1.4 Extremwerte

Definition

Sei $M \subset \mathbb{R}^n$ eine Teilmenge, $f : M \to \mathbb{R}$ stetig, $\mathbf{a} \in M$ ein Punkt.

f hat in $\mathbf{a}$ auf M ein relatives (oder lokales) ***Maximum*** bzw. ein relatives (oder lokales) ***Minimum***, wenn es eine offene Umgebung $U(\mathbf{a}) \subset \mathbb{R}^n$ gibt, so dass

$$f(\mathbf{x}) \leq f(\mathbf{a}) \quad (\text{bzw.} \quad f(\mathbf{x}) \geq f(\mathbf{a})\ )$$

für alle $\mathbf{x} \in U \cap M$ ist. In beiden Fällen spricht man von einem relativen (oder lokalen) ***Extremum***.

Gilt die Ungleichung sogar für alle $\mathbf{x} \in M$, so spricht man von einem ***absoluten*** (oder ***globalen***) Maximum oder Minimum.

4.1. Notwendiges Kriterium für relative Extremwerte

Sei $B \subset \mathbb{R}^n$ offen und $f : B \to \mathbb{R}$ in $\mathbf{a} \in B$ differenzierbar.

Besitzt f in $\mathbf{a}$ ein relatives Extremum, so ist $\nabla f(\mathbf{a}) = \mathbf{0}$.

Beweis: Für $i = 1, \ldots, n$ besitzt auch $g_i(t) := f(\mathbf{a} + t\mathbf{e}_i)$ in $t = 0$ ein lokales Extremum. Nach dem notwendigen Kriterium aus der Differentialrechnung einer Veränderlichen muss dann $(g_i)'(0) = 0$ sein. Es ist aber

$$(g_i)'(0) = \frac{\partial f}{\partial x_i}(\mathbf{a}), \text{ für } i = 1, \ldots, n.$$

Daraus folgt die Behauptung. ∎

Definition

Ist f in $\mathbf{a}$ differenzierbar und $\nabla f(\mathbf{a}) = \mathbf{0}$, so heißt $\mathbf{a}$ ein ***stationärer*** (oder ***kritischer***) Punkt von f.

Ein stationärer Punkt $\mathbf{a}$ von f heißt ***Sattelpunkt*** von f, falls es in jeder Umgebung von $\mathbf{a}$ Punkte $\mathbf{b}$ und $\mathbf{c}$ gibt, so dass $f(\mathbf{b}) < f(\mathbf{a}) < f(\mathbf{c})$ ist.

Beispiele dazu werden wir später betrachten. Ein hinreichendes Kriterium für die Existenz eines Extremwertes erhält man in einer Veränderlichen durch Untersuchung der höheren Ableitungen, insbesondere der zweiten Ableitung.

Wir kommen nun nicht umhin, die Taylorformel in n Veränderlichen zu beweisen. Für die Untersuchung von Extremwerten brauchen wir sie zumindest bis zur Ordnung 2.

Definition

Eine Funktion $f : B \to \mathbb{R}$ heißt auf B stetig differenzierbar, wenn f total differenzierbar ist und alle partiellen Ableitungen stetig sind. Dafür reicht aber schon aus, dass f stetig partiell differenzierbar ist. Deshalb nennen wir f auf B k–***mal stetig differenzierbar***, wenn f partielle Ableitungen bis zur Ordnung k besitzt, also Ableitungen der Form

$$f_{x_i} = \frac{\partial f}{\partial x_i}\,, \quad f_{x_i x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad \text{usw.,}$$

bis hin zu

$$\frac{\partial^{i_1+\cdots+i_n} f}{\partial x_1^{i_1} \partial x_2^{i_2} \ldots \partial x_n^{i_n}} \quad \text{mit} \quad i_1 + \cdots + i_n \le k,$$

und wenn alle partiellen Ableitungen der Ordnung k auf B noch stetig sind.

Die Menge aller k–mal stetig differenzierbaren Funktionen auf B wird mit dem Symbol $\mathcal{C}^k(B)$ bezeichnet.

Bemerkung: Ist $k \ge 1$ und $f \in \mathcal{C}^k(B)$, so ist f insbesondere in jedem Punkt von B total differenzierbar. Darüber hinaus ist f sogar „k–mal total differenzierbar“, aber dieser Begriff ist schwer zu erklären und nicht sehr intuitiv.

