14.5 Gradient

Für *skalare* Funktionen gibt es schließlich noch einen dritten Ableitungsbegriff, den des *Gradienten*. Ist $f : V \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so definiert die totale Ableitung

$$Df(a) : \ V \to \mathbb{R}, \quad h \mapsto Df(a)h$$

ein *lineares Funktional* auf V. Ist V nun ein *Hilbertraum* mit einem Skalarprodukt $\langle \cdot , \cdot \rangle$, so existiert aufgrund des Rieszschen Darstellungssatzes [3] genau ein Vektor $\phi \in V$ mit der Eigenschaft, dass

$$Df(a)h = \langle \phi , h \rangle , \qquad h \in V.$$

Man sagt, *ϕ stellt $Df(a)$ bezüglich $\langle \cdot , \cdot \rangle$ dar*. Dieser Vektor ist der *Gradient* von f im Punkt a.

15 **Definition** *Sei V ein Hilbertraum. Ist $f : V \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so ist der Gradient von f an der Stelle a der eindeutig bestimmte und mit $\nabla f(a)$ bezeichnete Vektor in V mit der Eigenschaft, dass*

$$Df(a) = \langle \nabla f(a), \cdot \rangle .$$

16 A. Sind $l \in V$ und $x_0 \in V$, so ist

$$L : \ V \to \mathbb{R}, \quad L(x) = \langle l, x - x_0 \rangle$$

differenzierbar mit $DL(x)h = \langle l, h \rangle$. Also ist

$$\nabla L(x) = l, \qquad x \in V.$$

B. Ist $A : V \to V$ linear und symmetrisch, also $\langle Av, w \rangle = \langle v, Aw \rangle$ für alle $v, w \in V$, so ist

$$f : \ V \to \mathbb{R}, \quad f(x) = \langle Ax, x \rangle$$

differenzierbar mit Ableitung $Df(x)h = 2 \langle Ax, h \rangle$. Somit ist

$$\nabla f(x) = 2Ax.$$

Der Gradient ist also die *Darstellung* der totalen Ableitung einer skalaren Funktion bezüglich eines Skalarproduktes. Ohne Bezug auf ein Skalarprodukt macht es keinen Sinn, von einem Gradienten zu sprechen. Besonders wichtig ist natürlich der Standardraum $\mathbb{R}^n$ mit dem Standardskalarprodukt.

16 **Satz** *Im Standardraum $\mathbb{R}^n$ mit dem Standardskalarprodukt ist der Gradient einer differenzierbaren Funktion $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}$ der Spaltenvektor*

$$\nabla f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix} = Df^\top . \quad \rtimes$$

Der Gradient ∇f ist also ein *Spaltenvektor*, während die Ableitung Df einer skalaren Funktion durch einen *Zeilenvektor* dargestellt wird. Dies ist ein wichtiger Unterschied, der zum Beispiel Folgen hat für deren Verhalten unter Koordinatentransformationen.

16 ⟨⟨⟨⟨⟨ Für den Zeilenvektor $Df(x)$ und den Spaltenvektor h gilt

$$Df(x)h = \langle Df^\top(x), h\rangle = \langle \nabla f(x), h\rangle$$

genau dann, wenn $\nabla f(x) = Df^\top(x)$. ⟩⟩⟩⟩⟩

16 *Bemerkung* Das Symbol ∇ selbst wird *Nablaoperator* genannt und bezeichnet in der klassischen Vektoranalysis den vektoriellen Differenzialoperator

$$\nabla = \begin{pmatrix} \partial_1 \\ \vdots \\ \partial_n \end{pmatrix}.$$

Der *Gradient* von f kann damit – rein formal interpretiert – verstanden werden als das Produkt des Vektors ∇ mit dem Skalar f:

$$\nabla f = \begin{pmatrix} \partial_1 \\ \vdots \\ \partial_n \end{pmatrix} f = \begin{pmatrix} \partial_1 f \\ \vdots \\ \partial_n f \end{pmatrix}.$$

Skalarprodukte und Kreuzprodukte von ∇ mit Vektorfunktionen sind ebenfalls erklärt, sie werden uns später – im dritten Band – als *Rotation* und *Divergenz* begegnen. Das Skalarprodukt von ∇ mit sich selbst ergibt den *Laplaceoperator*

$$\nabla \bullet \nabla = \sum_{j=1}^{n} \partial_j^2 =: \Delta. \quad \multimap$$

▪ Richtung des steilsten An- und Abstiegs

Sei $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}$ um den Punkt a definiert. Für jeden Einheitsvektor $e \in \mathbb{R}^n$ ist dann

$$f_e\colon \mathbb{R} \hookrightarrow \mathbb{R}, \quad t \mapsto f(a + te)$$

in einer Umgebung von 0 definiert und stetig differenzierbar. Wird die Ableitung $f_e'(0)$ für einen Vektor e *maximal*, so nennen wir e eine *Richtung des steilsten Anstiegs*. Wird $f_e'(0)$ *minimal*, so heißt e eine *Richtung des steilsten Abstiegs*.

