

Mathematik für Anwender II

Vorlesung 49

Extrema

Zu einer reellwertigen Funktion

$$f: G \longrightarrow \mathbb{R}$$

auf einer offenen Menge $G \subseteq \mathbb{R}^n$ interessieren wir uns, wie schon bei einem eindimensionalen Definitionsbereich, für die Extrema, also Maxima und Minima, der Funktion, und inwiefern man dies anhand der (höheren) Ableitungen (falls diese existieren) erkennen kann. Wir verallgemeinern zuerst die relevanten Definitionen auf die Situation, wo der Definitionsbereich ein beliebiger metrischer Raum ist.

DEFINITION 49.1. Sei (M, d) ein metrischer Raum und

$$f: M \longrightarrow \mathbb{R}$$

eine Funktion. Man sagt, dass f in einem Punkt $x \in M$ ein *lokales Maximum* besitzt, wenn es ein $\epsilon > 0$ gibt derart, dass für alle $x' \in M$ mit $d(x, x') < \epsilon$ die Abschätzung

$$f(x) \geq f(x')$$

gilt. Man sagt, dass f in $x \in M$ ein *lokales Minimum* besitzt, wenn es ein $\epsilon > 0$ gibt derart, dass für alle $x' \in M$ mit $d(x, x') < \epsilon$ die Abschätzung

$$f(x) \leq f(x')$$

gilt.

DEFINITION 49.2. Sei (M, d) ein metrischer Raum und

$$f: M \longrightarrow \mathbb{R}$$

eine Funktion. Man sagt, dass f in einem Punkt $x \in M$ ein *isoliertes lokales Maximum* besitzt, wenn es ein $\epsilon > 0$ gibt derart, dass für alle $x' \in M$ mit $d(x, x') \leq \epsilon$ und $x' \neq x$ die Abschätzung

$$f(x) > f(x')$$

gilt. Man sagt, dass f in $x \in M$ ein *isoliertes lokales Minimum* besitzt, wenn es ein $\epsilon > 0$ gibt derart, dass für alle $x' \in M$ mit $d(x, x') \leq \epsilon$ und $x' \neq x$ die Abschätzung

$$f(x) < f(x')$$

gilt.

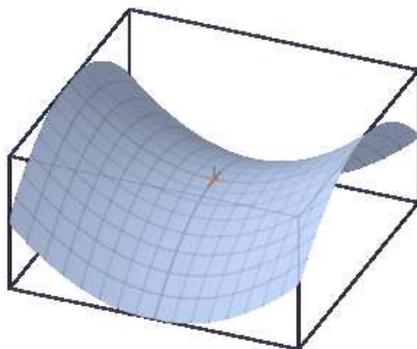
Ein *globales Maximum* liegt in $x \in M$ vor, wenn $f(x) \geq f(x')$ für alle $x' \in M$ ist.

BEISPIEL 49.3. Die Funktion

$$\mathbb{R}^2 \longrightarrow \mathbb{R}, (x, y) \longmapsto x^2 + y^2,$$

hat in $P = (0, 0)$ den Wert 0 und überall sonst positive Werte, daher liegt in P ein (isoliertes) globales Minimum vor.

Wenn die Funktion $f: M \rightarrow \mathbb{R}$ ein lokales Minimum im Punkt $P \in M$ besitzt, so gilt dies auch für die Einschränkung von f auf jede Teilmenge $N \subseteq M$, die P enthält. Beispielsweise muss ein (lokales) Minimum einer Funktion der Ebene auch auf jeder Geraden durch diesen Punkt ein (lokales) Minimum sein.



Dies heißt umgekehrt, dass wenn eine Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ auf einer Geraden L_1 durch P ein isoliertes lokales Maximum und auf einer anderen Geraden L_2 ein isoliertes lokales Minimum besitzt, dass dann kein lokales Extremum vorliegen kann. Solche Punkte nennt man *Sattelpunkt* oder *Passpunkt*, das Standardbeispiel ist das folgende.