Wir betrachten nun eine Funktion $f \in \mathcal{C}^2(B)$ und einen Punkt $\mathbf{a} \in B$. Für eine beliebige Richtung $\mathbf{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ und kleines $\varepsilon > 0$ sei $\boldsymbol{\alpha}_{\mathbf{h}} : (-\varepsilon, \varepsilon) \to B$ definiert durch $\boldsymbol{\alpha}_{\mathbf{h}}(t) := \mathbf{a} + t\mathbf{h}$ und $g(t) := f \circ \boldsymbol{\alpha}_{\mathbf{h}}(t) = f(\mathbf{a} + t\mathbf{h})$. Dann folgt aus der speziellen Kettenregel:

$$\begin{aligned} g'(t) &= \nabla f(\alpha_{\mathbf{h}}(t)) \bullet \mathbf{h} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a} + t\mathbf{h}) \cdot h_i. \end{aligned}$$

Da $\dfrac{\partial f}{\partial x_i}(\mathbf{x})$ nach Voraussetzung auf ganz B stetige partielle Ableitungen besitzt, also insbesondere total differenzierbar ist, ist auch $g'(t)$ ein weiteres Mal differenzierbar. Es gilt:

$$\begin{aligned} g''(t) &= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \circ \boldsymbol{\alpha}_{\mathbf{h}} \right)'(t) \cdot h_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i\, \partial x_j}(\boldsymbol{\alpha}_{\mathbf{h}}(t)) \cdot h_j \right) \cdot h_i \\ &= \sum_{i,j=1}^{n} h_i \cdot \frac{\partial^2 f}{\partial x_i\, \partial x_j}(\mathbf{a} + t\mathbf{h}) \cdot h_j. \end{aligned}$$

Definition

Sei f in der Nähe von $\mathbf{a} \in \mathbb{R}^n$ zweimal stetig differenzierbar. Dann heißt die symmetrische Matrix

$$H_f(\mathbf{a}) := \left(\frac{\partial^2 f}{\partial x_i \, \partial x_j}(\mathbf{a}) \;\Big|\; i,j = 1,\ldots,n \right)$$

die ***Hesse–Matrix*** von f in $\mathbf{a}$.

Wir haben gerade ausgerechnet, dass $g''(t) = \mathbf{h} \cdot H_f(\mathbf{a} + t\mathbf{h}) \cdot \mathbf{h}^\top$ ist.

Bemerkung: Die Symmetrie der Hesse–Matrix folgt aus der Vertauschbarkeit der 2. Ableitungen, und die ist nur gegeben, weil f in einer ganzen Umgebung von $\mathbf{a}$ zweimal stetig differenzierbar ist. Diese Voraussetzung ist also wichtig!

Im Falle $n = 2$ ist $H_f(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}$.

Nun können wir die benötigte Taylorformel formulieren und beweisen:

4.2. Taylorformel 2. Ordnung

Sei $B = B_r(\mathbf{a})$ eine offene Kugel um $\mathbf{a}$, $f : B \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gibt es eine auf $B_r(\mathbf{0})$ definierte Funktion R mit

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|^2} = 0,$$

so dass für $\|\mathbf{h}\| < r$ gilt:

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + \frac{1}{2}\mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^\top + R(\mathbf{h}).$$

Beweis: Ist $\mathbf{h} \in B_r(\mathbf{0})$, so liegt $\alpha_{\mathbf{h}}(t) = \mathbf{a} + t\mathbf{h}$ für $t \in [-1,1]$ in $B_r(\mathbf{a})$, und deshalb ist $g(t) := f \circ \alpha_{\mathbf{h}}(t)$ auf $[-1,1]$ definiert und zweimal stetig differenzierbar. Wir wenden auf g in $t = 0$ den Satz von der Taylorentwicklung in einer Veränderlichen an:

Danach gibt es eine (von t abhängige) Zahl c mit $0 < c < t$, so dass gilt:

$$g(t) = g(0) + g'(0) \cdot t + \frac{1}{2} g''(0) \cdot t^2 + \eta(t) \cdot t^2,$$

mit

$$\eta(t) := \frac{1}{2} \cdot (g''(c) - g''(0)), \text{ also } \lim_{t \to 0} \eta(t) = 0.$$

Setzen wir $t = 1$, so erhalten wir:

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + \eta(1) \\ &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + \frac{1}{2}\mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^\top + \eta(1). \end{aligned}$$

Das ist die gewünschte Taylorformel, mit

$$\begin{aligned} R(\mathbf{h}) &:= \eta(1) = \frac{1}{2}\mathbf{h} \cdot (H_f(\mathbf{a}+c\mathbf{h}) - H_f(\mathbf{a})) \cdot \mathbf{h}^\top \\ &= \frac{1}{2}\sum_{i,j=1}^{n} \left(\frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{a}+c\mathbf{h}) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{a}) \right) h_i h_j \end{aligned}$$

und $0 < c < 1$. Diesen Ausdruck müssen wir noch abschätzen.