16 **Satz** *Sei $f\colon \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar. Ist $\nabla f(a) \neq 0$, so bezeichnen $\nabla f(a)$ und $-\nabla f(a)$ die Richtungen des steilsten An- respektive Abstiegs von f im Punkt a. Beide Richtungen sind eindeutig.* ⋊

16 ⟨⟨⟨⟨⟨ Es ist

$$f_e'(0) = f(a+te)^{\cdot}|_{t=0} = Df(a)e = \langle \nabla f(a), e\rangle .$$

Aufgrund der Cauchy-Schwarzschen Ungleichung gilt

$$|\langle \nabla f(a), e\rangle| \leqslant |\nabla f(a)|\,|e| = |\nabla f(a)| .$$

Ist $\nabla f(a) \neq 0$, so tritt Gleichheit hierbei *genau dann* ein, wenn die Vektoren $\nabla f(a)$ und e kollinear sind $_{5.32}$. Somit wird $f_e'(0)$ maximal genau für $e \upuparrows \nabla f(a)$, und minimal genau für $e \updownarrows \nabla f(a)$. In allen anderen Richtungen ist die Ableitung nicht extremal. Das ist gerade die Behauptung. ⟩⟩⟩⟩⟩

14.6 Höhere Ableitungen

Wir betrachten nun höhere partielle Ableitungen. Existiert die partielle Ableitung einer Abbildung $f\colon \mathbb{R}^n \hookrightarrow W$ nach einer Koordinate x_k auf dem ganzen Definitionsgebiet von f, so erhalten wir wieder eine Abbildung $\partial_k f\colon \mathbb{R}^n \hookrightarrow W$. Man kann also eventuell ein weiteres Mal partiell differenzieren, sagen wir nach x_l. Dann erhalten wir eine *zweite partielle Ableitung*

$$\partial_l\partial_k f \coloneqq \partial_l(\partial_k f) = (f_{x_k})_{x_l} = f_{x_k x_l} = \frac{\partial^2 f}{\partial x_k \partial x_l}\colon\ \mathbb{R}^n \hookrightarrow W,$$

wobei wir hier verschiedene Schreibweisen für dieselbe Sache aufführen. Und so weiter ...

16 **Definition** *Für eine Abbildung $f\colon \mathbb{R}^n \hookrightarrow W$ ist die r-te partielle Ableitung*

$$\partial_{k_r}..\partial_{k_2}\partial_{k_1} f = f_{x_{k_1}x_{k_2}..x_{k_r}} = \frac{\partial^r f}{\partial x_{k_1}..\partial x_{k_r}}$$

rekursiv erklärt durch

$$\partial_{k_r}..\partial_{k_2}\partial_{k_1} f \coloneqq \partial_{k_r}(\partial_{k_{r-1}}..\partial_{k_1} f),$$

sofern alle Zwischenableitungen existieren. ⋊

16 ⊳ Für

$$f:\ \mathbb{R}^3 \to \mathbb{R}, \quad f(x,y,z) = x^2y^4 \sin z$$

ist beispielsweise

$$\begin{aligned} f_x &= 2xy^4 \sin z, & f_z &= x^2y^4 \cos z, \\ f_{xy} &= 8xy^3 \sin z, & f_{zy} &= 4x^2y^3 \cos z, \\ f_{xyz} &= 8xy^3 \cos z, & f_{zyx} &= 8xy^3 \cos z. \end{aligned}$$ ⊲

Im vorangehenden Beispiel kommt es auf die Reihenfolge der partiellen Ableitungen *nicht an*. Das ist aber nicht immer so.

