

16. FUNKTIONEN
VON
MEHREREN VARIABLEN

Reelle Funktionen auf dem $\mathbb{R}^2$

Wir betrachten Funktionen

$$f(x_1, x_2)$$

von zwei reellen Variablen x_1, x_2 , z.B.

$$f(x_1, x_2) = x_1^2 + x_2^2, \quad g(x_1, x_2) = x_1^2 - x_2^2$$

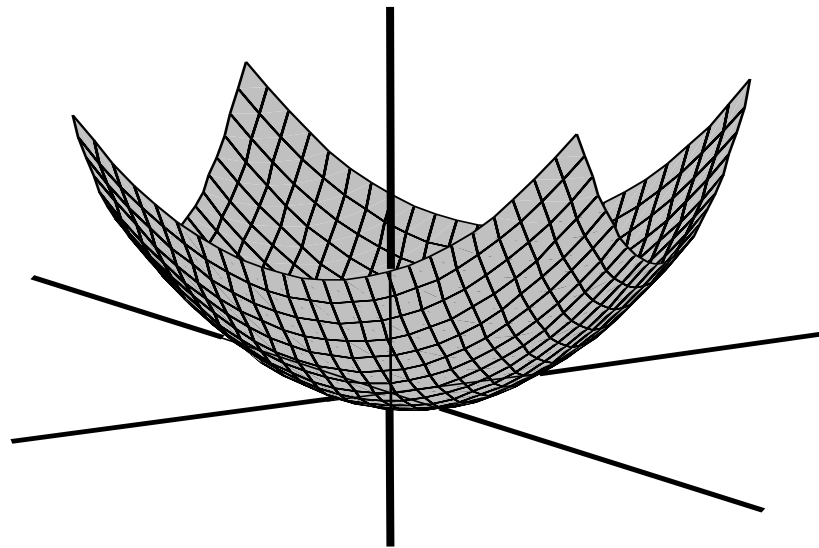
$$h(x_1, x_2) = x_1^3 - 3x_1x_2^2, \quad k(x, y) = \cos(\sqrt{x_1^2 + x_2^2})$$

Solche Funktionen lassen sich graphisch verschieden darstellen.

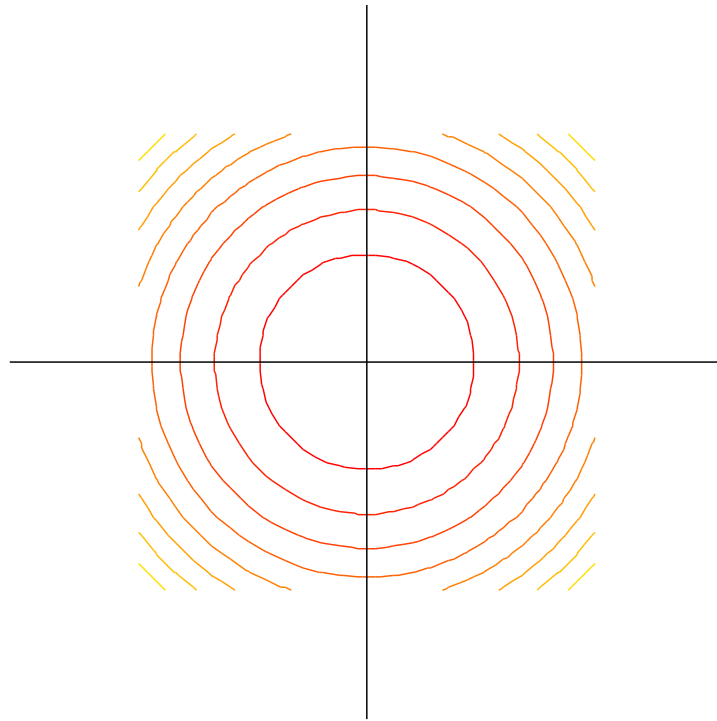
Als Fläche im $\mathbb{R}^3$ oder als Graph ihrer Höhenlinien.