Zunächst bemerken wir, dass $|h_i| = |\mathbf{h} \bullet \mathbf{e}_i| \le \|\mathbf{h}\| \cdot \|\mathbf{e}_i\| = \|\mathbf{h}\|$ ist.

Die Summe enthält n^2 Summanden, und da f zweimal stetig differenzierbar ist, die zweiten partiellen Ableitungen also stetig sind, gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass

$$\left| \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{a}+c\mathbf{h}) - \frac{\partial^2 f}{\partial x_i\,\partial x_j}(\mathbf{a}) \right| < \varepsilon$$

für $\mathbf{h} \in B_\delta(\mathbf{0})$ ist. Für solche $\mathbf{h}$ ist dann

$$|R(\mathbf{h})| < \frac{\varepsilon}{2} \cdot n^2 \cdot \|\mathbf{h}\|^2.$$

Daraus ergibt sich die gewünschte Limesbeziehung: $\lim\limits_{\mathbf{h}\to\mathbf{0}} \dfrac{R(\mathbf{h})}{\|\mathbf{h}\|^2} = 0\,.$ ■

Es gibt selbstverständlich auch Taylorformeln höherer Ordnung, darauf gehen wir am Ende dieses Abschnittes ein.

Ist nun f in $\mathbf{a}$ stationär, also $\nabla f(\mathbf{a}) = \mathbf{0}$ und

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^\top + R(\mathbf{h}),$$

so hängt das Verhalten von f in der Nähe von $\mathbf{a}$ im Wesentlichen von der Hesse-Matrix ab, denn $R(\mathbf{h})$ verschwindet ja in $\mathbf{a}$ von höherer Ordnung. Das führt uns zu einem ähnlichen hinreichenden Kriterium für Extremwerte, wie wir es aus der eindimensionalen Theorie kennen. Allerdings ist die Lage hier doch etwas komplizierter.

Ist $A \in M_n(\mathbb{R})$ eine symmetrische Matrix und $\varphi_A(\mathbf{x},\mathbf{y}) := \mathbf{x} \cdot A \cdot \mathbf{y}^\top$ die zugehörige symmetrische Bilinearform, so nennt man bekanntlich die Funktion

$$q(\mathbf{h}) = q_A(\mathbf{h}) := \mathbf{h} \cdot A \cdot \mathbf{h}^\top$$

die zugehörige ***quadratische Form***. Es ist

$$q(t\mathbf{h}) = t^2 \cdot q(\mathbf{h}) \text{ für } t \in \mathbb{R} \text{ und } \mathbf{h} \in \mathbb{R}^n.$$

Insbesondere ist natürlich $q(\mathbf{0}) = 0$.

Definition

Eine quadratische Form $q(\mathbf{h})$ heißt

$$\begin{aligned}
\textbf{\textit{positiv semidefinit}} &:\iff && q(\mathbf{h}) \geq 0 \text{ für alle } \mathbf{h}, \\
\textbf{\textit{positiv definit}} &:\iff && q(\mathbf{h}) > 0 \text{ für alle } \mathbf{h} \neq \mathbf{0}, \\
\textbf{\textit{negativ semidefinit}} &:\iff && q(\mathbf{h}) \leq 0 \text{ für alle } \mathbf{h}, \\
\textbf{\textit{negativ definit}} &:\iff && q(\mathbf{h}) < 0 \text{ für alle } \mathbf{h} \neq \mathbf{0}, \\
\textbf{\textit{indefinit}} &:\iff && \exists\, \mathbf{h}_1, \mathbf{h}_2 \text{ mit } q(\mathbf{h}_1) < 0 < q(\mathbf{h}_2).
\end{aligned}$$

