „Stammbaum“ der partiellen Ableitungen von f

In beiden Fällen ist $f_{xy} = f_{yx}$, es scheint also ohne Belang zu sein, in welcher Reihenfolge wir ableiten. Ist das Zufall?

Der Satz von Schwarz: Es ist egal, in welcher Reihenfolge wir ableiten

Dass es bei fast allen praktischen Anwendungen nicht auf die Reihenfolge der partiellen Ableitung ankommt, ist die Aussage des wichtigen Satzes von Schwarz.

Satz von Schwarz

Sei $D \subset \mathbb{R}^n$ offen, $f : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar (d. h. alle möglichen partiellen Ableitungen bis zur zweiten Ordnung seien stetig). Dann gilt

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i} \quad \text{für alle} \quad \mathbf{x} \in D, \quad i, j = 1, \dots, n,$$

d. h., die gemischten Ableitungen sind vertauschbar.

Quasi alle Funktionen, die uns im „täglichen Leben“ begegnen, erfüllen die Voraussetzungen des Satzes von Schwarz. Im mathematischen Hintergrund 2.1 findet der interessierte Leser ein Beispiel einer Funktion, bei der $f_{xy} \neq f_{yx}$ ist.

Durch wiederholte Anwendung des Satzes von Schwarz lässt sich zeigen, dass auch zwei partielle Ableitungen der Ordnung $m > 2$ gleich sind, wenn f eine m -mal stetig differenzierbare Funktion ist und beide Male jeweils gleich oft nach den einzelnen Variablen abgeleitet wurde, völlig egal in welcher Reihenfolge dies geschehen ist. Beispielsweise gilt

$$f_{xxy} = f_{xyx} = f_{yxx}.$$

Wir fassen dieses Ergebnis zusammen:

Für eine Funktion f , deren partielle Ableitungen bis zur m -ten Ordnung stetig sind, sind diese partiellen Ableitungen nicht abhängig davon, in welcher Reihenfolge partiell differenziert wurde.

2.1 Mathematischer Hintergrund: Eine Funktion mit $f_{xy} \neq f_{yx}$

Wir betrachten die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ aus Abb. 2.5 mit

$$f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

Wir haben bereits im mathematischen Hintergrund 1.1 gezeigt, dass diese Funktion stetig ist. Nun schauen wir uns ihre partiellen Ableitungen an: Da $f(x, 0) = f(0, y) = 0$ ist, verschwinden in $(0, 0)$ die partiellen Ableitungen: $f_x(0, 0) = f_y(0, 0) = 0$. In allen anderen Punkten, d. h. $(x, y) \neq (0, 0)$, ist

$$f_x = y \left(\frac{x^2 - y^2}{x^2 + y^2} + \frac{4x^2 y^2}{(x^2 + y^2)^2} \right)$$

und

$$f_y = x \left(\frac{x^2 - y^2}{x^2 + y^2} - \frac{4x^2 y^2}{(x^2 + y^2)^2} \right).$$

Also ist

$$f_{xy}(0, 0) = \lim_{y \rightarrow 0} \frac{f_x(0, y) - f_x(0, 0)}{y} = \lim_{y \rightarrow 0} -\frac{y^3}{y^3} = -1$$

und

$$f_{yx}(0, 0) = \lim_{x \rightarrow 0} \frac{f_y(x, 0) - f_y(0, 0)}{x} = \lim_{x \rightarrow 0} \frac{x^3}{x^3} = 1.$$

In diesem (zugegebenermaßen ziemlich weit hergeholten) Beispiel stimmen also f_{xy} und f_{yx} nicht überein. Folglich müssen die Voraussetzungen des Satzes von Schwarz verletzt sein. In der Tat sind hier zweite Ableitungen in $(0, 0)$ nicht stetig.

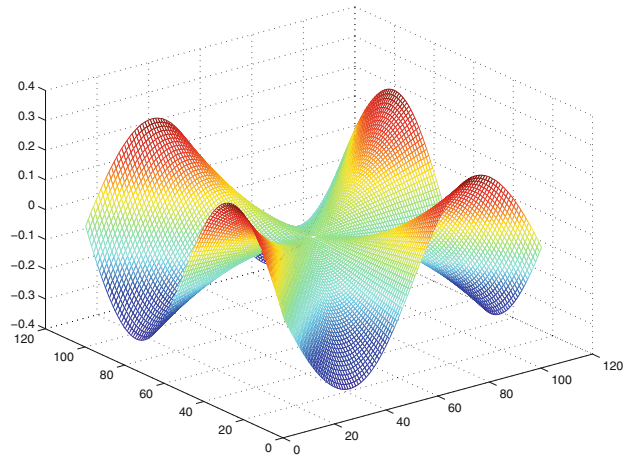


Abb. 2.5 Die Funktion f sieht harmlos aus, hat es aber in sich

2.2 Differenzierbarkeit

Eine differenzierbare Funktion von einer Variablen konnten wir mithilfe ihrer Ableitung linearisieren, also die Tangente in