$$f(x_1, x_2) = x_1^2 + x_2^2$$

Paraboloid

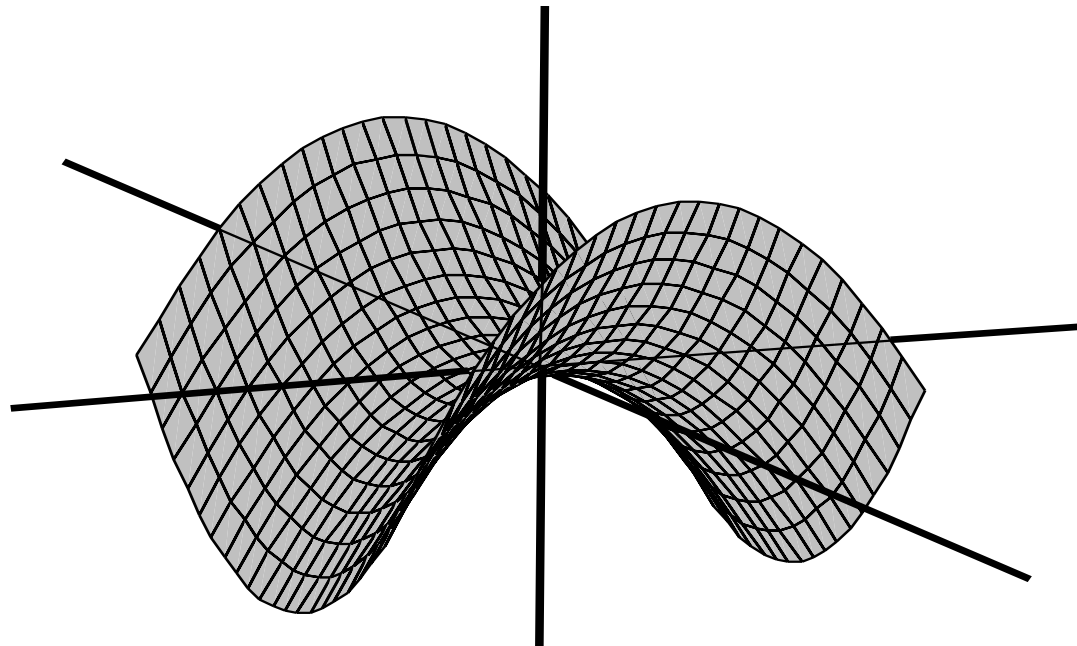


$$f(x_1, x_2) = x_1^2 + x_2^2 \quad \text{Paraboloid}$$

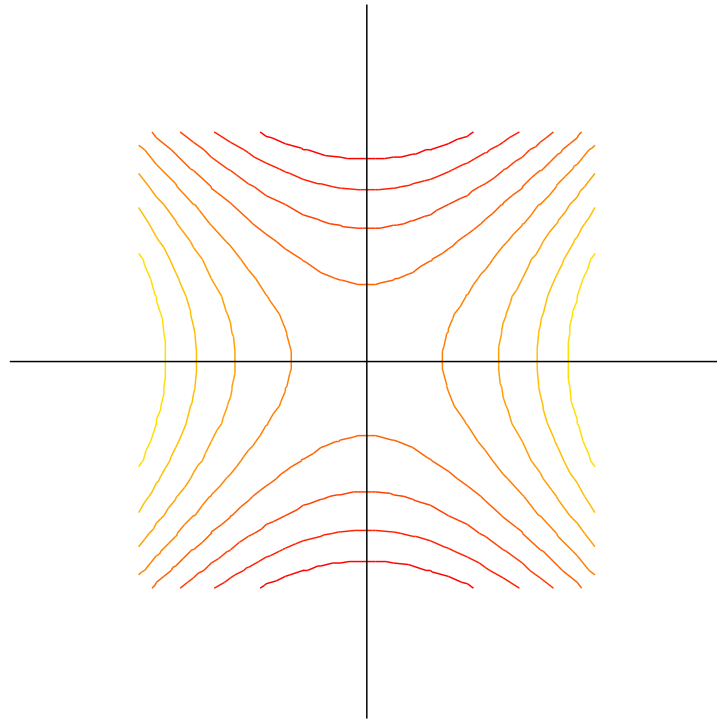


$$g(x_1, x_2) = x_1^2 - x_2^2$$

Sattelfläche

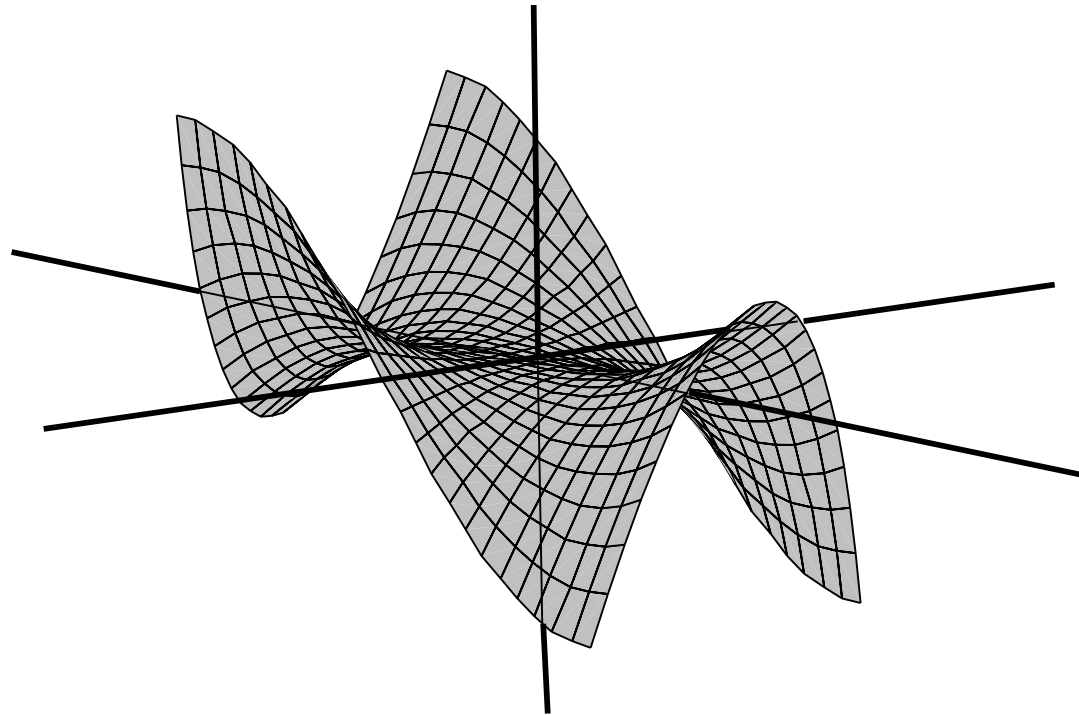


$$g(x_1, x_2) = x_1^2 - x_2^2 = \Re(x_1 + ix_2)^2$$

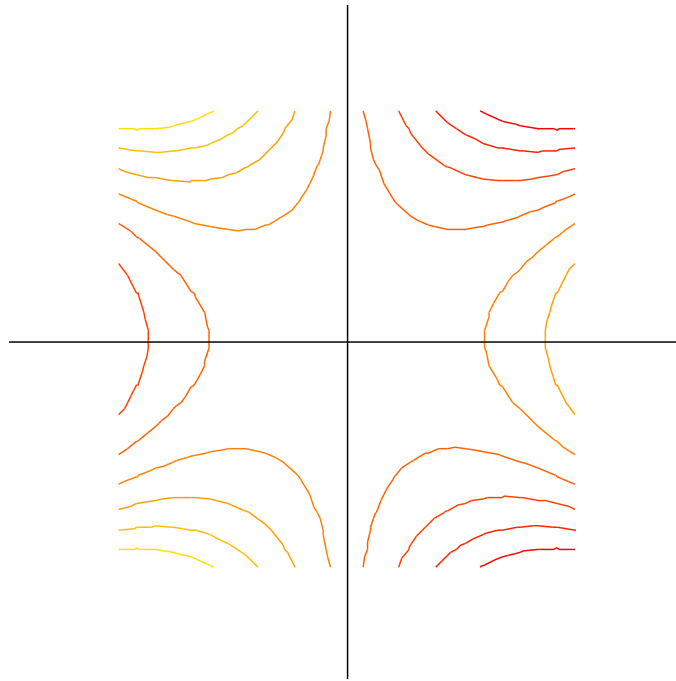


$$h(x_1, x_2) = x_1^3 - 3x_1x_2^2$$

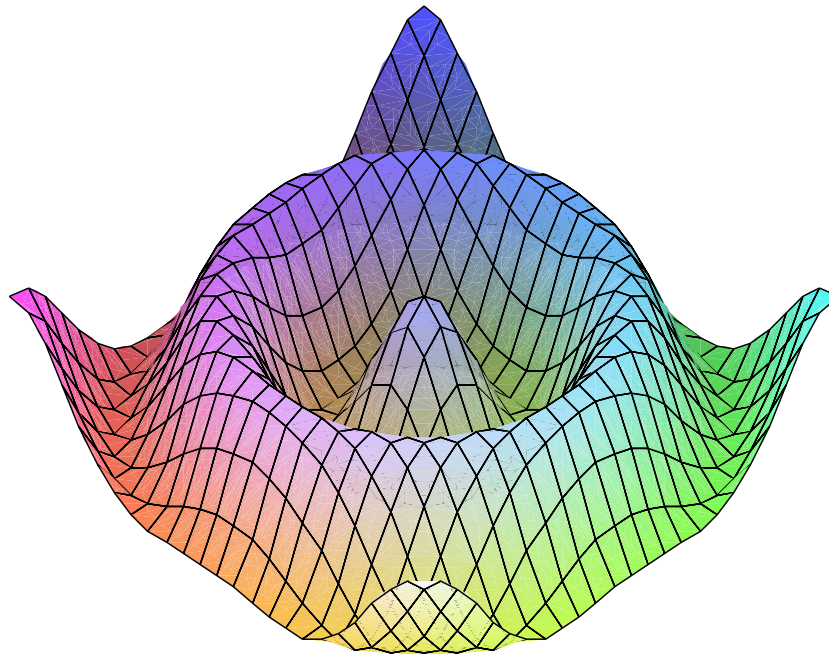
Affensattel



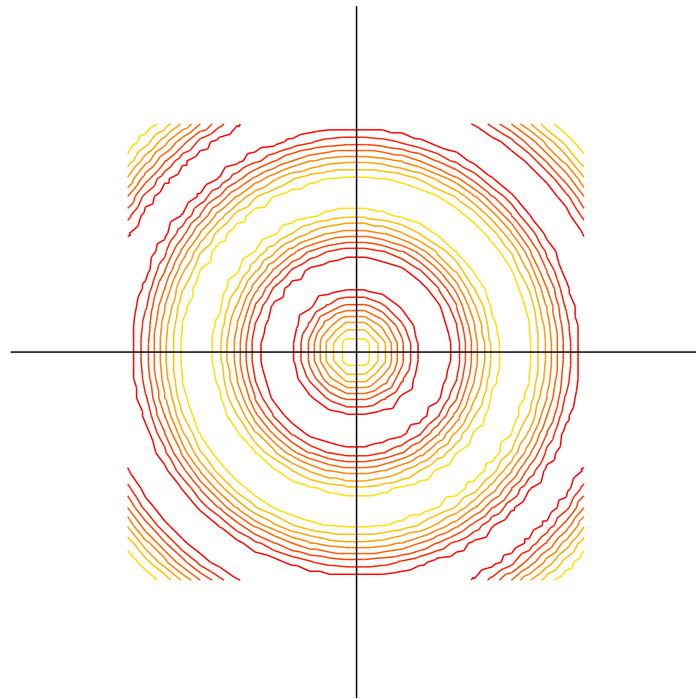
