2.2 Lokale Extrema

Wir betrachten jetzt skalare Funktionen $f : V \hookrightarrow \mathbb{R}$. Den Graphen einer solchen Funktion kann man sich als Höhenprofil über ihrem Definitionsbereich vorstellen – zumindest wenn $\dim V = 2$. Ist f differenzierbar, so besitzt dieses Profil in jedem Punkt eine *Tangentialebene.*

Definition *Ist $f : V \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so heißt der Graph der affinen Funktion*

$$T : V \to \mathbb{R}, \quad z = f(a) + Df(a)(x-a)$$

die Tangentialebene an den Graphen von f im Punkt $(a, f(a))$.

A. Die Tangentialebene des linearen Funktionals $x \mapsto \langle l, x \rangle$ im Punkt a gegeben durch

$$z = \langle l, a \rangle + \langle l, x-a \rangle = \langle l, x \rangle,$$

ist also identisch mit dem Graphen der Funktion.

B. Die quadratische Form $x \mapsto \langle Ax, x \rangle / 2$ hat in x_0 die Tangentialebene

$$z = \frac{1}{2} \langle Ax_0, x_0 \rangle + \langle Ax_0, x - x_0 \rangle = \langle Ax_0, x - x_0/2 \rangle.$$

Im Standardfall ist die Tangentialebene eine Hyperebene im Raum $\mathbb{R}^n \times \mathbb{R}$, die sich durch den Gradienten wie folgt beschreiben lässt.

Satz *Ist $f : \mathbb{R}^n \hookrightarrow \mathbb{R}$ im Punkt a differenzierbar, so ist*

$$N(a) := (-\nabla f(a), 1) \in \mathbb{R}^n \times \mathbb{R}$$

der Normalenvektor der Tangentialebene an den Graphen von f über a.

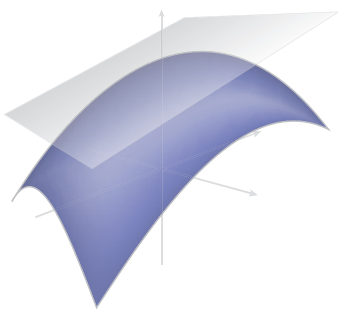


Abb 1

Eine Tangentialebene

⟨⟨⟨⟨⟨ Ausgedrückt mithilfe des Gradienten lautet die Gleichung der Tangentialebene $z = f(a) + \langle \nabla f(a), x - a \rangle$, oder

$$\langle -\nabla f(a), x - a \rangle + 1 \cdot (z - f(a)) = 0.$$

Dies entspricht der *Normalengleichung* dieser Ebene im Raum $\mathbb{R}^n \times \mathbb{R}$ mit Koordinaten (x, z) und dem Standardskalarprodukt $\langle \cdot, \cdot \rangle_{n+1}$ und dem Normalenvektor

$$N(a) = (-\nabla f(a), 1).$$ ⟩⟩⟩⟩⟩

▪ Extremalstellen

Für skalare Funktionen ist es sinnvoll, nach der Existenz lokaler Extrema zu fragen. Diese sind genau wie für Funktionen einer Variablen erklärt.

Definition *Eine Funktion $f : V \hookrightarrow \mathbb{R}$ besitzt im Punkt a ein lokales Minimum, wenn es eine Umgebung U von a im Definitionsbereich von f gibt, so dass*

$$f(a) \leqslant f(x), \qquad x \in U.$$

Das lokale Minimum heißt strikt, wenn sogar

$$f(a) < f(x), \qquad x \in U \setminus \{a\}.$$

Der Punkt a selbst heißt eine Minimalstelle von f. Entsprechend sind lokales Maximum und Maximalstelle erklärt. ⋊

Minima und Maxima werden gemeinsam als *Extrema* bezeichnet, und Minimal- und Maximalstellen gemeinsam als *Extremalstellen*. Für diese gilt der Satz von Fermat 8.8 entsprechend auch hier.