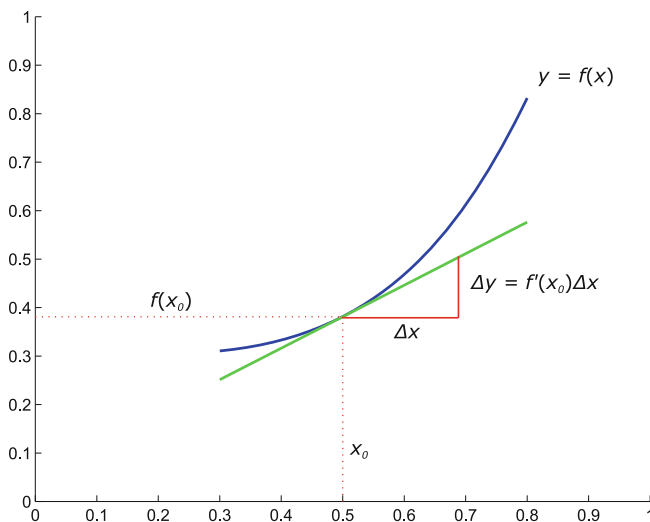


Abb. 2.6 Linearisierung einer Funktion von einer Variablen durch die Tangente

einem bestimmten Punkt bestimmen. (Abb. 2.6). Die Tangentengleichung lautete

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Anders ausgedrückt: Die Ableitung $f'(x_0)$ dient als Faktor, der eine Änderung $\Delta x = x - x_0$ in x -Richtung *linear* in eine Änderung in y -Richtung (Funktionswertänderung auf der Tangente) „übersetzt“:

$$\Delta y = f'(x_0) \cdot \Delta x.$$

Zur Erinnerung: Eine Funktion f von einer Variablen ist genau dann differenzierbar an der Stelle x_0 , wenn wir im Punkt $(x_0, f(x_0))$ die Tangente an das Schaubild von f bestimmen können. Dies ist dann möglich, wenn das Schaubild an der Stelle x_0 glatt ist, also keinen „Knick“ aufweist.

Die Tangentialebene linearisiert Funktionen von zwei Variablen

Wir wenden uns Funktionen von zwei Variablen zu. Das Schaubild einer solchen Funktion f ist bekanntlich ein Flächenstück im Raum. Wenn wir sie an einer bestimmten Stelle $(x_0, y_0) \in \mathbb{R}^2$ linearisieren wollen, läuft das anschaulich darauf hinaus, die

Tangentialebene an das Schaubild in $(x_0, y_0, f(x_0, y_0))$ zu bestimmen.

Definition: Differenzierbarkeit im \mathbb{R}^2

Eine Funktion f von zwei Variablen heißt differenzierbar an der Stelle (x_0, y_0) , wenn im Punkt $(x_0, y_0, f(x_0, y_0))$ die eindeutige Tangentialebene an das Schaubild von f existiert. Dies ist der Fall, wenn das Schaubild an der Stelle (x_0, y_0) glatt ist, also keine „Spitzen“ oder „Grate“ wie in Abb. 2.7 aufweist.

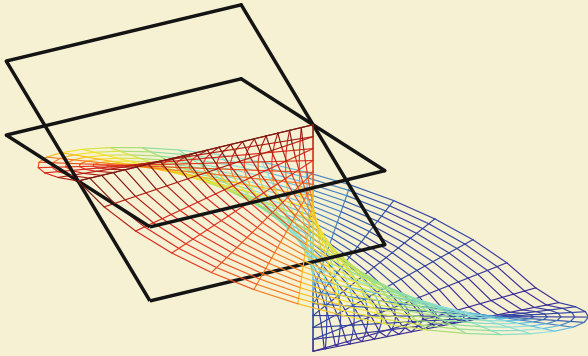


Abb. 2.7 An Spitzen oder Graten kann man keine Tangentialebenen legen, sie würden „wegkippen“