$$h(x_1, x_2) = \Re(x_1 + ix_2)^3$$



$$k(x_1, x_2) = \cos(\sqrt{x_1^2 + x_2^2})$$



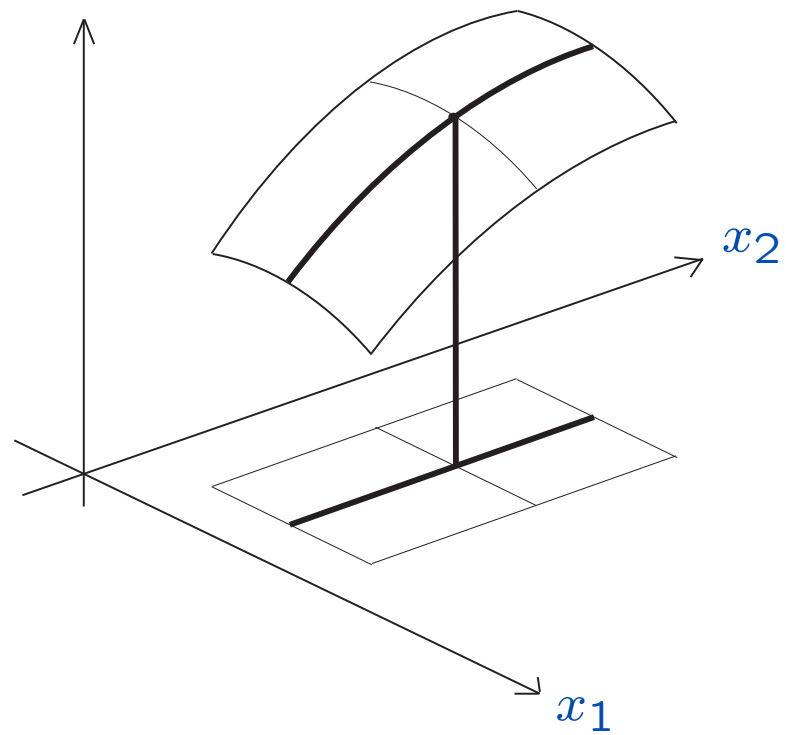
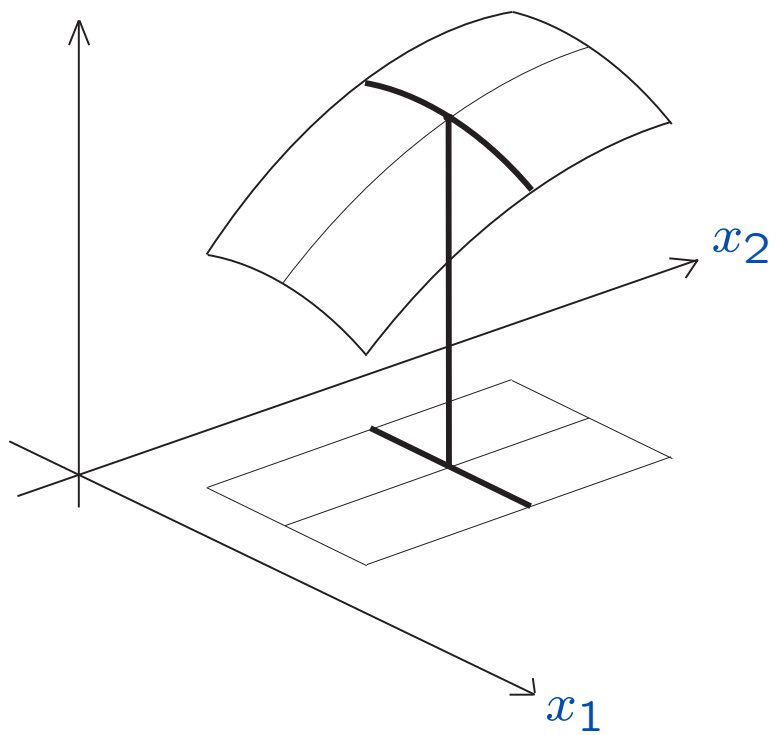
$$k(x_1, x_2) = \cos(\sqrt{x_1^2 + x_2^2})$$



Partielle Ableitungen:

In Funktionen von mehreren Variablen finden sich Funktionen von einer Variablen, wenn man die anderen Variablen festhält:

$f(x_1, x_2)$ kann man bei festem x_2 als Funktion in x_1 und bei festem x_1 als Funktion in x_2 betrachten.