Es sei an folgendes Ergebnis aus der linearen Algebra erinnert: Alle Eigenwerte einer symmetrischen Matrix $A \in M_n(\mathbb{R})$ sind reell und es gibt im $\mathbb{R}^n$ eine Orthonormalbasis von Eigenvektoren von A. Sind nun

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

die n (reellen) Eigenwerte von A und ist $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ die zugehörige ON-Basis von Eigenvektoren von A, so kann man jeden Vektor $\mathbf{h} \in \mathbb{R}^n$ in der Form

$$\mathbf{h} = h_1\mathbf{a}_1 + \cdots + h_n\mathbf{a}_n$$

darstellen, und es folgt:

$$q_A(\mathbf{h}) := \mathbf{h} \cdot A \cdot \mathbf{h}^\top = \sum_{i,j} h_i h_j (\mathbf{a}_i \cdot A \cdot \mathbf{a}_j^\top) = \sum_{i,j} h_i h_j \mathbf{a}_i \cdot (\lambda_j \mathbf{a}_j^\top) = \sum_{i=1}^{n} \lambda_i (h_i)^2.$$

Daraus kann man sofort ablesen:

$$\begin{aligned}
q_A \textbf{ positiv definit} \quad &\iff \quad \mathbf{h} \cdot A \cdot \mathbf{h}^\top > 0 \text{ für alle } \mathbf{h} \neq \mathbf{0} \\
&\iff \quad \sum_{i=1}^{n} \lambda_i (h_i)^2 > 0 \text{ für alle } (h_1, \ldots, h_n) \neq (0, \ldots, 0) \\
&\iff \quad \lambda_1, \ldots, \lambda_n > 0.
\end{aligned}$$

Genauso sieht man, dass q_A genau dann **negativ definit** ist, wenn alle $\lambda_i < 0$ sind, und genau dann **indefinit**, wenn es ein i und ein j mit $\lambda_i < 0 < \lambda_j$ gibt.

Im Falle $n = 2$ gibt es noch ein einfacheres Kriterium:

4.3. Definitheit im Falle der Dimension 2

Sei $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in M_2(\mathbb{R})$ *eine symmetrische Matrix. Dann gilt:*

1. *Ist* $\det(A) < 0$*, so ist* q_A *indefinit.*
2. *Ist* $\det(A) > 0$ *und* $a > 0$*, so ist* q_A *positiv definit.*
3. *Ist* $\det(A) > 0$ *und* $a < 0$*, so ist* q_A *negativ definit.*

Beweis: Sei $\Delta := \det(A) = ad - b^2$. Zur Berechnung der Eigenwerte brauchen wir noch das charakteristische Polynom:

$$p_A(x) = \det \begin{pmatrix} a-x & b \\ b & d-x \end{pmatrix} = (a-x)(d-x) - b^2 = x^2 - (a+d)x + \Delta.$$

Die Eigenwerte λ_1, λ_2 von A sind die beiden Nullstellen dieses quadratischen Polynoms. Die Gleichungen von Vieta liefern

$$\lambda_1 + \lambda_2 = a + d \quad \text{und} \quad \lambda_1 \cdot \lambda_2 = \Delta.$$

Ist $\Delta < 0$, so haben die beiden Eigenwerte verschiedenes Vorzeichen, und q_A ist indefinit. Ist $\Delta > 0$, so sind λ_1 und λ_2 beide $\neq 0$, und sie haben das gleiche Vorzeichen. Außerdem ist $ad = \Delta + b^2 > 0$. Ist nun $a > 0$, so ist auch $d > 0$ und damit $\lambda_1 + \lambda_2 > 0$. In diesem Fall ist q_A positiv definit. Genauso folgt aus $a < 0$, dass q_A negativ definit ist. ∎

Jetzt wenden wir die Theorie der quadratischen Formen auf die Extremwertbestimmung an.

4.4. Hinreichendes Kriterium für Extremwerte

Sei $B \subset \mathbb{R}^n$ offen, $f \in \mathcal{C}^2(B)$. Weiter sei $\mathbf{a} \in B$ ein stationärer Punkt von f, also $\nabla f(\mathbf{a}) = \mathbf{0}$.