Die Idee ist in Abb. 2.8 skizziert. Bewegen wir uns nur in x -Richtung von der Stelle (x_0, y_0) weg, so ändert sich der Funktionswert auf der x -Tangente (in der Abbildung rot) um

$$f_x(x_0, y_0)(x - x_0).$$

Bewegen wir uns hingegen in y -Richtung von der Stelle (x_0, y_0) weg, so ändert sich der Funktionswert auf der y -Tangente (in der Abbildung blau) um

$$f_y(x_0, y_0)(y - y_0).$$

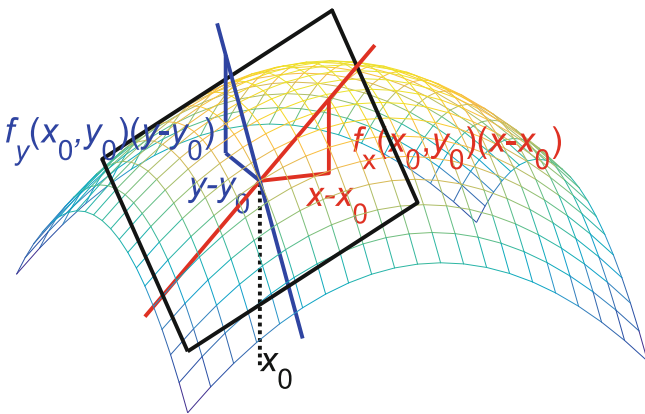


Abb. 2.8 Linearisierung einer Funktion von zwei Variablen durch die Tangentialebene. Diese wird aufgespannt durch die Tangenten in x -Richtung (rot) und in y -Richtung (blau)

Wenn wir diese beiden Änderungen zusammenfassen und zum Funktionswert in (x_0, y_0) addieren, erhalten wir die Gleichung der Tangentialebene:

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Definition: Tangentialebene

Ist die Funktion f an der Stelle (x_0, y_0) differenzierbar, so lautet die Gleichung der Tangentialebene

$$T : z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Mithilfe der Tangentialebene können wir also das Änderungsverhalten einer Funktion in der Nähe einer bestimmten Stelle (x_0, y_0) beschreiben und Näherungswerte für die Funktionswerte bestimmen.

Beispiel

Wir bestimmen die Tangentialebene der Funktion f mit $f(x, y) = x^3 y^2$ an der Stelle $(x_0, y_0) = (1, 2)$. Dann ist

$$f_x = 3x^2 y^2, \quad f_y = 2x^3 y,$$

und damit

$$f(1, 2) = 4, \quad f_x(1, 2) = 12, \quad f_y(1, 2) = 4.$$

Also ist die Tangentialebene in $(1, 2)$

$$T : z = 4 + 12(x - 1) + 4(y - 2).$$

Wenn wir uns nun z. B. für den Funktionswert $f(0.9, 2.1)$ interessieren, können wir mit der Tangentialebene sehr einfach (sogar im Kopf) einen Näherungswert berechnen: $f(0.9, 2.1) \approx T(0.9, 2.1) = 4 + 12(0.9 - 1) + 4(2.1 - 2) = 4 - 1.2 + 0.4 = 3.2$. Der wahre Funktionswert wäre $f(0.9, 2.1) = 0.9^3 \cdot 2.1^2 = 3.21489$ gewesen – dazu braucht man aber einen Taschenrechner. ◀

Von der Tangentialebene zur Ableitung

Wir schreiben die Gleichung der Tangentialebene nun etwas um, indem wir die partiellen Ableitungen in der 1×2 -Matrix $A = (f_x(x_0, y_0) \quad f_y(x_0, y_0))$ zusammenfassen:

$$\begin{aligned} T : z &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &= f(x_0, y_0) + A \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}. \end{aligned}$$

Die Situation ist hier ganz ähnlich wie bei einer Funktion von einer Variablen: Zum Funktionswert am „Aufpunkt“ wird eine lineare Änderung addiert, die sich jetzt als Produkt der Matrix A mit dem „Abweichungsvektor“

$$\mathbf{h} = \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

ergibt. Die Matrix A übernimmt nun also die Rolle der Ableitung, und wir können schreiben:

$$\begin{aligned} T : z &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &= f(x_0, y_0) + f'(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\ &= f(x_0, y_0) + f'(x_0, y_0) \cdot \mathbf{h}. \end{aligned}$$