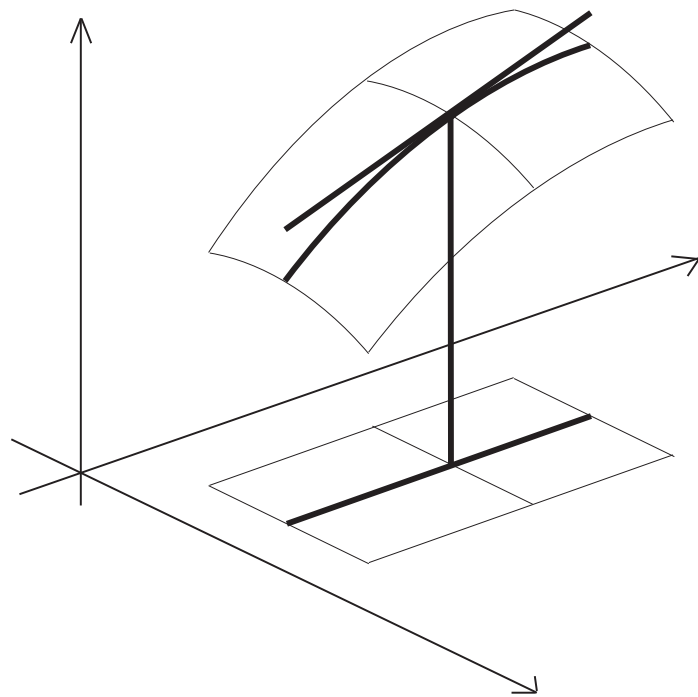
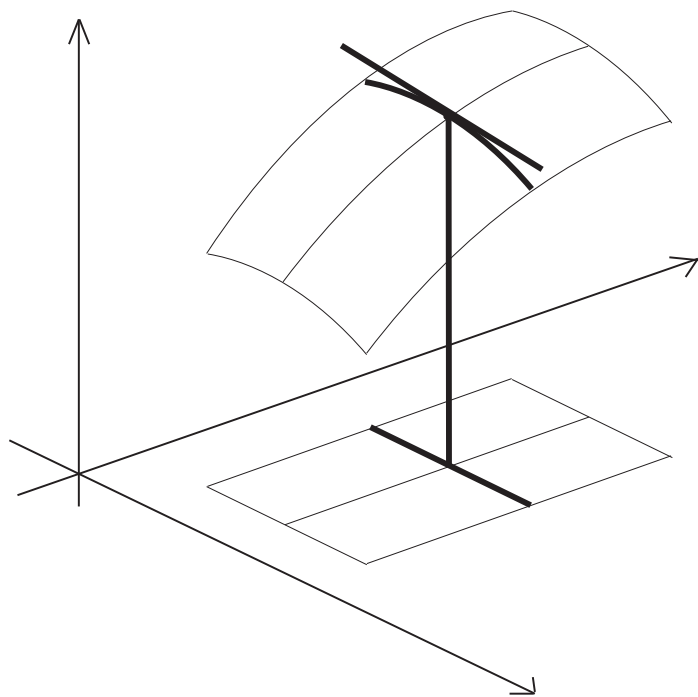


Wir können dann wie früher Ableitungen bilden, mit festgehaltenem x_2 oder festgehaltenem x_1 :

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \lim_{h \rightarrow 0} \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h}$$

$$\frac{\partial f}{\partial x_2}(x_1, x_2) = \lim_{h \rightarrow 0} \frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h}$$

Diese Ableitungen heißen *partielle Ableitungen*. (Voraussetzung ist natürlich, dass diese Ableitungen existieren.)



Partielle Ableitungen können zu denselben Zwecken benutzt werden wie gewöhnliche, etwa zum Bestimmen von Extrempunkten (lokalen Maxima und Minima) und allgemeiner stationären Punkten.

Definition.

(x_1, x_2) heißt *stationärer Punkt* von f , falls

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{\partial f}{\partial x_2}(x_1, x_2) = 0$$

gilt.

Beispiele:

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

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

$(x_0, y_0) = (0, 0)$ ist stationärer Punkt, hier ein *Minimum*.

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

$$\frac{\partial f}{\partial x}(x, y) = 2x, \frac{\partial f}{\partial y}(x, y) = -2y$$

$(x_0, y_0) = (0, 0)$ ist stationärer Punkt, hier ein *Sattelpunkt*.

3. Lineare Regression: An Datenpunkte $x_1, \dots, x_n$ sollen Punkte auf einer Geraden $\alpha + \beta t_1, \dots, \alpha + \beta t_n$ angepasst werden. Bestimme α, β so, dass der Abstand im $\mathbb{R}^n$

$$f(\alpha, \beta) = (x_1 - \alpha - \beta t_1)^2 + \dots + (x_n - \alpha - \beta t_n)^2$$

minimal wird. Es gilt

$$\begin{aligned}\frac{\partial f}{\partial \alpha}(\alpha, \beta) &= -2(x_1 - \alpha - \beta t_1) - \dots - 2(x_n - \alpha - \beta t_n) \\ \frac{\partial f}{\partial \beta}(\alpha, \beta) &= -2(x_1 - \alpha - \beta t_1)t_1 - \dots - 2(x_n - \alpha - \beta t_n)t_n\end{aligned}$$

Die Gleichungen $\frac{\partial f}{\partial \alpha}(\alpha, \beta) = \frac{\partial f}{\partial \beta}(\alpha, \beta) = 0$ sind die schon früher abgeleiteten „Normalgleichungen“ für α, β .

Totale Differenzierbarkeit:

Wir fassen $f(x_1, x_2)$ nun als Funktion auf dem $\mathbb{R}^2$ und schreiben

$$f(x_1, x_2) = f(\mathbf{x}) \quad \text{mit} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$$

Damit schließen wir an die Lineare Algebra an.

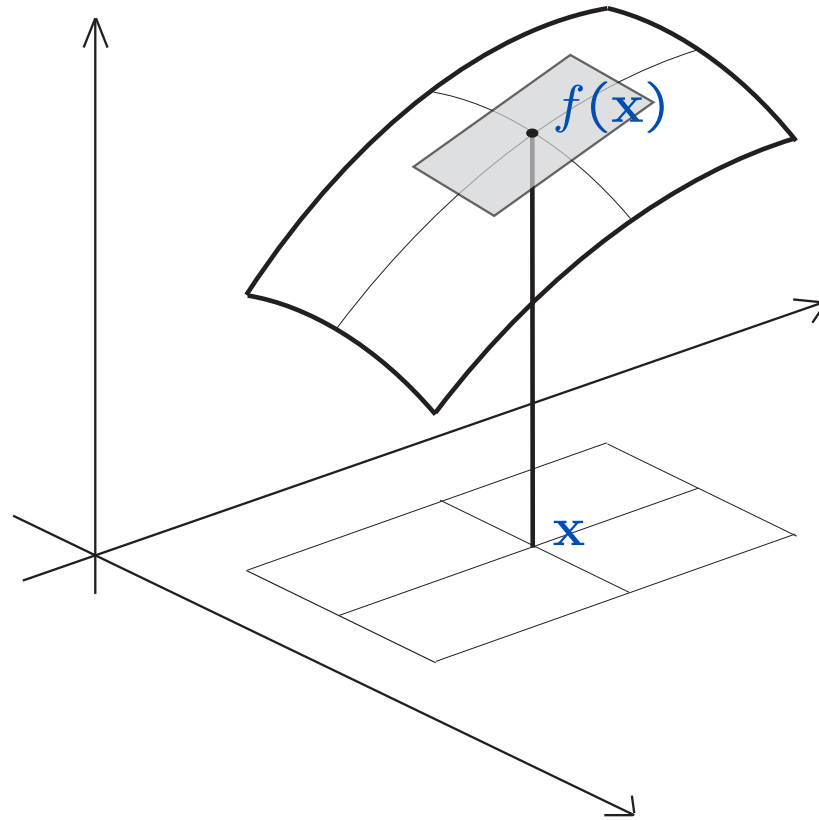
Definition.