1. *Ist $H_f(\mathbf{a})$ positiv definit, so besitzt f in $\mathbf{a}$ ein relatives Minimum.*
2. *Ist $H_f(\mathbf{a})$ negativ definit, so besitzt f in $\mathbf{a}$ ein relatives Maximum.*
3. *Ist $H_f(\mathbf{a})$ indefinit, so liegt in $\mathbf{a}$ ein Sattelpunkt vor.*

Beweis:
1) Sei $H_f(\mathbf{a})$ positiv definit und $q(\mathbf{h}) := \mathbf{h} \cdot H_f(\mathbf{a}) \cdot \mathbf{h}^\top$. Da f in $\mathbf{a}$ stationär ist, ergibt die Taylorformel:

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2} q(\mathbf{h}) + R(\mathbf{h}).$$

Die Funktion q ist stetig und nach Voraussetzung > 0 außerhalb des Nullpunktes. Insbesondere nimmt sie auf der abgeschlossenen und beschränkten und daher kompakten Menge

$$S^{n-1} := \{\mathbf{x} \in \mathbb{R}^n \,:\, \|\mathbf{x}\| = 1\}$$

ein Minimum $m > 0$ an. Daher gilt für beliebiges $\mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$:

$$q(\mathbf{h}) = \|\mathbf{h}\|^2 \cdot q\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right) \geq m \cdot \|\mathbf{h}\|^2.$$

Ist jetzt ein ε mit $0 < \varepsilon < m/2$ vorgegeben und dazu ein $r = r(\varepsilon)$ so gewählt, dass

$$|R(\mathbf{h})| \le \varepsilon \cdot \|\mathbf{h}\|^2 \text{ für } \mathbf{h} \in B_r(\mathbf{0})$$

ist, so ist

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}q(\mathbf{h}) + R(\mathbf{h}) \ge (\frac{m}{2} - \varepsilon) \cdot \|\mathbf{h}\|^2 \ge 0$$

für alle $\mathbf{h} \in B_r(\mathbf{0})$.

Also ist $f(\mathbf{a}+\mathbf{h}) \ge f(\mathbf{a})$ für kleines $\mathbf{h}$, und es liegt ein relatives Minimum in $\mathbf{a}$ vor.

2) Der Fall des Maximums kann durch Übergang von f zu $-f$ auf (1) zurückgeführt werden.

3) Ist q indefinit, so gibt es in jeder Umgebung von $\mathbf{0}$ Vektoren $\mathbf{h}_1$ und $\mathbf{h}_2$ mit $q(\mathbf{h}_1) < 0 < q(\mathbf{h}_2)$. Die Funktionen

$$\begin{aligned} f_1(t) &:= f(\mathbf{a}+t\mathbf{h}_1) \\ \text{und} \quad f_2(t) &:= f(\mathbf{a}+t\mathbf{h}_2) \end{aligned}$$

sind dann definiert und zweimal differenzierbar, und es gilt:

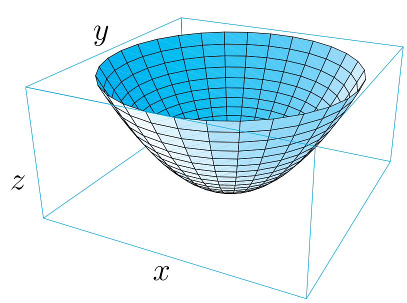
$$(f_1)'(0) = (f_2)'(0) = 0, \quad (f_1)''(0) = q(\mathbf{h}_1) < 0 \quad \text{und} \quad (f_2)''(0) = q(\mathbf{h}_2) > 0.$$

Also besitzt f_1 in $t = 0$ ein isoliertes Maximum und f_2 in $t = 0$ ein isoliertes Minimum. Das bedeutet, dass f beliebig nahe bei $\mathbf{a}$ sowohl Werte $< f(\mathbf{a})$ als auch Werte $> f(\mathbf{a})$ annimmt. Damit liegt ein Sattelpunkt vor. ∎

Bemerkung: Ist $H_f(\mathbf{a})$ nur semidefinit, so kann man keine genaue Aussage machen!

4.5. Beispiele

A. Sei $f(x,y) := x^2 + y^2$. Dann ist $\nabla f(x,y) = (2x, 2y)$, also $(0,0)$ der einzige stationäre Punkt von f.



Da $f(0,0) = 0$ und allgemein $f(x,y) \ge 0$ ist, liegt ein absolutes Minimum vor.