Definition: Ableitung im \mathbb{R}^2

Ist die Funktion f an der Stelle (x_0, y_0) differenzierbar, so ist ihre Ableitung die 1×2 -Matrix

$$f'(x_0, y_0) = \begin{pmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \end{pmatrix}.$$

Wie schon im Eindimensionalen „übersetzt“ also die Ableitungsmatrix eine Änderung \mathbf{h} im Funktionsargument linear in die näherungsweise Änderung des Funktionswertes:

$$\Delta z = f' \cdot \mathbf{h}.$$

Die Ableitung für Funktionen von mehreren Variablen ist eine Matrix

Dieses Konzept lässt sich nun leicht auf Funktionen mit noch mehr Variablen verallgemeinern.

Definition: Differenzierbarkeit im \mathbb{R}^n

Eine Funktion $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ heißt differenzierbar an der Stelle $\mathbf{x}_0 \in D$, wenn sie dort linearisierbar ist, d. h., die eindeutige **Tangentialnäherung** (die „ n -dimensionale Tangentialebene“) existiert.

$$T : z = f(\mathbf{x}_0) + f'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$$

existiert.

Achtung $f'(\mathbf{x}_0)$ ist eine $(1 \times n)$ -Matrix und $\mathbf{x} - \mathbf{x}_0$ ein Vektor. Das Produkt in der obigen Darstellung ist also im Sinne eines Matrix-Vektor-Produkts zu verstehen. \blacktriangleleft

Für Funktionen mit einer Variablen ist die Tangentialnäherung eine Gerade (nämlich die Tangente), für Funktionen mit

zwei Variablen eine Ebene (die oben erklärte Tangentialebene). Allgemein ist die Tangentialnäherung im \mathbb{R}^n eine $n - 1$ -dimensionale Ebene, eine sog. **Hyperebene** – allerdings kann man sich dieses Gebilde für Funktionen mit mehr als zwei Variablen nicht mehr anschaulich vorstellen. Dennoch erlaubt uns die Tangentialnäherung, die Funktion f in der Nähe von \mathbf{x}_0 durch eine lineare Funktion zu approximieren.

Beispiel

Für die Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ mit $f(x, y, z) = \frac{x + y^2}{z}$ ist

$$f_x = \frac{1}{z}, \quad f_y = \frac{2y}{z} \quad \text{und} \quad f_z = -\frac{x + y^2}{z^2}.$$

An der Stelle $\mathbf{x}_0 = (x_0, y_0, z_0) = (1, 3, 2)$ haben wir $f(1, 3, 2) = 5$ sowie die partiellen Ableitungen

$$f_x(1, 3, 2) = 0.5, \quad f_y(1, 3, 2) = 3, \quad f_z(1, 3, 2) = -2.5.$$

Als Tangentialnäherung erhalten wir

$$\begin{aligned} T : u &= 5 + \begin{pmatrix} 0.5 & 3 & -2.5 \end{pmatrix} \cdot \begin{pmatrix} x - 1 \\ y - 3 \\ z - 2 \end{pmatrix} \\ &= 0.5 + 0.5x + 3y - 2.5z. \end{aligned}$$

Eine Approximation für den Funktionswert beispielsweise an der Stelle $\mathbf{x} = (1.1, 2.9, 2.1)$ ist damit

$$u = 0.5 + 0.5 \cdot 1.1 + 3 \cdot 2.9 - 2.5 \cdot 2.1 = 4.5.$$

Das ist eine ziemlich gute Näherung für den wahren Funktionswert

$$f(1.1, 2.9, 2.1) = \frac{1.1 + 2.9^2}{2.1} = 4.529. \quad \blacktriangleleft$$

Nun ist es nicht mehr schwierig, den Begriff der Ableitung auf Funktionen von noch mehr Variablen zu verallgemeinern. Um den Abweichungsvektor $\mathbf{h} = \mathbf{x} - \mathbf{x}_0 \in \mathbb{R}^n$ linear in eine Funktionswertänderung zu übersetzen, müssen wir ihn mit einer $1 \times n$ -Matrix multiplizieren, die analog zur Ableitungsmatrix für Funktionen von zwei Variablen aufgebaut ist.