Eine Abbildung $f : D \rightarrow \mathbb{R}$ mit $D \subset \mathbb{R}^2$ heißt *total differenzierbar* im Punkte $\mathbf{x} \in \mathbb{R}^2$, falls es eine Linearform $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$ gibt, so dass für Vektoren $\mathbf{u}$ gilt

$$f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + \ell(\mathbf{u}) + o(|\mathbf{u}|) \quad \text{für } \mathbf{u} \rightarrow \mathbf{0}$$

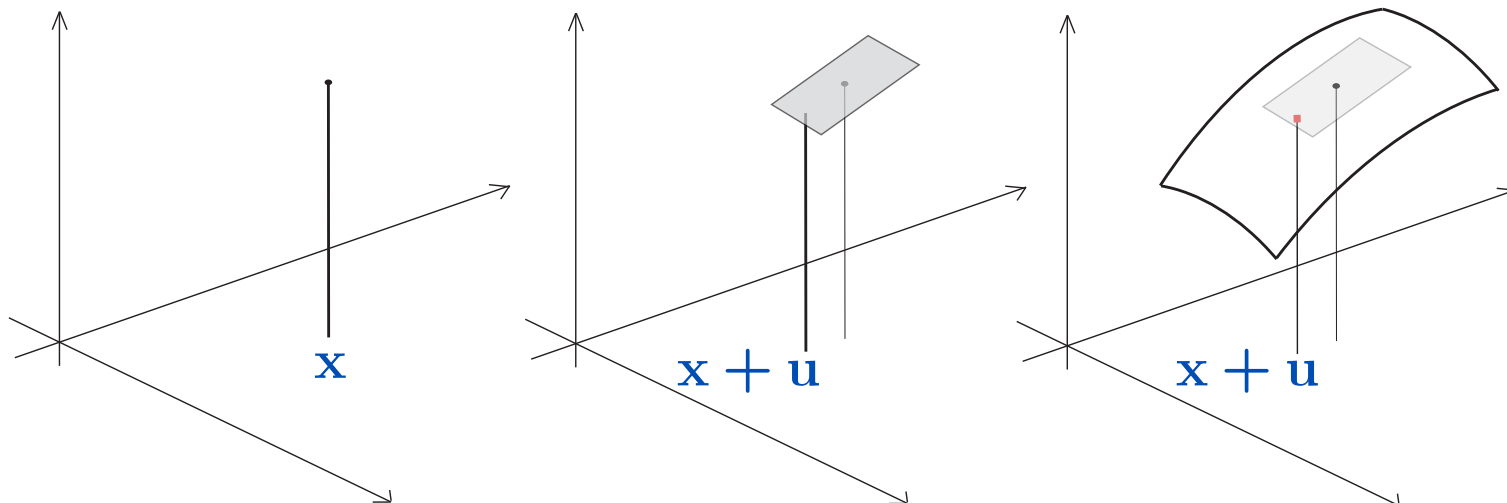
$\mathbf{u} \rightarrow \mathbf{0}$ bedeutet dasselbe wie $|\mathbf{u}| \rightarrow 0$. Eine gebräuchliche Schreibweise für die (von $\mathbf{x}$ abhängige) Linearform ℓ ist

$$\ell = Df(\mathbf{x})$$



$$f(\mathbf{x} + \mathbf{u}) =$$

$$f(\mathbf{x}) \quad + \quad \ell(\mathbf{u}) \quad + \quad o(|\mathbf{u}|)$$



Kurz gesprochen:

Wir haben f *lokal linearisiert*, approximativ um $\mathbf{x}$ herum.

Die Linearform ℓ hat die Gestalt

$$\ell(\mathbf{u}) = \mathbf{a}\mathbf{u} = a_1u_1 + a_2u_2$$

mit einem Zeilenvektor $\mathbf{a} = (a_1, a_2)$ (der von $\mathbf{x}$ abhängt). Wie kann man $\mathbf{a}$ berechnen?

Es gilt

$$a_1 = \frac{\partial f}{\partial x_1}(x_1, x_2) \quad \text{und} \quad a_2 = \frac{\partial f}{\partial x_2}(x_1, x_2)$$

bzw.

$$\mathbf{a} = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}) \right)$$

Zum Beweis wählen wir $\mathbf{u} = (h, 0)^T$, so dass für $h \rightarrow 0$

$$\begin{aligned} f(x_1 + h, x_2) &= f(\mathbf{x}) + \ell(\mathbf{u}) + o(|\mathbf{u}|) \\ &= f(x_1, x_2) + a_1 h + o(|h|) \end{aligned}$$

Also ist a_1 die Ableitung von $f(x_1, x_2)$ nach x_1 bei festem x_2 , also tatsächlich

$$a_1 = \frac{\partial f}{\partial x_1}(x_1, x_2)$$

Beispiel: Für $f(x_1, x_2) = x_1^2 - x_2^2$ ist

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1, \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = -2x_2$$

und wir erhalten die lineare Approximation in $\mathbf{x} = (2, 3)^T$ als

$$f(x_1 + u_1, x_2 + u_2) \approx -5 + 4u_1 - 6u_2$$

z.B.

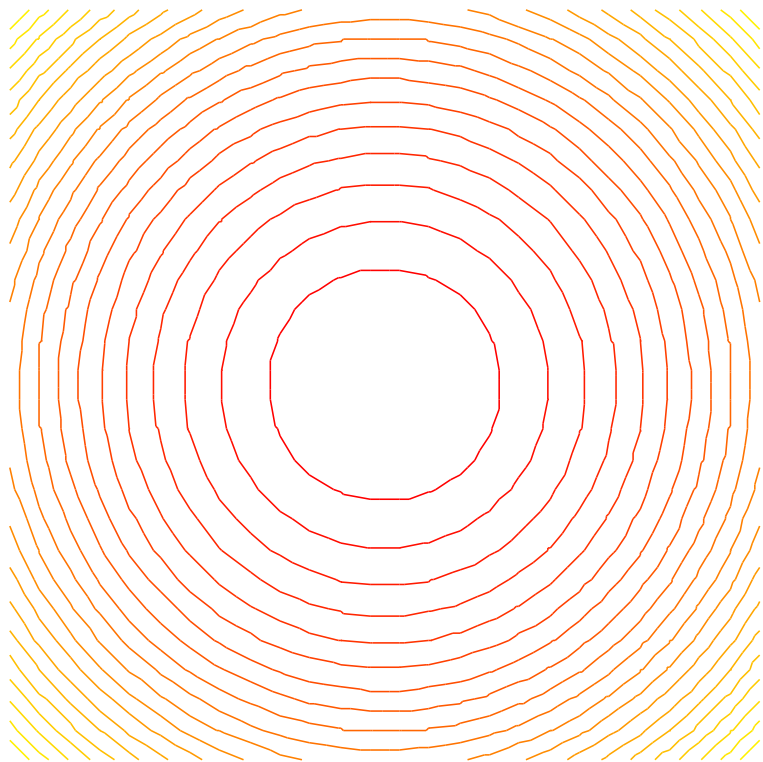
$$\begin{aligned} f(2.02, 2.99) &= -4.8597 \\ &\approx -5 + 4 \cdot 0.02 - 6 \cdot (-0.01) = -4.86 \end{aligned}$$

Der Gradient.