BEISPIEL 49.4. Wir betrachten das Verhalten der Funktion

$$\mathbb{R}^2 \longrightarrow \mathbb{R}, (x, y) \longmapsto x^2 - y^2.$$

in $P = (0, 0)$. Die Einschränkung dieser Funktion auf der durch $y = 0$ gegebenen Geraden (also auf der x -Achse) ist die Funktion $x \mapsto x^2$, die in P ein (isoliertes) globales Minimum besitzt. Die Einschränkung dieser Funktion auf der durch $x = 0$ gegebenen Geraden (also auf der y -Achse) ist die Funktion $y \mapsto -y^2$, die in P ein (isoliertes) globales Maximum besitzt. Daher kann f in P kein Extremum besitzen. Auf den durch $y = x$ und $y = -x$ gegebenen Geraden ist die Funktion die Nullfunktion.

BEISPIEL 49.5. Wir betrachten im \mathbb{R}^2 die beiden Kreise K_1 und K_2 , wobei K_1 den Mittelpunkt $(0, 1)$ und Radius 1 und K_2 den Mittelpunkt $(0, 2)$ und Radius 2 habe. K_1 liegt innerhalb von K_2 , und die beiden Kreise berühren

sich in $P = (0, 0)$. Durch diese beiden Kreise wird die Ebene (neben den zwei Kreislinien selbst) in drei offene Gebiete aufgeteilt: Das Innere des Kreises $K_1 (= A)$, die große offene Kreisscheibe ohne die kleine abgeschlossene Kreisscheibe ($= B$) und das Äußere von $K_2 (= C)$. Der innere Kreis K_1 wird als Nullstelle der Funktion $f_1(x, y) = x^2 + (y - 1)^2 - 1$ beschrieben. Im Innern von K_1 ist diese Funktion negativ, auf K_1 hat sie den Wert 0 und außerhalb davon hat sie positive Werte. Entsprechendes gilt für K_2 und die Funktion $f_2(x, y) = x^2 + (y - 2)^2 - 4$. Wir setzen

$$\begin{aligned} f(x, y) &:= f_1(x, y) \cdot f_2(x, y) \\ &= (x^2 + (y - 1)^2 - 1) \cdot (x^2 + (y - 2)^2 - 4) \\ &= (x^2 + y^2 - 2y) \cdot (x^2 + y^2 - 4y) \\ &= x^4 + y^4 + 2x^2y^2 - 6y^3 - 6x^2y + 8y^2. \end{aligned}$$

Diese Funktion nimmt auf den beiden Kreisen den Wert 0 an, sie ist auf A positiv, auf B negativ und auf C wieder positiv.

Die Funktion f besitzt in P kein lokales Minimum, da sie dort den Wert 0 besitzt und da jede beliebig kleine Umgebung $U(P, \epsilon)$ den Bereich B trifft, wo f negative Werte besitzt. Die Einschränkung der Funktion auf jede Gerade durch den Nullpunkt besitzt aber dort ein lokales Minimum. Sei dazu G eine solche Gerade. Wenn G die x -Achse ist, so verläuft diese Gerade (bis auf P selbst) in C , wo f nur positive Werte annimmt, sodass in P ein (sogar globales) Minimum vorliegt. Sei also G eine von der x -Achse verschiedene Gerade durch P . Die eine Hälfte der Geraden verläuft ganz in C , wo die Funktion positiv ist. Die andere Hälfte verläuft, ausgehend von P , zuerst in A , dann in B und schließlich in C . Da die Funktion auf A positiv ist, kann man ein Teilintervall $[-\delta, \delta]$ der Gerade wählen derart, dass dieses Teilstück (abgesehen von P) nur in A und C verläuft. Auf diesem Teilintervall nimmt die Funktion in P den Wert 0 und sonst überall positive Werte annimmt. Daher besitzt die eingeschränkte Funktion ein lokales Minimum. Das dabei zu wählende δ hängt natürlich wesentlich von der Steigung der Geraden ab, es gibt kein gemeinsames δ für alle Geraden.