7 **Satz von Fermat** *Besitzt die Funktion $f : V \hookrightarrow \mathbb{R}$ im Punkt a eine Extremalstelle und ist sie dort total differenzierbar, so ist $Df(a) = 0$.* ⋊

⟨⟨⟨⟨⟨ Der Definitionsbereich von $f : V \hookrightarrow \mathbb{R}$ ist vereinbarungsgemäß offen. Somit ist f auf einer offenen Umgebung von a erklärt. Für jeden Vektor $v \neq 0$ ist dann $\varphi : t \mapsto f(a + tv)$ auf einem offenen Intervall um 0 erklärt und besitzt in $t = 0$ ein lokales Extremum. Da φ dort auch differenzierbar ist, gilt nach dem klassischen Satz von Fermat 8.8 $\dot{\varphi}(0) = 0$, also

$$0 = f(a + tv)^{\cdot}\Big|_{t=0} = Df(a)v.$$

Da dies für jeden Vektor v gilt, ist $Df(a) = 0$. ⟩⟩⟩⟩⟩

▹ Für ein beliebiges Skalarprodukt $\langle \cdot, \cdot \rangle$ besitzt die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = \frac{1}{2} \langle x, x \rangle$$

aufgrund der Definitheit des Skalarprodukts bei $x = 0$ ein striktes Minimum. Und in der Tat verschwindet genau dort auch $Df(x) = \langle x, \cdot \rangle$.

Definition *Ist $f : V \hookrightarrow \mathbb{R}$ im Punkt c total differenzierbar und $Df(c) = 0$, so heißt c ein kritischer oder stationärer Punkt von f.*

Der Satz von Fermat besagt also auch für skalare Funktionen in höheren Dimensionen, dass eine Extremalstelle im Innern *notwendig* ein kritischer Punkt ist, wenn die Funktion dort total differenzierbar ist. Die Tangentialebene an dieser Stelle ist dann ›horizontal‹. Dies gilt allerdings nicht mehr, wenn die Extremstelle am Rand des Definitionsbereichs liegt. Ebensowenig ist ein kritischer Punkt notwendigerweise eine Extremstelle.

Definite Matrizen

In der eindimensionalen Theorie haben wir *hinreichende* Bedingungen für die Existenz eines lokalen Extremums mithilfe der zweiten Ableitung formuliert 8.14. In höheren Dimensionen wird die zweite Ableitung aber nicht mehr durch eine reelle Zahl, sondern im Standardfall – den wir von nun an betrachten – durch die Hessematrix dargestellt. Es geht also darum, geeignete hinreichende Bedingungen bezüglich solcher Operatoren zu formulieren.

Sei $S(n)$ der Raum aller reellen symmetrischen $n \times n$-Matrizen

$$A = (A_{kl})_{1 \leqslant k,l \leqslant n}, \qquad A^\top = A.$$

Dies ist ein reeller Vektorraum der Dimension $n(n+1)/2$.

Definition *Eine Matrix $A \in S(n)$ heißt*

(i) *positiv definit, geschrieben $A > 0$, falls*

$$\langle Av, v \rangle > 0, \qquad 0 \neq v \in \mathbb{R}^n,$$

(ii) *positiv semidefinit, geschrieben $A \geqslant 0$, falls*

$$\langle Av, v \rangle \geqslant 0, \qquad v \in \mathbb{R}^n,$$

(iii) *negativ definit, geschrieben $A < 0$, falls $-A > 0$,*
(iv) *negativ semidefinit, geschrieben $A \leqslant 0$, falls $-A \geqslant 0$,*
(v) *indefinit, geschrieben $A \gtrless 0$, im verbleibenden Fall.*

Eine Matrix $A \in S(n)$ ist also indefinit, wenn $\langle Av, v \rangle$ das Vorzeichen wechselt. Alle diese Fälle treten bereits bei Diagonalmatrizen auf. — Im definiten Fall gilt noch eine stärkere Aussage.

Lemma *Für eine Matrix $A \in S(n)$ sind äquivalent:*

2.10

(i) *Es ist* $A > 0$.
(ii) *Es gibt ein* $\mu > 0$, *so dass* $\langle Av, v\rangle \geqslant \mu\,|v|^2$ *für alle* $v \in \mathbb{R}^n$.
(iii) *Es gibt ein* $\mu > 0$, *so dass* $A - \mu E_n \geqslant 0$. ⋊

⟨⟨⟨⟨⟨ (i) ⇒ (ii) Die quadratische Form

$$Q : \mathbb{R}^n \to \mathbb{R}, \quad Q(v) = \langle Av, v\rangle$$

ist stetig und nach Voraussetzung auf der Einheitssphäre $\mathbb{S}^{n-1}$ positiv. Wegen der Kompaktheit von $\mathbb{S}^{n-1}$ nimmt sie ihr Minimum an, so dass

$$\mu \coloneqq \inf_{u \in \mathbb{S}^{n-1}} Q(u) > 0.$$

Für ein beliebiges $0 \neq v \in \mathbb{R}^n$ gilt dann mit $u = v/\,|v|$ die Abschätzung

$$\langle Av, v\rangle = |v|^2\, Q(u) \geqslant \mu\,|v|^2.$$

Dies bleibt auch für $v = 0$ gültig, womit (ii) gezeigt ist.