Tatsächlich ist

$$H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

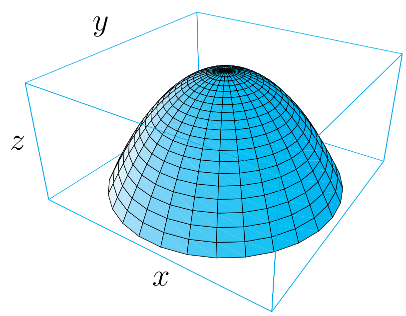
Offensichtlich ist $H_f(x,y)$ positiv definit. Das hinreichende Kriterium bestätigt also, dass f im Nullpunkt ein lokales Minimum besitzt.

B. Sei $f(x,y) := 1 - x^2 - y^2$. Dann ist $\nabla f(x,y) = (-2x, -2y)$ und wieder $(0,0)$ der einzige stationäre Punkt.

Die Hesse-Matrix

$$H_f(x,y) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

ist offensichtlich negativ definit. Also liegt im Nullpunkt ein Maximum vor.

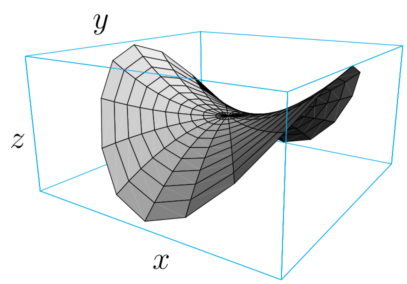


C. Sei $f(x,y) := x^2 - y^2$.

In diesem Falle ist $\nabla f(x,y) = (2x, -2y)$ und

$$H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Da $\det H_f(x,y) < 0$ ist, liegt im Nullpunkt ein Sattelpunkt vor.



D. Sei $f(x,y) := x^4 - 2x^2 + 2x^2y^2 - y^2$. Dann ist

$$\nabla f(x,y) = (4x^3 - 4x + 4xy^2, 4x^2y - 2y) = \big(4x(x^2 - 1 + y^2), 2y(2x^2 - 1)\big).$$

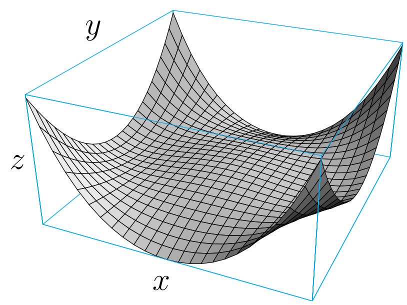
Zunächst bestimmen wir die kritischen Punkte. Sei also $\nabla f(x,y) = (0,0)$.

- Ist $x = 0$, so muss auch $y = 0$ sein.
- Ist $y = 0$ und $x \neq 0$, so muss $x^2 - 1 = 0$, also $x = \pm 1$ sein.
- Ist $x \neq 0$ und $y \neq 0$, so muss $x^2 + y^2 = 1$ und $2x^2 = 1$ sein. Dann ist $x^2 = y^2 = 1/2$, also $x = \pm 1/\sqrt{2}$ und $y = \pm x$.

Das ergibt die sieben kritischen Punkte

$(0,0)$, $\pm(1,0)$, $\pm\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

und $\pm\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.



Als Hesse-Matrix ergibt sich

$$H_f(x,y) = \begin{pmatrix} 12x^2 - 4 + 4y^2 & 8xy \\ 8xy & 4x^2 - 2 \end{pmatrix}.$$

Setzen wir die kritischen Punkte ein, so erhalten wir:

$$H_f(0,0) = \begin{pmatrix} -4 & 0 \\ 0 & -2 \end{pmatrix} \text{ ist negativ definit,}$$

$$H_f(\pm 1,0) = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix} \text{ ist positiv definit,}$$

$$H_f\Big(\pm\,(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})\Big) = \begin{pmatrix} 4 & 4 \\ 4 & 0 \end{pmatrix} \text{ ist indefinit}$$

$$\text{und}\quad H_f\Big(\pm\,(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}})\Big) = \begin{pmatrix} 4 & -4 \\ -4 & 0 \end{pmatrix} \text{ ist ebenfalls indefinit.}$$

Demnach liegt im Nullpunkt ein Maximum vor, in den Punkten $(1,0)$ und $(-1,0)$ Minima und in den anderen kritischen Punkten Sattelpunkte.