Der Gradient

Wenn eine Funktion $f: V \rightarrow \mathbb{R}$ total differenzierbar ist, so ist das totale Differential in einem Punkt eine lineare Abbildung von V nach \mathbb{R} . Für solche linearen Abbildungen gibt es einen eigenen Namen.

DEFINITION 49.6. Sei K ein Körper und sei V ein K -Vektorraum. Eine lineare Abbildung

$$V \longrightarrow K$$

heißt auch eine *Linearform* auf V .

Wenn $G \subseteq \mathbb{R}^n$ ist, so bilden die partiellen Ableitungen in einem Punkt $P \in G$ eine Matrix mit einer einzigen Zeile, die bei stetigen partiellen Ableitungen das totale Differential repräsentiert. Eine solche Matrix kann man aber ebenso auch als ein n -Tupel in \mathbb{R} und damit als einen Vektor in \mathbb{R}^n auffassen.

LEMMA 49.7. *Es sei $(V, \langle -, - \rangle)$ ein euklidischer Vektorraum und*

$$f: V \longrightarrow \mathbb{R}$$

eine Linearform. Dann gibt es einen eindeutig bestimmten Vektor $w \in V$ mit

$$f(v) = \langle w, v \rangle.$$

Wenn u_1, \dots, u_n eine Orthonormalbasis von V und $f(u_i) = a_i$ ist, so ist dieser Vektor gleich $w = \sum_{i=1}^n a_i u_i$.

Beweis. Die Aussage folgt aus dem Zusatz. Sei also eine Orthonormalbasis u_1, \dots, u_n gegeben und sei $w = \sum_{i=1}^n a_i u_i$. Dann ist für jedes j

$$\langle w, u_j \rangle = \left\langle \sum_{i=1}^n a_i u_i, u_j \right\rangle = a_j = f(u_j).$$

D.h. die beiden linearen Abbildungen $v \mapsto \langle w, v \rangle$ und f stimmen auf einer Basis überein, sind also nach Satz 9.5 identisch. Für jeden anderen Vektor $w' = \sum_{i=1}^n b_i u_i$ ist der Wert der zugehörigen Linearform an mindestens einem Basisvektor u_j von $f(u_j)$ verschieden, daher liegt Eindeutigkeit vor. \square

DEFINITION 49.8. Sei $(V, \langle -, - \rangle)$ ein euklidischer Vektorraum, $G \subseteq V$ offen und

$$f: G \longrightarrow \mathbb{R}$$

eine in $P \in G$ differenzierbare Funktion. Dann nennt man den eindeutig bestimmten Vektor $w \in V$ mit

$$(Df)_P(v) = \langle w, v \rangle$$

für alle $v \in V$ den *Gradienten* von f in P . Er wird mit

$$\text{grad } f(P)$$

bezeichnet.

Man beachte, dass wir durchgehend die endlichdimensionalen Vektorräume mit einem Skalarprodukt versehen, um topologische Grundbegriffe wie Konvergenz und Stetigkeit zur Verfügung zu haben, dass diese Begriffe aber nicht von dem gewählten Skalarprodukt abhängen. Dem entgegen hängt aber der Gradient von dem gewählten Skalarprodukt ab.

Bei $V = \mathbb{R}^n$, versehen mit dem Standardskalarprodukt, ist der Gradient einfach gleich

$$\text{grad } f(P) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(P) \\ \vdots \\ \frac{\partial f}{\partial x_n}(P) \end{pmatrix}.$$