(ii) ⇒ (iii) Aus der Voraussetzung folgt

$$\langle (A - \mu E_n)v, v\rangle = \langle Av, v\rangle - \mu\,|v|^2 \geqslant 0, \qquad v \in \mathbb{R}^n.$$

Somit ist $A - \mu E_n \geqslant 0$.

(iii) ⇒ (i) Mit der letzten Ungleichung gilt

$$\langle Av, v\rangle \geqslant \mu\,|v|^2 > 0, \qquad v \neq 0,$$

also $A > 0$. ⟩⟩⟩⟩⟩

Die Definitheitseigenschaften einer symmetrischen Matrix lassen sich direkt aus ihrem Spektrum ablesen, das ja reell ist.

8 **Lemma** *Sind* $\lambda_1 \leqslant \lambda_2 \leqslant .. \leqslant \lambda_n$ *die Eigenwerte von* $A \in S(n)$, *so gilt:*
(i) $A > 0 \;\Leftrightarrow\; \lambda_1 > 0$,
(ii) $A \geqslant 0 \;\Leftrightarrow\; \lambda_1 \geqslant 0$,
(iii) $A < 0 \;\Leftrightarrow\; \lambda_n < 0$,
(iv) $A \leqslant 0 \;\Leftrightarrow\; \lambda_n \leqslant 0$,
(v) $A \gtrless 0 \;\Leftrightarrow\; \lambda_1\lambda_n < 0$. ⋊

⟨⟨⟨⟨⟨ Eine symmetrische Matrix A besitzt ein Orthonormalsystem von Eigenvektoren $w_1, .., w_n$ zu den Eigenwerten $\lambda_1, .., \lambda_n$. Es ist also

$$Aw_k = \lambda_k w_k, \qquad \langle w_k, w_l\rangle = \delta_{kl}.$$

Für $v = v_1 w_1 + .. + v_n w_n$ ist dann $Av = \lambda_1 v_1 w_1 + .. + \lambda_n v_n w_n$ und

$$\langle Av, v\rangle = \sum_{i,j=1}^{n} \lambda_i v_i v_j \,\langle w_i, w_j\rangle = \sum_{i=1}^{n} \lambda_i v_i^2.$$

Daraus ergeben sich alle Behauptungen. Zum Beispiel ist für $v \neq 0$

$$\langle Av, v\rangle \geqslant \lambda_1 \sum_{i=1}^{n} v_i^2 = \lambda_1 \left|v\right|^2 > 0 \quad \Leftrightarrow \quad \lambda_1 > 0. \quad ⟩⟩⟩⟩⟩$$

Speziell für symmetrische 2×2-Matrizen ergibt sich hieraus folgendes

9 **Korollar** *Für eine reelle symmetrische Matrix* $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ *gilt:*

(i) $A > 0 \quad \Leftrightarrow \quad \det A > 0 \wedge a > 0,$

(ii) $A < 0 \quad \Leftrightarrow \quad \det A > 0 \wedge a < 0,$

(iii) $A \gtrless 0 \quad \Leftrightarrow \quad \det A < 0.$ ⋊

⟨⟨⟨⟨⟨ Bezeichnen λ_1 und λ_2 die beiden Eigenwerte von A, so gilt bekanntlich $\det A = \lambda_1\lambda_2$. Ist $\det A < 0$, so haben beide Eigenwerte entgegengesetztes Vorzeichen, und A ist indefinit $_8$. Ist $\det A > 0$, so ist A definit, und $a = \langle Ae_1, e_1\rangle$ entscheidet über das Vorzeichen. ⟩⟩⟩⟩⟩

Nun benötigen wir noch eine Stetigkeitsaussage für Matrizenfunktionen.