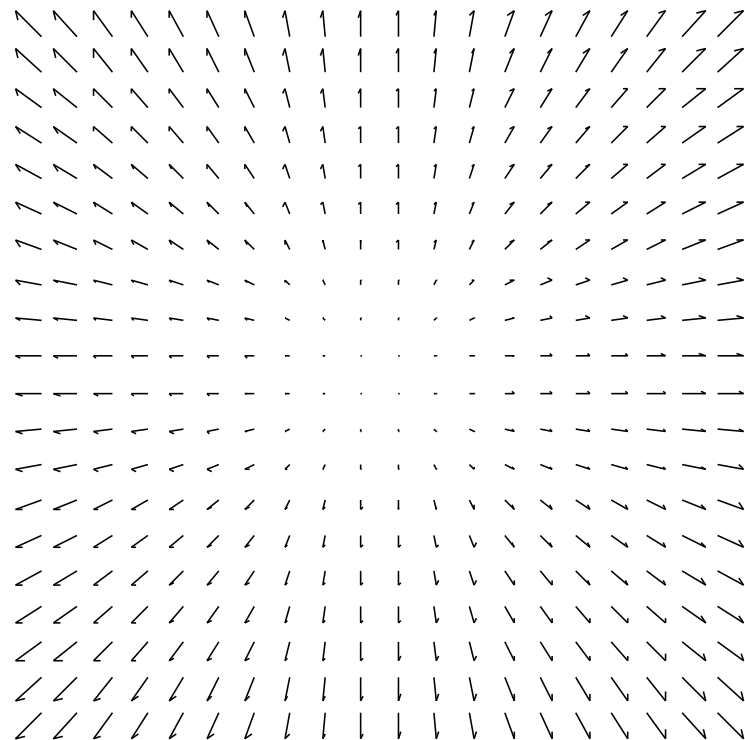
Die lineare Approximation benutzt den Zeilenvektor $\left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x})\right)$.
Durch Transposition erhält man den *Gradienten* von f :

$$\text{grad} f = \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix}$$

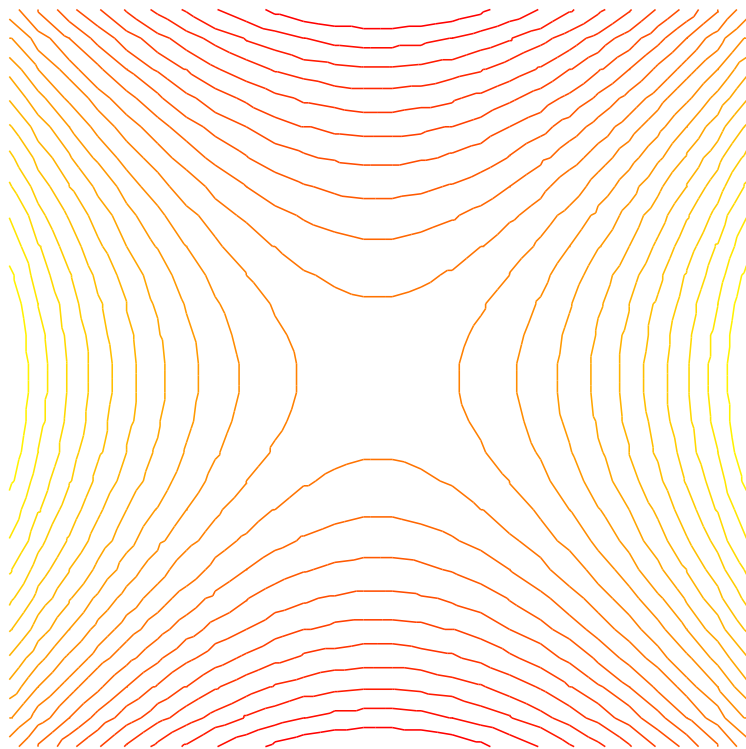
$\nabla f(\mathbf{x})$ ist für jedes $\mathbf{x}$ ein Vektor, ein Element des $\mathbb{R}^2$.