BEMERKUNG 49.9. Zu einer differenzierbaren Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ lässt sich der Gradient (bezüglich des Standardskalarproduktes) einfach durch partielles Differenzieren berechnen. Es wäre aber eine künstliche Einschränkung, nur diese Situation zu betrachten. Um dies zu illustrieren sei beispielsweise

$$f: \mathbb{R}^3 \longrightarrow \mathbb{R}$$

eine differenzierbare Funktion und $E \subset \mathbb{R}^3$ eine Ebene, die etwa als Lösungsmenge der linearen Gleichung $5x - 4y + 9z = 0$ gegeben sei. Dann induziert das Standardskalarprodukt des \mathbb{R}^3 durch Einschränkung ein Skalarprodukt auf E . Diese Ebene ist zwar isomorph zu \mathbb{R}^2 , es ergibt aber keinen Sinn, das eingeschränkte Skalarprodukt als Standardskalarprodukt anzusprechen. Der Gradient G zu f in einem Punkt $P \in \mathbb{R}^3$ lässt sich direkt mit den partiellen Ableitungen zu den drei Raumkoordinaten berechnen. Bei $P \in E$ wird im Allgemeinen der Gradient *nicht* auf E liegen. Die eingeschränkte Funktion

$$f|_E: E \longrightarrow \mathbb{R}$$

ist aber ebenfalls differenzierbar und besitzt daher einen Gradienten \tilde{G} , der auf E liegt, und dieser lässt sich nicht über partielle Ableitungen berechnen, da es auf E keine Standardbasis gibt. Übrigens ist \tilde{G} die orthogonale Projektion von G auf E .

SATZ 49.10. Sei $(V, \langle -, - \rangle)$ ein euklidischer Vektorraum, sei $G \subseteq V$ offen und sei

$$f: G \longrightarrow \mathbb{R}$$

eine in $P \in G$ differenzierbare Funktion. Dann gelten folgende Aussagen.

- (1) Für jeden Vektor $v \in V$ ist

$$|(Df)_P(v)| \leq \|v\| \cdot \|\text{grad } f(P)\|.$$

- (2) Dabei gilt Gleichheit genau dann, wenn v linear abhängig zum Gradienten ist.
 (3) Sei $\text{grad } f(P) \neq 0$. Unter allen Vektoren $v \in V$ mit $\|v\| = 1$ ist die Richtungsableitung in Richtung des normierten Gradienten maximal, und zwar gleich der Norm des Gradienten.

Beweis. (1) folgt wegen

$$(Df)_P(v) = \langle v, \text{grad } f(P) \rangle$$

direkt aus der Abschätzung von Cauchy-Schwarz. (2) ergibt sich aus den Zusätzen zur Abschätzung von Cauchy-Schwarz, siehe Aufgabe ***** (3). Aus (1) und (2) folgt, dass

$$\begin{aligned} \left| \left\langle \text{grad } f(P), \pm \frac{\text{grad } f(P)}{\|\text{grad } f(P)\|} \right\rangle \right| &= \left| (Df)_P \left(\pm \frac{\text{grad } f(P)}{\|\text{grad } f(P)\|} \right) \right| \\ &= \|\text{grad } f(P)\| \end{aligned}$$

gilt, und dass diese beiden Vektoren die einzigen Vektoren der Norm 1 sind, für die diese Gleichung gilt. Wenn man links die Betragstriche weglässt, so

gilt die Gleichheit für $\frac{\text{grad } f(P)}{|\text{grad } f(P)|}$ nach wie vor, da das Skalarprodukt positiv definit ist. \square

Der Gradient gibt demnach die Richtung an, in die die Funktion den stärksten Anstieg hat. In die entgegengesetzte Richtung liegt entsprechend der steilste Abstieg vor.

Lokale Extrema von Funktionen in mehreren Variablen

Wir wollen mit den Mitteln der Differentialrechnung Kriterien erarbeiten, in welchen Punkten eine Funktion

$$f: G \longrightarrow \mathbb{R}$$

ein lokales Minimum oder ein lokales Maximum annimmt. Wenn man sich den Graph einer solchen Funktion als ein Gebirge über der Grundmenge G vorstellt, so geht es also um die Gipfel und die Senken des Gebirges. Der folgende Satz liefert ein notwendiges Kriterium für die Existenz eines lokalen Extremums, das das entsprechende Kriterium in einer Variablen verallgemeinert.