Definition: Die Ableitung im \mathbb{R}^n

Ist die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differenzierbar im Punkt $\mathbf{x}_0 \in \mathbb{R}^n$, so ist ihre Ableitung dort eine $1 \times n$ -Matrix, nämlich

$$f'(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}.$$

Diese Ableitungsmatrix übersetzt wiederum die Änderung \mathbf{h} im Argumentvektor (x_1, \dots, x_n) in eine Funktionswertänderung

$$f'(\mathbf{x}_0) \cdot \mathbf{h}$$

auf einer verallgemeinerten „Tangentialebene“, auch wenn wir uns diese für Funktionen von mehr als drei Variablen nicht mehr bildlich vorstellen können.

Differenzierbarkeit ist mehr als die Existenz der partiellen Ableitungen

Bisher haben wir Differenzierbarkeit ziemlich anschaulich als „Linearisierbarkeit“ definiert und uns überlegt, dass eine differenzierbare Funktion „glatt“ sein muss, ihr Schaubild also beispielsweise keine Sprünge, Kanten, Spitzen oder Grate haben darf. Wie können wir nun erkennen, ob eine gegebene Funktion an einem bestimmten Punkt \mathbf{x}_0 differenzierbar ist? Es leuchtet ein, dass zunächst einmal alle partiellen Ableitungen in \mathbf{x}_0 existieren müssen, wenn f dort differenzierbar sein soll – sonst könnten wir die Ableitungsmatrix A gar nicht hinschreiben. Allerdings reicht das noch nicht aus. Im mathematischen Hintergrund 2.2 zeigen wir eine Funktion, deren partielle Ableitungen zwar alle existieren, die aber trotzdem nicht differenzierbar ist.

Dieses Beispiel zeigt, dass Differenzierbarkeit im \mathbb{R}^n mehr beinhaltet als nur die Existenz der partiellen Ableitung – wir müssen noch weitere Forderungen stellen. Leider erweist sich deren praktische Anwendung als ziemlich unhandlich. Wir behandeln daher für den interessierten Leser die exakte Definition im mathematischen Hintergrund 2.3 und erwähnen hier nur ein hinreichendes Kriterium.

Hinreichendes Kriterium für Differenzierbarkeit im \mathbb{R}^n

Sind alle partiellen Ableitungen der Funktion f auf der offenen Menge $D \subset \mathbb{R}^n$ stetig, so ist f in jedem Punkt in D differenzierbar.

Die meisten Funktionen, die uns im täglichen Leben begegnen, besitzen glücklicherweise stetige partielle Ableitungen.

Beispiel

Für die bereits bekannte Funktion f mit $f(x, y, z) = \frac{x + y^2}{z}$ sind die partiellen Ableitungen

$$f_x = \frac{1}{z}, \quad f_y = \frac{2y}{z} \quad \text{und} \quad f_z = -\frac{x + y^2}{z^2}$$

auf dem gesamten Definitionsbereich stetig, also ist f dort differenzierbar. ◀

Das totale Differenzial ist eine Näherung für die Funktionswertänderung

Betrachten wir die Gleichung der Tangentialnäherung etwas genauer, so stellen wir fest, dass stets zum Funktionswert im Punkt $\hat{\mathbf{x}}$ (dem „Aufpunkt“ der Tangentialebene) der „Zuwachs“ auf der Tangentialebene, nämlich

$$dz = f'(\hat{\mathbf{x}}) \cdot (\mathbf{x} - \hat{\mathbf{x}})$$

oder, ausgeschrieben,

$$dz = f_{x_1}(\hat{\mathbf{x}}) \underbrace{(x_1 - \hat{x}_1)}_{\Delta x_1} + f_{x_2}(\hat{\mathbf{x}}) \underbrace{(x_2 - \hat{x}_2)}_{\Delta x_2} + \dots + f_{x_n}(\hat{\mathbf{x}}) \underbrace{(x_n - \hat{x}_n)}_{\Delta x_n}$$

addiert wird. Die Terme $\Delta x_1 = x_1 - \hat{x}_1$ usw. geben an, wie weit wir uns in den einzelnen Achsen vom „Aufpunkt“ $\hat{\mathbf{x}}$ entfernt haben. Sie werden auch als „Zuwächse“ in den einzelnen Variablen bezeichnet – was genau genommen nicht ganz korrekt ist, weil sie ja auch negativ sein können.

Dieser „Zuwachs auf der Tangentialebene“ ist so wichtig, dass er einen eigenen Namen bekommen hat.