16 ⊳ Betrachte

$$f:\ \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} \dfrac{x^3y - xy^3}{x^2+y^2}, & (x,y) \neq 0, \\ 0, & (x,y) = 0. \end{cases}$$

Wegen $f(0,\cdot) \equiv 0$ gilt

$$f_x(0,y) = \lim_{h\to 0} f(h,y)/h = -y,$$

und wegen $f(\cdot,0) \equiv 0$ gilt

$$f_y(x,0) = \lim_{h\to 0} f(x,h)/h = x.$$

Somit ist

$$f_{xy}(0,y) = -1, \qquad f_{yx}(x,0) = 1.$$

Im Punkt $(0,0)$ stimmen diese beiden Ableitungen somit nicht überein. ⊲

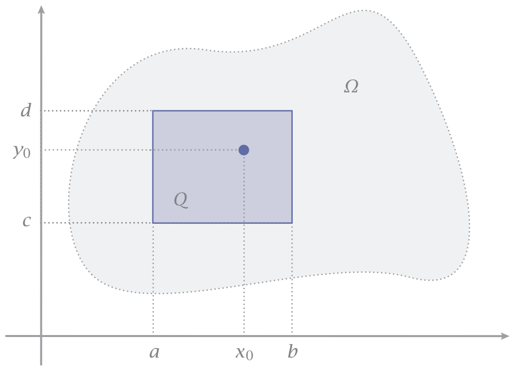
Zum Glück reicht die *Stetigkeit* der partiellen Ableitungen, um sie unabhängig von deren Reihenfolge zu machen. Die Quintessenz ist das

17 **Lemma von Schwarz** *Sei $\Omega \subset \mathbb{R}^n$ offen, $f:\ \Omega \to \mathbb{R}^m$ stetig differenzierbar, und seien x und y zwei beliebige Koordinaten in Ω. Existiert die zweite Ableitung f_{xy} auf Ω und ist sie dort stetig, so existiert auch f_{yx} auf Ω, und es gilt*

$$f_{xy} = f_{yx}.$$ ⋊

Abb 5

Zum Lemma von Schwarz



17 ⟨⟨⟨⟨⟨ Da nur die beiden Koordinaten x und y involviert sind und alle anderen fixiert werden können, beschränken wir uns auf den Fall $f = f(x,y)$. Fixiere einen Punkt $(x_0, y_0) \in \Omega$ und wähle Intervalle $[a,b]$ um x_0 und $[c,d]$ um y_0 so, dass

$$Q \coloneqq [a,b] \times [c,d] \subset \Omega.$$

Da f stetig differenzierbar in x ist, gilt

$$f(x,y) - f(a,y) = \int_a^x f_x(t,y)\,\mathrm{d}t, \qquad (x,y) \in Q.$$

Nach Voraussetzung ist f_{xy} auf Q stetig. Aufgrund des anschließend bewiesenen Lemmas 18 definiert das Integral daher eine in y differenzierbare Funktion, deren Ableitung man durch Differenzieren ›unter dem Integral‹ erhält. Da auch die linke Seite nach y differenzierbar ist, gilt also

$$f_y(x,y) - f_y(a,y) = \int_a^x f_{xy}(t,y)\,\mathrm{d}t.$$

Da der Integrand f_{xy} nach Voraussetzung stetig ist, definiert dieses Integral eine nach x differenzierbare Funktion. Somit ist auch $f_y(x,y)$ nach x differenzierbar, und es gilt

$$\partial_x f_y(x,y) = \frac{\partial}{\partial x} \int_a^x f_{xy}(t,y)\,\mathrm{d}t = f_{xy}(x,y).$$

Das war zu zeigen. ⟩⟩⟩⟩⟩

Das folgende Lemma macht eine Aussage darüber, wann ein sogenanntes *parameterabhängiges Integral* eine differenzierbare Funktion definiert, deren Ableitung man durch ›Differenziation unter dem Integral‹ erhält.

14.26

18 **Lemma** *Sei $Q = [a,b] \times [c,d]$ mit Koordinaten (x,y). Ist $f\colon Q \to \mathbb{R}$ stetig, nach y partiell differenzierbar und f_y ebenfalls stetig auf Q, so ist auch die Funktion*

$$\phi\colon \ [c,d] \to \mathbb{R}, \quad \phi(y) = \int_a^b f(t,y)\,\mathrm{d}t$$

differenzierbar, und es gilt

$$\phi' = \int_a^b f_y(t,y)\,\mathrm{d}t. \qquad ⋊$$

18 ⟨⟨⟨⟨⟨ Sei zuerst y ein innerer Punkt von $[c,d]$. Aufgrund des Hauptsatzes der Differenzial- und Integralrechnung $_{10.16}$ ist

$$\begin{aligned}\phi(y+h) - \phi(y) &= \int_a^b (f(t,y+h) - f(t,y))\,\mathrm{d}t \\ &= \int_a^b \left(\int_0^1 f_y(t,y+\theta h)h\,\mathrm{d}\theta\right)\mathrm{d}t.\end{aligned}$$