10 **Lemma** *Die Matrixfunktion* $A\colon \mathbb{R}^n \hookrightarrow S(n)$ *sei stetig und* $A(c) > 0$. *Dann existiert eine Umgebung* U *von* c, *so dass* $A(x) > 0$ *für alle* $x \in U$. ⋊

⟨⟨⟨⟨⟨ Es existiert ein $\mu > 0$ $_8$, so dass

$$\langle A(c)v, v\rangle \geqslant 2\mu \left|v\right|^2, \qquad v \in \mathbb{R}^n.$$

Aufgrund der Stetigkeit von A existiert dazu eine Umgebung U von c, so dass

$$\|A(x) - A(c)\| \leqslant \mu, \qquad x \in U.$$

bezüglich der induzierten Operatornorm $\|\cdot\|$. Mit der Schwarzschen Ungleichung folgt

$$\begin{aligned} &\left|\langle A(x)v, v\rangle - \langle A(c)v, v\rangle\right| \\ &\quad \leqslant \left|(A(x) - A(c))v\right| \left|v\right| \leqslant \|A(x) - A(c)\| \left|v\right|^2 \leqslant \mu \left|v\right|^2. \end{aligned}$$

Also ist

$$\left|\langle A(x)v, v\rangle\right| \geqslant \left|\langle A(c)v, v\rangle\right| - \mu \left|v\right|^2 \geqslant 2\mu \left|v\right|^2 - \mu \left|v\right|^2 = \mu \left|v\right|^2.$$

Somit gilt $_8$ $A(x) > 0$ für $x \in U$. ⟩⟩⟩⟩⟩

Hinweis Das Lemma wird *falsch* mit $\geqslant$ anstelle von $>$. ⊸

▪ Lokale Extremstellen

Nun zurück zum eigentlichen Problem, der Charakterisierung von Extremalstellen. Wir formulieren die Ergebnisse für *Minimalstellen*, für Maximalstellen sind die Aussagen entsprechend abzuwandeln.

11 **Satz** *Ist c eine lokale Minimalstelle einer C^2-Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$, so ist*

$$Hf(c) \geqslant 0. \quad ⋊$$

⟨⟨⟨⟨⟨ Für jedes $v \in \mathbb{R}^n$ besitzt die Hilfsfunktion $\varphi\colon t \mapsto f(c+tv)$ in $t=0$ ein lokales Minimum. Also gilt

$$\varphi''(0) = \partial_v^2 f(c) = \langle Hf(c)v, v\rangle \geqslant 0.$$

Da dies für alle $v \in \mathbb{R}^n$ gilt, folgt die Behauptung. ⟩⟩⟩⟩⟩

Dieser Satz formuliert eine notwendige, aber nicht hinreichende Bedingung, wie bereits die kubische Parabel $t \mapsto t^3$ zeigt. Um die Existenz einer Minimalstelle zu *garantieren*, braucht es etwas mehr.

12 **Satz** *Sei c ein kritischer Punkt einer C^2-Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$. Gilt*

$$Hf(x) \geqslant 0$$

in einer Umgebung U von c, so ist c eine lokale Minimalstelle. Gilt sogar

$$Hf(c) > 0,$$

so ist c eine strikte lokale Minimalstelle. ⋊

⟨⟨⟨⟨⟨ Auf einer kleinen kugelförmigen Umgebung U von c gilt [5]

$$f(c+h) = f(c) + \frac{1}{2}\langle Hf(\xi)h, h\rangle$$

mit einem $\xi \in [c, c+h]$. Ist U hinreichend klein, so gilt $\langle Hf(\xi)h, h\rangle \geqslant 0$ nach Voraussetzung und somit

$$f(x) \geqslant f(c), \qquad x \in U.$$

Also ist c eine Minimalstelle.

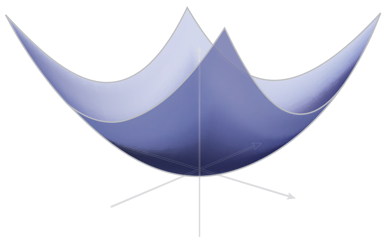
Gilt $Hf(c) > 0$, so ist sogar $\langle Hf(\xi)h, h\rangle > 0$ für alle $\xi \in U$ und $h \neq 0$, wenn U hinreichend klein ist [10]. Also folgt mit demselben Argument

$$f(x) > f(c), \qquad x \in U \setminus \{c\},$$

und c ist eine strikte Minimalstelle. ⟩⟩⟩⟩⟩

Abb 2

Das Paraboloid $x^2 + y^2$



■ **Typische Fälle**

▷ *Der definite Fall* Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

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

Es ist

$$\nabla f = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \qquad Hf = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Also ist 0 der einzige kritische Punkt von f. Da außerdem $Hf > 0$ auf ganz $\mathbb{R}^2$, ist 0 eine lokale strikte Minimalstelle. Der Graph von f ist ein nach oben geöffnetes *Paraboloid* Abb 2 .