Definition: Totales Differenzial

Ist $f: \mathbb{R}^n \rightarrow \mathbb{R}$ an der Stelle $\hat{\mathbf{x}}$ differenzierbar, so heißt

$$dz := f_{x_1}(\hat{\mathbf{x}}) \Delta x_1 + f_{x_2}(\hat{\mathbf{x}}) \Delta x_2 + \dots + f_{x_n}(\hat{\mathbf{x}}) \Delta x_n$$

totales Differenzial von f an der Stelle $\hat{\mathbf{x}}$ zu den Zuwächsen Δx_1 bis Δx_n .

Für eine Funktion von 2 Veränderlichen ist also das totale Differenzial im Punkt (x_0, y_0) zu den Zuwächsen Δx und Δy gegeben durch

$$dz = f_x(x_0, y_0) \Delta x + f_y(x_0, y_0) \Delta y$$

und beschreibt die *Änderung des Funktionswertes auf der Tangentialebene*.

Dieser Zuwachs ist zumindest für kleine Zuwächse Δx und Δy eine recht gute Näherung für die *tatsächliche Änderung des Funktionswertes*

$$\Delta z = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0).$$

Allerdings ist das totale Differenzial meist erheblich einfacher auszuwerten als die Funktion selbst, weshalb man das Änderungsverhalten von Funktionen gerne mithilfe des totalen Differenzials untersucht.

Beispiel

Wir betrachten nochmals die Funktion f mit $f(x, y) = x^3 y^2$ an der Stelle $(x_0, y_0) = (1, 2)$. Dann ist das totale Differenzial im Punkt $(1, 2)$ zu den Zuwächsen Δx und Δy gegeben durch

$$dz = 12(x - 1) + 4(y - 2) = 12\Delta x + 4\Delta y. \quad \blacktriangleleft$$

2.2 Mathematischer Hintergrund: Existierende partielle Ableitungen, aber keine Differenzierbarkeit

Die Funktion f aus Abb. 2.9 mit

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0), \end{cases}$$

ist bereits eine alte Bekannte. Wir wissen, dass sie in $(0, 0)$ nicht stetig ist. Deshalb kann sie dort erst recht nicht differenzierbar sein. Aber es ist $f(x, 0) = f(0, y) = 0$, also

$$f_x(0, 0) = \lim_{x \rightarrow 0} \frac{f(x, 0) - f(0, 0)}{x} = \lim_{x \rightarrow 0} \frac{0}{x} = 0$$

und

$$f_y(0, 0) = \lim_{y \rightarrow 0} \frac{f(0, y) - f(0, 0)}{y} = \lim_{y \rightarrow 0} \frac{0}{y} = 0.$$

Also ist f in $(0, 0)$ nicht differenzierbar, obwohl dort beide partiellen Ableitungen existieren.

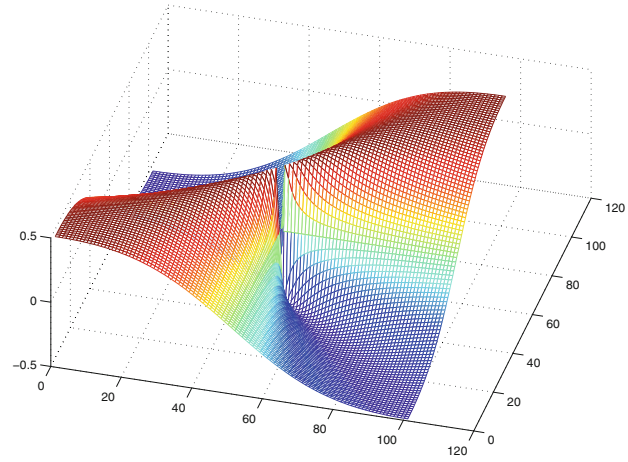


Abb. 2.9 In $(0, 0)$ ist f nicht stetig, geschweige denn differenzierbar

2.3 Mathematischer Hintergrund: Differenzierbarkeit im \mathbb{R}^n

Wir betrachten einen Punkt $\mathbf{x}_0 = (x_1, \dots, x_n)$, in dem alle partiellen Ableitungen existieren, sowie einen „Abweichungsvektor“ $\mathbf{h} = (h_1, \dots, h_n)$ und versuchen, f zu linearisieren:

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{\partial f}{\partial x_1}(\mathbf{x}_0)h_1 + \dots + \frac{\partial f}{\partial x_n}(\mathbf{x}_0)h_n + r(\mathbf{h}).$$