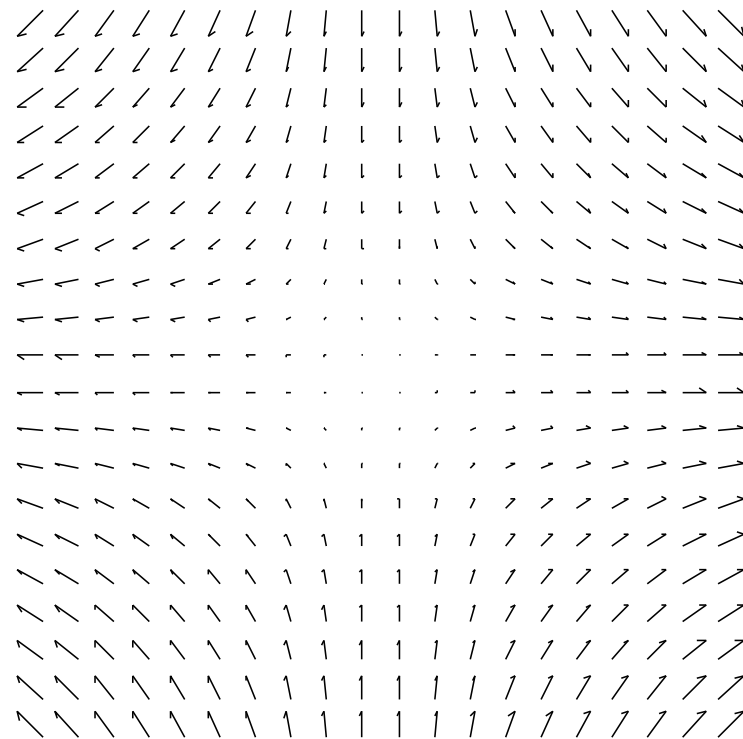
$$f(x_1, x_2) = x_1^2 + x_2^2$$



$$\nabla f(x_1, x_2) = (2x_1, 2x_2)^T$$



$$f(x_1, x_2) = x_1^2 - x_2^2$$

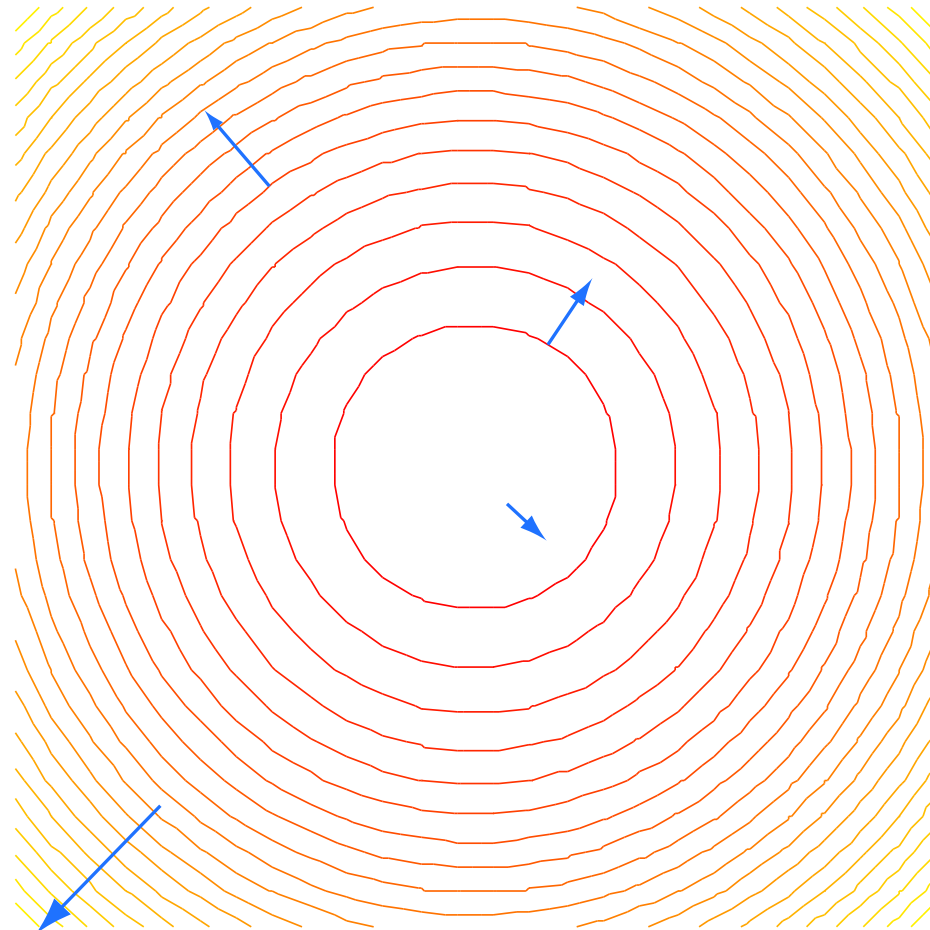


$$\nabla f(x_1, x_2) = (2x_1, -2x_2)^T$$

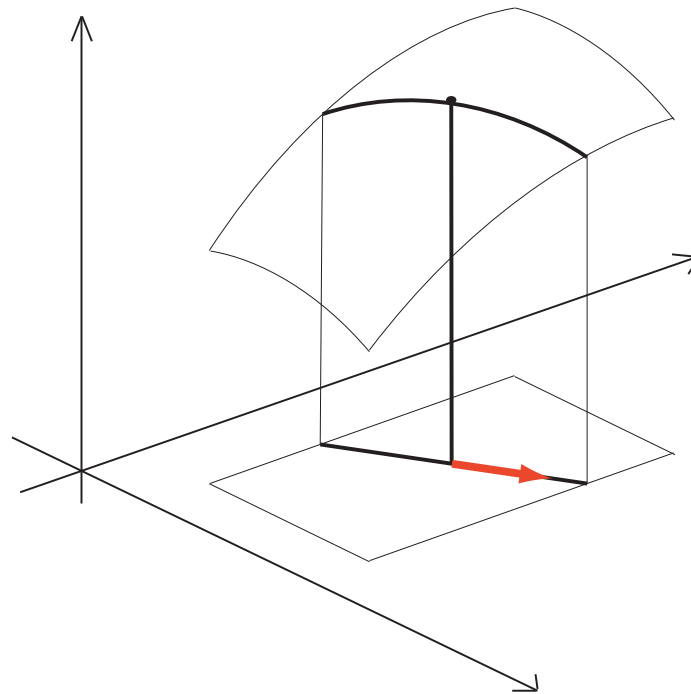
Der Gradient $\nabla f(\mathbf{x})$ gibt die Richtung und die Größe der steilsten Steigung der Funktion f an der Stelle $\mathbf{x}$ an.

Beispiel.

Für $f(x_1, x_2) = x_1^2 + x_2^2$ gilt $\nabla f(x_1, x_2) = (2x_1, 2x_2)^T$. Die Richtung ist weg vom Ursprung, und die Größe ist $2\sqrt{x_1^2 + x_2^2}$.



Zum Beweis berechnen wir die *Richtungsableitung* in die Richtung $\mathbf{b}$, mit einem Vektor $\mathbf{b} = (b_1, b_2)^T \in \mathbb{R}^2$ der Länge 1:



Für reelle Zahlen $h > 0$ gilt aufgrund der totalen Differenzierbarkeit mit $\mathbf{u} = h\mathbf{b}$

$$f(\mathbf{x} + h\mathbf{b}) = f(\mathbf{x}) + h\ell(\mathbf{b}) + o(h) \quad \text{für } h \rightarrow 0$$

Die „Steigung in Richtung $\mathbf{b}$ “ ist

$$\ell(\mathbf{b}) = a_1 b_1 + a_2 b_2 = \frac{\partial f}{\partial x_1}(\mathbf{x})b_1 + \frac{\partial f}{\partial x_2}(\mathbf{x})b_2$$

und mit dem Gradienten ausgedrückt ist die Steigung in Richtung $\mathbf{b}$ gleich

$$\ell(\mathbf{b}) = \frac{\partial f}{\partial x_1}(\mathbf{x})b_1 + \frac{\partial f}{\partial x_2}(\mathbf{x})b_2 = \langle \nabla f(\mathbf{x}), \mathbf{b} \rangle$$

also

$$f(\mathbf{x} + h\mathbf{b}) = f(\mathbf{x}) + h\langle \nabla f(\mathbf{x}), \mathbf{b} \rangle + o(h) \quad \text{für } h \rightarrow 0$$

Wann ist die Steigung

$$\langle \nabla f(\mathbf{x}), \mathbf{b} \rangle$$

maximal?

Nach der Cauchy-Schwarz-Ungleichung gilt für diese Richtungssteigung wegen $|\mathbf{b}| = 1$

$$\langle \nabla f(\mathbf{x}), \mathbf{b} \rangle \leq |\nabla f(\mathbf{x})| \cdot |\mathbf{b}| = |\nabla f(\mathbf{x})|$$

und speziell für den Vektor $\mathbf{b} = \frac{1}{|\nabla f(\mathbf{x})|} \nabla f(\mathbf{x})$ in Richtung des Gradienten der Länge 1 ist dieser Ausdruck maximal:

$$\langle \nabla f(\mathbf{x}), \mathbf{b} \rangle = \frac{1}{|\nabla f(\mathbf{x})|} \langle \nabla f(\mathbf{x}), \nabla f(\mathbf{x}) \rangle = |\nabla f(\mathbf{x})|$$

Also:

In Richtung des Vektors $\nabla f(\mathbf{x})$ steigt die Funktion f im Punkt $\mathbf{x}$ am stärksten an,

und diese Steigung hat die Größe

$$|\nabla f(\mathbf{x})|$$

In stationären Punkten verschwindet der Gradient.

Extrema unter Nebenbedingungen.