Für die Funktion $-f$ ist 0 entsprechend eine lokale strikte Maximalstelle, und ihr Graph ein nach unten geöffnetes Paraboloid. ◁

Ist die Hessische in einem kritischen Punkt semidefinit, aber nicht definit, so ist mindestens einer ihrer Eigenwerte Null. In diesem Fall kann der Graph von f sehr unterschiedliche Gestalt annehmen.

▷ *Der semidefinite Fall* Betrachte die Funktionen $f, g, h : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x, y) = x^2 + y^4, \qquad g(x, y) = x^2, \qquad h(x, y) = x^2 + y^3.$$

In allen Fällen ist 0 der einzige kritische Punkt mit Hessematrix

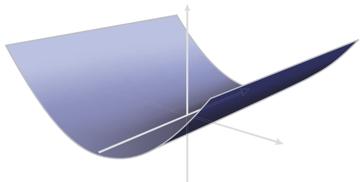
$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \geqslant 0.$$

Die Funktionen verhalten sich um 0 jedoch sehr unterschiedlich:
(i) f hat ein striktes Minimum, sehr ähnlich zu Abbildung 2,
(ii) g hat ein nichtisoliertes Minimum wie in Abbildung 3,
(iii) h hat kein Minimum, sondern bildet einen *Affensattel* wie in Abbildung 4. ◁

Der semidefinite Fall ist in vielen Fällen schwierig zu behandeln. Oft kann man ihn aber als eine ›nicht typische‹ oder ›entartete‹ Situation betrachten, die

Abb 3

Die Rinne x^2



›normalerweise‹ nicht auftritt. Betrachtet man nur die ›nichtentarteten‹ Fälle, so wird die Situation meist sehr viel übersichtlicher.

Definition *Ein kritischer Punkt c einer C^2-Funktion $f: \mathbb{R}^n \hookrightarrow \mathbb{R}$ heißt nichtdegeneriert oder nichtentartet, falls*

$$\det Hf(c) \neq 0.$$

Andernfalls heißt er entartet oder degeneriert. ⋊

In einem nichtentarteten kritischen Punkt kann die Hessische nicht semidefinit sein, da 0 kein Eigenwert ist. Sie ist entweder definit oder *indefinit*. Dieser Fall hat ebenfalls einen eigenen Namen.

Definition *Ein nichtentarteter kritischer Punkt c einer C^2-Funktion f heißt Sattelpunkt, falls die Hessische $Hf(c)$ indefinit ist.* ⋊

Im Falle einer Funktion *zweier* Variablen ist dies äquivalent mit der Bedingung $\det Hf(c) < 0$. In höheren Dimensionen ist diese Bedingung jedoch nur hinreichend, aber nicht notwendig, denn die Anzahl positiver wie negativer Eigenwerte kann positiv und gerade und die Determinante damit positiv sein.

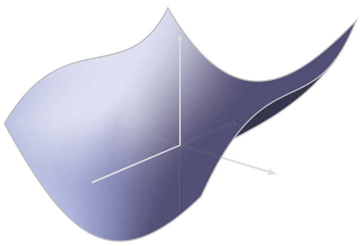
▹ *Sattelpunkt im $\mathbb{R}^2$* Betrachte die Funktion $f: \mathbb{R}^2 \to \mathbb{R}$ mit

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

Es ist

Abb 4

Der Affensattel $x^2 + y^3$



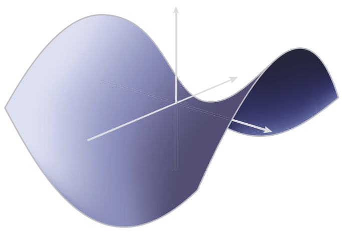


Abb 5

Der Sattel $x^2 - y^2$

$$\nabla f = \begin{pmatrix} 2x \\ -2y \end{pmatrix}, \qquad Hf = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Hier ist 0 wieder der einzige kritische Punkt von f, und es ist $Hf \gtrless 0$. Es liegt somit ein Sattelpunkt vor Abb 5. ◂