Anhang: Taylorformel für Funktionen von mehreren Veränderlichen:

Sei f in der Nähe von $\mathbf{x}_0 \in \mathbb{R}^n$ genügend oft differenzierbar. Wir betrachten den Weg $\boldsymbol{\alpha}(t) := \mathbf{x}_0 + t\mathbf{h}$, mit $\mathbf{h} := \mathbf{x} - \mathbf{x}_0$, und untersuchen die Funktion

$$g(t) := f \circ \boldsymbol{\alpha}(t) = f(\mathbf{x}_0 + t\mathbf{h}).$$

Auf jeden Fall ist

$$g'(t) = \nabla f(\mathbf{x}_0 + t\mathbf{h}) \bullet \mathbf{h} = \sum_{\nu=1}^{n} h_\nu \frac{\partial f}{\partial x_\nu}(\mathbf{x}_0 + t\mathbf{h}).$$

Wir wollen die höheren Ableitungen von g berechnen.

Sei P der Differentialoperator

$$P = \mathbf{h} \bullet \nabla := h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\,.$$

Dann ist $(Pf) \circ \boldsymbol{\alpha} = (f \circ \boldsymbol{\alpha})'$, und per Induktion folgt: $(P^k f) \circ \boldsymbol{\alpha} = (f \circ \boldsymbol{\alpha})^{(k)}$.

Der Induktionsschritt sieht dabei folgendermaßen aus:

$$\begin{aligned} (P^{k+1} f) \circ \boldsymbol{\alpha} &= P(P^k f) \circ \boldsymbol{\alpha} = ((P^k f) \circ \boldsymbol{\alpha})' \\ &= ((f \circ \boldsymbol{\alpha})^{(k)})' = (f \circ \boldsymbol{\alpha})^{(k+1)}. \end{aligned}$$

Um $g^{(k)}(t) = (\mathbf{h} \bullet \nabla)^k f(\mathbf{x}_0 + t\mathbf{h})$ zu berechnen, brauchen wir die folgende Formel:

4.6. Satz

Für $x_1, \ldots, x_n \in \mathbb{R}$ und $k \in \mathbb{N}$ ist $\quad (x_1 + \cdots + x_n)^k = \displaystyle\sum_{\nu_1 + \cdots + \nu_n = k} \frac{k!}{\nu_1! \cdots \nu_n!} x_1^{\nu_1} \cdots x_n^{\nu_n}.$

Beweis: (Induktion nach n)

Der Induktionsanfang ist trivial. Zum Induktionsschluss:

$$\begin{aligned}(x_1+\cdots+x_{n+1})^k &= ((x_1+\cdots+x_n)+x_{n+1})^k\\ &= \sum_{m+\nu_{n+1}=k}\frac{k!}{m!\nu_{n+1}!}(x_1+\cdots+x_n)^m x_{n+1}^{\nu_{n+1}}\\ &= \sum_{m+\nu_{n+1}=k}\frac{k!}{m!\nu_{n+1}!}\sum_{\nu_1+\cdots+\nu_n=m}\frac{m!}{\nu_1!\cdots\nu_n!}x_1^{\nu_1}\cdots x_n^{\nu_n}\cdot x_{n+1}^{\nu_{n+1}}\\ &= \sum_{\nu_1+\cdots+\nu_{n+1}=k}\frac{k!}{\nu_1!\cdots\nu_{n+1}!}x_1^{\nu_1}\cdots x_{n+1}^{\nu_{n+1}}.\end{aligned}$$

■

Da die h_i Konstanten und die partiellen Ableitungen vertauschbar sind, kann man $(\mathbf{h}\bullet\nabla)^k$ nach der gleichen Formel wie der für den Ausdruck $(x_1+\cdots+x_n)^k$ berechnen. Es folgt:

$$g^{(k)}(t) = (\mathbf{h}\bullet\nabla)^k f(\mathbf{x}_0+t\mathbf{h}) = k!\sum_{|\nu|=k}\frac{1}{\nu!}D^\nu f(\mathbf{x}_0+t\mathbf{h})\cdot\mathbf{h}^\nu\,.$$

Dabei ist $\nu! := \nu_1!\cdots\nu_n!$, $|\nu| := \nu_1+\cdots+\nu_n$ und $D^\nu f := D_1^{\nu_1}D_2^{\nu_2}\cdots D_n^{\nu_n}f$, sowie $\mathbf{h}^\nu := h_1^{\nu_1}\cdots h_n^{\nu_n}$ für einen Vektor $\mathbf{h}=(h_1,\ldots,h_n)$.