Wenn nun für kleiner werdendes \mathbf{h} der Rest $r(\mathbf{h})$ so schnell gegen 0 geht, dass sogar $\lim_{|\mathbf{h}| \rightarrow 0} \frac{r(\mathbf{h})}{|\mathbf{h}|} = 0$ ist, also

$$\lim_{|\mathbf{h}| \rightarrow 0} \frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \frac{\partial f}{\partial x_1}(\mathbf{x}_0)h_1 - \dots - \frac{\partial f}{\partial x_n}(\mathbf{x}_0)h_n}{|\mathbf{h}|} = 0,$$

so heißt f differenzierbar in \mathbf{x}_0 . Die Ableitung ist dann die $(1, n)$ -Matrix

$$A = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}_0) \quad \frac{\partial f}{\partial x_2}(\mathbf{x}_0) \quad \dots \quad \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \right).$$

Wir haben hier ein Konzept „herübergerettet“, das uns bereits bei der Ableitung im Eindimensionalen begegnet ist: Auch dort musste die Abweichung $r(h)$ zwischen dem Funktionswert $f(x_0 + h)$ und der Näherung $f(x_0) + f'(x_0) \cdot h$ so schnell gegen 0 gehen, dass sogar $r(h)/h \rightarrow 0$ geht, wenn h immer kleiner wird. Wir kommen darauf in Abschn. 2.3 zurück, wenn wir die Taylor-Formel ausführlich behandeln.

An diesem Beispiel ist einfach zu erkennen: Das totale Differenzial ist nichts weiter als die Tangentialnäherung, bei welcher der „Funktionswert im Aufpunkt“ weggelassen wird. Uns interessiert in diesem Fall nicht der Funktionswert auf der Tangentialebene, sondern nur die Abweichung (der „Zuwachs“) bezüglich des Aufpunktes.

Das totale Differenzial ist nützlich bei der Fehlerrechnung

Das totale Differenzial spielt eine wichtige Rolle bei der Fehlerrechnung in der Physik. Dabei geht es darum, den Einfluss

von Fehlern in den Messdaten auf das Ergebnis einer Formel abzuschätzen. Hängt unsere Formel beispielsweise von zwei Messgrößen - sagen wir x und y - ab, für die wir die Werte x_0 und y_0 messen, so erhalten wir das Ergebnis $f(x_0, y_0)$. Leider sind Messungen stets mit Fehlern behaftet. Die tatsächlichen Werte $x = x_0 + \Delta x$ und $y = y_0 + \Delta y$ weichen um gewisse Beträge Δx bzw. Δy ab, die wir natürlich nicht kennen, sondern nur abschätzen können. Zu diesen Werten gehört das eigentlich richtige Ergebnis $f(x_0 + \Delta x, y_0 + \Delta y)$. Wir wollen nun die Abweichung (d. h. den absoluten Fehler)

$$|\Delta z| = |f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)|$$

abschätzen. (Die Betragsstriche setzen wir deshalb, weil uns nur die Größe, nicht aber das Vorzeichen des Messfehlers interes-

siert). Hier kommt nun das totale Differenzial ins Spiel. Für hinreichend kleine Abweichungen dürfen wir nämlich Δz durch dz ersetzen:

$$\begin{aligned} |\Delta z| &\approx |dz| = |f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y| \\ &\leq |f_x(x_0, y_0)| |\Delta x| + |f_y(x_0, y_0)| |\Delta y|. \end{aligned}$$

Im letzten Schritt haben wir die Dreiecksungleichung angewendet. Die Fehler in den Messdaten werden also mit den Beträgen der partiellen Ableitungen gewichtet und aufaddiert. Die Betragsstriche bewirken, dass sich Fehlerbestandteile mit verschiedenem Vorzeichen nicht einfach „wegheben“ können – schließlich sind wir an einer Worst-Case-Abschätzung interessiert.

Wir demonstrieren dieses Vorgehen zunächst an der Messung des Volumens eines Kegels.

Beispiel

Ein kegelförmiger Behälter mit Radius r und Höhe h hat bekanntlich das Volumen $V(r, h) = \frac{\pi}{3} r^2 h$. Der Radius wird mit $r = 10 \pm 0.1$ cm gemessen, die Höhe mit 50 ± 0.3 cm. In welchem Bereich liegt das tatsächliche Volumen?