SATZ 49.11. *Sei V ein endlichdimensionaler reeller Vektorraum und $G \subseteq V$ eine offene Teilmenge. Es sei*

$$f: G \longrightarrow \mathbb{R}$$

eine Funktion, die im Punkt $P \in G$ ein lokales Extremum besitzt. Dann gelten folgende Aussagen.

- (1) *Wenn f in P in Richtung $v \in V$ differenzierbar ist, so ist*

$$(D_v f)(P) = 0.$$

- (2) *Wenn f in P total differenzierbar ist, so verschwindet das totale Differential, also*

$$(Df)_P = 0.$$

Beweis. (1) Zu $v \in V$ betrachten wir die Funktion

$$h: I \longrightarrow \mathbb{R}, t \longmapsto h(t) = f(P + tv),$$

wobei I ein geeignetes reelles Intervall ist. Da die Funktion f in P ein lokales Extremum besitzt, besitzt die Funktion h in $t = 0$ ein lokales Extremum. Nach Voraussetzung ist h differenzierbar und nach Fakt ***** ist $h'(0) = 0$. Diese Ableitung stimmt aber mit der Richtungsableitung überein, also ist

$$(D_v f)(P) = h'(0) = 0.$$

- (2) folgt aus (1) aufgrund von Fakt *****. \square

Ein lokales Extremum kann also nur in einem sogenannten kritischen Punkt einer Funktion auftreten.

DEFINITION 49.12. Sei V ein endlichdimensionaler reeller Vektorraum, $G \subseteq V$ offen und

$$f: G \longrightarrow \mathbb{R}$$

eine differenzierbare Funktion. Dann heißt $P \in G$ ein *kritischer Punkt* von f (oder ein *stationärer Punkt*), wenn

$$(Df)_P = 0$$

ist. Anderfalls spricht man von einem *regulären Punkt*.

Die Hesse-Form

Wir sind natürlich auch an hinreichenden Kriterien für das Vorliegen von lokalen Extrema interessiert. Wie schon im eindimensionalen Fall muss man sich die zweiten Ableitungen anschauen, wobei die Situation natürlich dadurch wesentlich verkompliziert wird, dass es zu je zwei Richtungsvektoren v und w eine zweite Richtungsableitung $D_{vw} = D_v D_w$ gibt. Die zweite Richtungsableitung wird dadurch handhabbar, dass man sie in die sogenannte Hesse-Form bzw. Hesse-Matrix zusammenfasst.

DEFINITION 49.13. Es sei V ein endlichdimensionaler reeller Vektorraum, $G \subseteq V$ eine offene Menge und

$$f: G \longrightarrow \mathbb{R}$$

eine zweimal stetig differenzierbare Funktion. Zu $P \in G$ heißt die Abbildung

$$\text{Hess}_P f: V \times V \longrightarrow \mathbb{R}, (u, v) \longmapsto D_u D_v f(P),$$

die *Hesse-Form* im Punkt $P \in G$.

DEFINITION 49.14. Es sei V ein endlichdimensionaler reeller Vektorraum, $G \subseteq V$ eine offene Menge und

$$f: G \longrightarrow \mathbb{R}$$

eine zweimal stetig differenzierbare Funktion. Es sei eine Basis $v_i, i = 1, \dots, n$, von V gegeben mit den zugehörigen Richtungsableitungen $D_i := D_{v_i}, i = 1, \dots, n$. Zu $P \in G$ heißt dann die Matrix

$$\begin{pmatrix} D_1 D_1 f(P) & \cdots & D_1 D_n f(P) \\ \vdots & \ddots & \vdots \\ D_n D_1 f(P) & \cdots & D_n D_n f(P) \end{pmatrix}$$

die *Hesse-Matrix* zu f im Punkt P bezüglich der gegebenen Basis.

Die Hesse-Form zu einem festen Punkt P ordnet also zwei Vektoren eine reelle Zahl zu, und sie ist durch ihre Hesse-Matrix vollständig beschrieben. Damit ordnet sie sich in das Konzept von symmetrischen Bilinearformen ein.

Abbildungsverzeichnis

Quelle = Saddle point.png , Autor = Nashev, Lizenz =

2