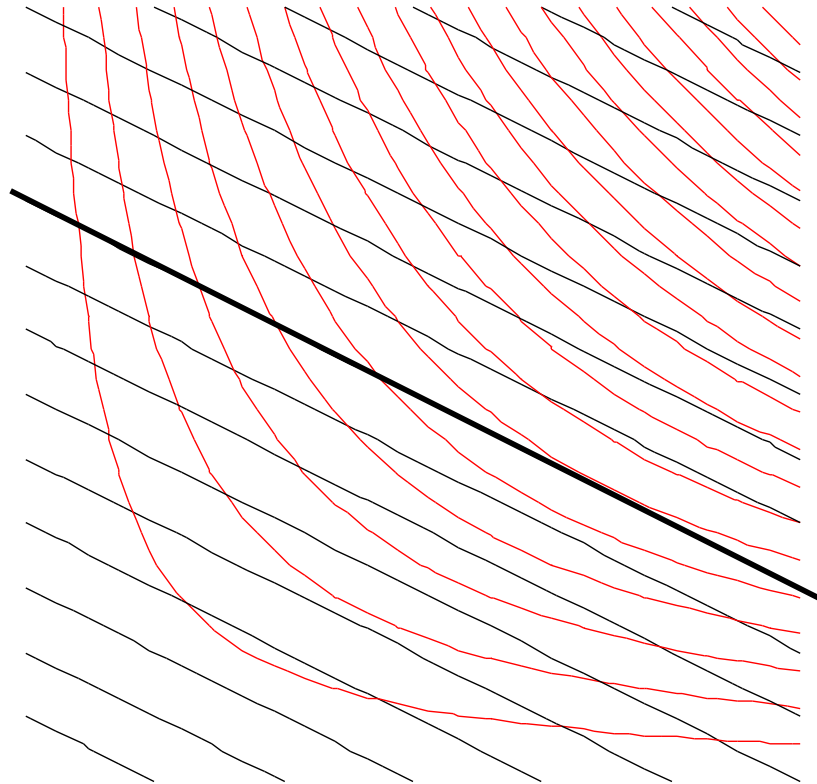
Aufgabe: Maximiere

$$f(x_1, x_2) = x_1 x_2$$

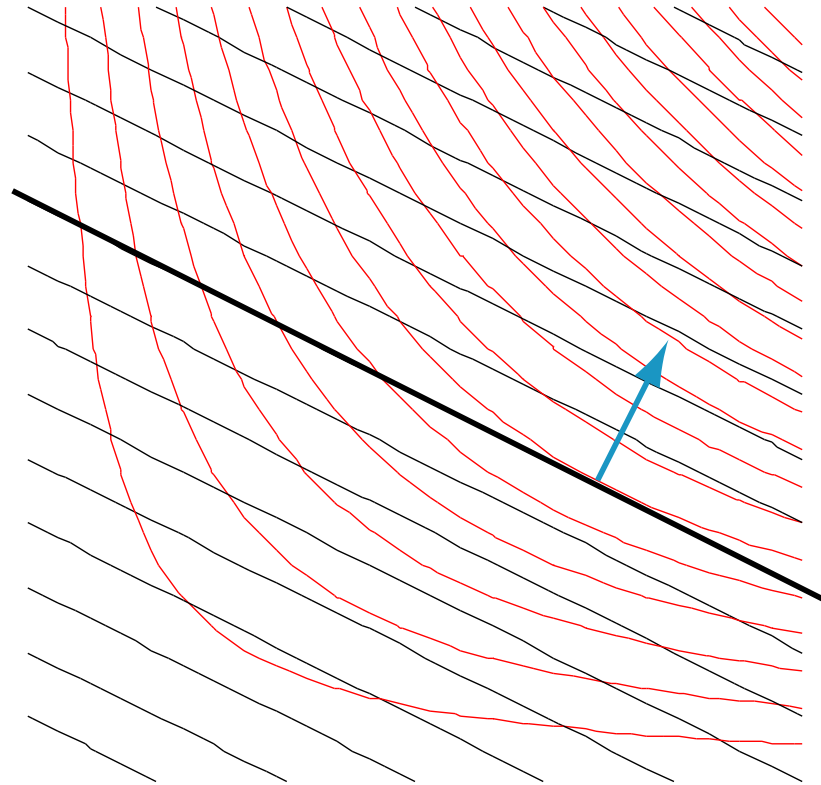
unter der Nebenbedingung

$$g(x_1, x_2) = x_1 + 2x_2 = 5$$

Höhenlinien von f und g :



An der Stelle $(x_1, x_2)^T$ des Extremums zeigen ∇f und ∇g offenbar in dieselbe (oder entgegengesetzte) Richtung:



$$\nabla f(x_1, x_2) = \lambda \nabla g(x_1, x_2)$$

$-\lambda$ heißt *Lagrange-Multiplikator* (das Minuszeichen ist Konvention).

Es gilt hier

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \quad \nabla g(x_1, x_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

und die Gleichung

$$\nabla f(x_1, x_2) = \lambda \nabla g(x_1, x_2)$$

geht über in $x_2 = \lambda, x_1 = 2\lambda$, die zusammen mit der Nebenbedingung $x_1 + 2x_2 = 5$

$$\lambda = 1, \quad x_1 = 2, \quad x_2 = 1$$

ergeben. Das Maximum wird im Punkt $(x_1, x_2)^T = (1, 1)^T$ angenommen und hat den Wert 1.

Definition:

Eine Funktion

$$f : D \rightarrow \mathbb{R}^m \quad \text{mit } D \subset \mathbb{R}^n$$

heißt im Punkt $\mathbf{x} \in D$ *total differenzierbar*, falls es eine (im Allgemeinen von $\mathbf{x}$ abhängige) lineare Abbildung $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ gibt, so dass gilt

$$f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + L(\mathbf{u}) + o(|\mathbf{u}|) \quad \text{für } \mathbf{u} \rightarrow 0$$

Man nennt L das *Differential* von f und schreibt $L = Df(\mathbf{x})$.

Zerlegen wir die Funktion f in ihre Koordinaten gemäß

$$f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T$$

so lassen sich die partiellen Ableitungen $\partial f_i / \partial x_j$, $1 \leq i \leq m$, $1 \leq j \leq n$ bilden. Die lineare Abbildung L ist dann durch die Matrix

$$J_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

gegeben, die *Jacobimatrix* von f .

Kettenregel.

Ist für die Funktion

$$h = g \circ f$$

sowohl f im Punkt $\mathbf{x}$ als auch g im Punkt $\mathbf{y} = f(\mathbf{x})$ total differenzierbar, so ist auch h in $\mathbf{x}$ total differenzierbar. Die Verkettungseigenschaft überträgt sich auf das Differential:

$$Dh(\mathbf{x}) = Dg(\mathbf{y}) \circ Df(\mathbf{x})$$

bzw. auf die Jacobimatrix

$$J_{g \circ f}(\mathbf{x}) = J_g(\mathbf{y}) J_f(\mathbf{x})$$

Ist $m = 1$, also f eine Abbildung in die reellen Zahlen, so kann man wieder den Gradienten

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

bilden. Er hat analoge Eigenschaften zum Fall $n = 2$.