Damit folgt

$$\begin{aligned}&\frac{1}{h}(\phi(y+h) - \phi(y)) - \int_a^b f_y(t,y)\,\mathrm{d}t \\ &\quad = \int_a^b \int_0^1 f_y(t,y+\theta h)\,\mathrm{d}\theta\,\mathrm{d}t - \int_a^b \int_0^1 f_y(t,y)\,\mathrm{d}\theta\,\mathrm{d}t \\ &\quad = \int_a^b \int_0^1 (f_y(t,y+\theta h) - f_y(t,y))\,\mathrm{d}\theta\,\mathrm{d}t.\end{aligned}$$

Wegen der Kompaktheit von Q ist h_y auf Q *gleichmäßig stetig* $_{7.32}$. Zu jedem $\varepsilon > 0$ existiert also ein $\delta > 0$, so dass für *alle* $t \in [a,b]$ und $\theta \in [0,1]$

$$|f_y(t,y+\theta h) - f_y(t,y)| < \varepsilon, \qquad |h| < \delta.$$

Für diese h erhalten wir

$$\begin{aligned}&\left|\frac{1}{h}(\phi(y+h) - \phi(y)) - \int_a^b f_y(t,y)\,\mathrm{d}t\right| \\ &\quad \leqslant \int_a^b \int_0^1 |f_y(t,y+\theta h) - f_y(t,y)|\,\mathrm{d}\theta\,\mathrm{d}t \\ &\qquad \leqslant \int_a^b \int_0^1 \varepsilon\,\mathrm{d}\theta\,\mathrm{d}t = \varepsilon(b-a).\end{aligned}$$

Also gilt

$$\lim_{h\to 0} \frac{\phi(y+h) - \phi(y)}{h} = \int_a^b f_y(t,y)\,\mathrm{d}t.$$

Der Fall eines Randpunktes y wird entsprechend behandelt. ⟩⟩⟩⟩⟩

Ein entsprechender Satz gilt für die Vertauschbarkeit höherer partieller Ableitungen, wenn die entsprechenden Stetigkeitsbedingungen erfüllt sind. Auch aus diesem Grund sind Funktionen mit *stetigen* Ableitungen so wichtig.

18 **Definition** *Sei $\Omega \subset \mathbb{R}^n$ offen und $r \geqslant 1$. Dann bezeichnet $C^r(\Omega, \mathbb{R}^m)$ den Raum aller Abbildungen $f\colon \Omega \to \mathbb{R}^m$, die auf Ω sämtliche partiellen Ableitungen bis zur Ordnung r besitzen und diese dort auch stetig sind. Eine Abbildung $f \in C^r(\Omega, \mathbb{R}^m)$ heißt von der Klasse C^r oder C^r-Abbildung.* ⋊

Weiter setzt man

$$C^\infty(\Omega, \mathbb{R}^m) := \bigcap_{r \geqslant 1} C^r(\Omega, \mathbb{R}^m),$$

den Raum der unendlich oft partiell differenzierbaren Abbildungen $\Omega \to \mathbb{R}^m$. Diese werden auch *glatt* genannt. Alle diese Räume sind *lineare Vektorräume*. Im Fall skalarer Funktionen schreibt man noch kürzer

$$C^r(\Omega) := C^r(\Omega, \mathbb{R}).$$

Der folgende Satz ist dann eine leichte Übung.

18 **Satz** *Ist $f \in C^r(\Omega, \mathbb{R}^m)$, so ist jede partielle Ableitung von f der Ordnung r unabhängig von der Reihenfolge der Differenziationen.* ⋊

▪ Totale Ableitungen

Wir haben partielle Ableitungen höherer Ordnung im Standardfall definiert, jedoch nicht die entsprechenden totalen Ableitungen $D^2 f$, $D^3 f$, … *Dies werden wir hier auch nicht tun* – weil wir es nicht unmittelbar benötigen, und weil dies auch konzeptionell komplizierter ist.

Im Prinzip geht es um Folgendes. Ist $f\colon V \hookrightarrow W$ total differenzierbar, so ist

$$Df\colon\ V \hookrightarrow L(V, W).$$

Nun ist der Zielraum wieder ein Banachraum. Also ist auch die zweite totale Ableitung aufgrund unserer allgemeinen Definition erklärt – wenn sie existiert –, und es ist

$$D^2 f = D(Df)\colon\ V \hookrightarrow L(V, L(V, W)) \simeq L(V \times V, W).$$

Somit ist $D^2 f$ eine *bilineare Abbildung* von V nach W.

Und so weiter … Die r-te totale Ableitung von f kann identifiziert werden mit einer Abbildung

$$D^r f\colon\ V \hookrightarrow L(V \times \cdots \times V, W),$$

die jedem Punkt im Definitionsbereich eine r-lineare Abbildung von V nach W zuordnet. Auf die Details verzichten wir hier.