Ist f k-mal differenzierbar, so nennt man

$$T_k f(\mathbf{x};\mathbf{x}_0) := \sum_{|\nu|\le k}\frac{1}{\nu!}D^\nu f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)^\nu$$

das k-te ***Taylorpolynom*** von f in $\mathbf{x}_0$.

4.7. Satz (Taylorentwicklung)

Sei $B\subset\mathbb{R}^n$ eine offene konvexe Menge, $\mathbf{x}_0\in B$ und $f: B\to\mathbb{R}$ eine k-mal stetig differenzierbare Funktion. Dann gibt es eine Darstellung $f = T_k f + R_k$, wobei gilt:

1. $\displaystyle\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{R_k(\mathbf{x})}{\|\mathbf{x}-\mathbf{x}_0\|^k}=0.$
2. *Ist f sogar $(k+1)$-mal differenzierbar, so gibt es zu jedem $\mathbf{x}\in B$ ein $\xi\in[0,1]$, so dass gilt:*

$$R_k(\mathbf{x}) = \sum_{|\nu|=k+1}\frac{1}{\nu!}D^\nu f(\mathbf{x}_0+\xi(\mathbf{x}-\mathbf{x}_0))(\mathbf{x}-\mathbf{x}_0)^\nu.$$

Beweis: Wir betrachten zunächst den Fall, dass f sogar $(k+1)$-mal differenzierbar ist. Sei $\boldsymbol{\alpha}(t) := \mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)$. Dann ist auch $g(t) := f \circ \boldsymbol{\alpha}(t)$ $(k+1)$-mal differenzierbar. Die Taylorformel in einer Veränderlichen liefert zu jedem t ein $\xi = \xi(t)$ zwischen 0 und t, so dass gilt:

$$g(t) = \sum_{i=0}^{k} \frac{g^{(i)}(0)}{i!} t^i + \frac{1}{(k+1)!} g^{(k+1)}(\xi) t^{k+1} .$$

Setzen wir $t = 1$, so erhalten wir

$$f(\mathbf{x}) = \sum_{|\nu| \le k} \frac{1}{\nu!} D^\nu f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^\nu + \sum_{|\nu| = k+1} \frac{1}{\nu!} D^\nu f(\mathbf{x}_0 + \xi(\mathbf{x} - \mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0)^\nu .$$

Ist f nur k-mal stetig differenzierbar, so setzen wir $\mathbf{h} := \mathbf{x} - \mathbf{x}_0$ und erhalten

$$\begin{aligned} f(\mathbf{x}) &= T_{k-1} f(\mathbf{x}; \mathbf{x}_0) + \sum_{|\nu| = k} \frac{1}{\nu!} D^\nu f(\mathbf{x}_0 + \xi \mathbf{h}) \mathbf{h}^\nu \\ &= T_k f(\mathbf{x}; \mathbf{x}_0) + \sum_{|\nu| = k} \frac{1}{\nu!} \big(D^\nu f(\mathbf{x}_0 + \xi \mathbf{h}) - D^\nu f(\mathbf{x}_0) \big) \mathbf{h}^\nu . \end{aligned}$$

Setzen wir $\varphi_\nu(\mathbf{h}) := \dfrac{1}{\nu!} \big(D^\nu f(\mathbf{x}_0 + \xi \mathbf{h}) - D^\nu f(\mathbf{x}_0) \big)$, so erhalten wir

$$f(\mathbf{x}) = T_k f(\mathbf{x}; \mathbf{x}_0) + \sum_{|\nu| = k} \varphi_\nu(\mathbf{h}) \mathbf{h}^\nu .$$

Für $|\nu| = k$ ist $\frac{|\mathbf{h}^\nu|}{\|\mathbf{h}\|^k} = \frac{|h_1|^{\nu_1} \cdots |h_n|^{\nu_n}}{\|\mathbf{h}\|^{\nu_1} \cdots \|\mathbf{h}\|^{\nu_n}} \le 1$. Daraus folgt:

$$\Big| \sum_{|\nu| = k} \varphi_\nu(\mathbf{h}) \mathbf{h}^\nu \Big| / \|\mathbf{h}\|^k \le \sum_{|\nu| = k} |\varphi_\nu(\mathbf{h})| \to 0 \text{ für } \mathbf{h} \to \mathbf{0},$$

wegen der Stetigkeit von $D^\nu f$ in $\mathbf{x}_0$. ∎