Mit den gemessenen Werten ergibt sich das Volumen $V = 5236 \text{ cm}^3$. Für die fehlerbedingte Abweichung nach unten und oben erhalten wir

$$\Delta V \approx dV = \frac{\partial V}{\partial r} \Delta r + \frac{\partial V}{\partial h} \Delta h,$$

also

$$\begin{aligned} |\Delta V| &\leq \left| \frac{2\pi}{3} rh \right| |\Delta r| + \left| \frac{\pi}{3} r^2 \right| |\Delta h| \\ &= \left| \frac{2\pi}{3} 50 \cdot 10 \right| \cdot 0.1 + \left| \frac{\pi}{3} 10^2 \right| \cdot 0.3 = 136.1 \text{ cm}^3. \end{aligned}$$

Also liegt das tatsächliche Volumen im Intervall

$$5236 \pm 136.1 \text{ cm}^3.$$

Fast wie im Eindimensionalen: Die Ableitungsregeln

Die einfachsten Ableitungsregeln im Eindimensionalen sind die Summen- und die Faktorregel. Auch für Produkte und Quotienten reellwertiger Funktionen gelten entsprechende Ableitungsregeln.

Summen- und Faktorregel

Sind die Funktionen $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ an der Stelle \mathbf{x} differenzierbar, so gilt dies auch für $f + g$ und $\lambda \cdot f$, und es gilt

$$(f + g)'(\mathbf{x}) = f'(\mathbf{x}) + g'(\mathbf{x})$$

sowie

$$(\lambda \cdot f)'(\mathbf{x}) = \lambda \cdot f'(\mathbf{x}), \lambda \in \mathbb{R}.$$

Ebenso einfach verallgemeinern wir die Produkt- und die Quotientenregel.

Produkt- und Quotientenregel

Sind die reellwertigen Funktionen f, g an der Stelle $\mathbf{x} \in \mathbb{R}^n$ differenzierbar, so gilt dies auch für $f \cdot g$, und wir haben

$$(f \cdot g)'(\mathbf{x}) = f(\mathbf{x})g'(\mathbf{x}) + g(\mathbf{x})f'(\mathbf{x}).$$

Gilt zudem noch $g(\mathbf{x}) \neq 0$, so ist

$$\left(\frac{f}{g} \right)'(\mathbf{x}) = \frac{g(\mathbf{x})f'(\mathbf{x}) - f(\mathbf{x})g'(\mathbf{x})}{g^2(\mathbf{x})}.$$

Beispiel

Wir betrachten die Funktionen f und g mit $f(x, y) = x^2 + y^2$ und $g(x, y) = x - y$. Es gilt $f'(x, y) = \begin{pmatrix} 2x & 2y \end{pmatrix}$ und $g'(x, y) = \begin{pmatrix} 1 & -1 \end{pmatrix}$. Damit ist mit der Produktregel

$$\begin{aligned} (f \cdot g)'(x, y) &= (x^2 + y^2) \begin{pmatrix} 1 & -1 \end{pmatrix} + (x - y) \begin{pmatrix} 2x & 2y \end{pmatrix} \\ &= \begin{pmatrix} 3x^2 - 2xy + y^2 & -x^2 + 2xy - 3y^2 \end{pmatrix}. \end{aligned}$$

Bei direkter Rechnung mit

$$(f \cdot g)(x, y) = x^3 + xy^2 - yx^2 - y^3$$

erhalten wir

$$(f \cdot g)'(x, y) = \begin{pmatrix} 3x^2 - 2yx + y^2 & 2xy - x^2 - 3y^2 \end{pmatrix}.$$

Für den Quotienten $f/g = (x^2 + y^2)/(x - y)$ gilt entsprechend

$$\begin{aligned} \left(\frac{f}{g} \right)'(x, y) &= \frac{(x - y) \begin{pmatrix} 2x & 2y \end{pmatrix} - (x^2 + y^2) \begin{pmatrix} 1 & -1 \end{pmatrix}}{(x - y)^2} \\ &= \frac{1}{(x - y)^2} \begin{pmatrix} x^2 - 2xy - y^2 & x^2 + 2xy - y^2 \end{pmatrix}. \end{aligned}$$

Mathematik für Ingenieure: Verstehen – Rechnen –
Anwenden

Band 2: Analysis in mehreren Variablen,

Differenzialgleichungen, Optimierung

Göllmann, L.; Hübl, R.; Pulham, S.; Ritter, S.; Schon, H.;

Schöffler, K.; Voß, U.; Vossen, G.

2017, XIV, 481 S. 343 Abb., 308 Abb. in Farbe.,

Softcover

ISBN: 978-3-662-53864-7