

8. Funktionen mehrerer Veränderlicher

*Das Ziel des Rechnens ist Einsicht, nicht Zahlen.
[Richard Hamming, 1915-1998]*

8.1. Worum geht es?

Bisher hatten wir bei der Differentiation nur Funktionen einer Veränderlichen betrachtet. Bei den meisten Problemen der realen Welt treten aber mehrere Veränderliche auf:

- Eine Fläche in der Computergrafik kann durch $Z = f(x, y)$ beschrieben werden
- Der Gewinn eines Unternehmens ist eine Funktion der Umsätze aller seiner n Produkte und m Kostenstellen: $G = G(u_1, u_2, \dots, u_n, k_1, \dots, k_m)$
- Der Druck p ist eine Funktion von Temperatur T und Volumen V : $p = p(T, V) = \frac{RT}{V}$.

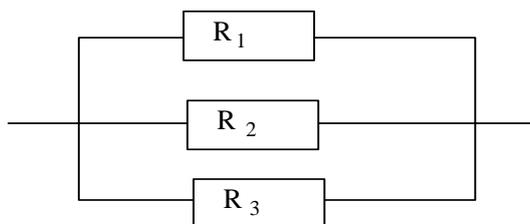
Wir werden uns in diesem Kapitel mit der Definition und der Differentiation solcher Funktionen beschäftigen. Damit können wir dann folgende Probleme und Anwendungen lösen:

- Wie stellt man Funktionen mehrerer Veränderlicher dar?
- Optimierung: Wie findet man **Extremwerte**? Anwendungsfall: Wie findet man die **Regressionsgerade** $y = ax + b$ für eine Menge von Punkten (x_i, y_i) ?
- Wie differenziert man eine Funktion mehrerer Veränderlicher (**partielle** Differentiation und **totales Differential**)?
- Optimierung mit Nebenbedingungen: Die Methode der **Lagrange-Multiplikatoren**.

Da man bei den meisten Realwelt-Optimierungsaufgaben an mehreren (vielen) "Stellschrauben" drehen kann, sind solche Probleme von großer praktischer Bedeutung.

8.2. Definition einer Funktion mehrerer Veränderlicher

Wir hatten bereits in Abschnitt 7.5 die Multiplikation eines Vektors mit einer Matrix als spezielle, nämlich lineare, Abbildung eines Vektors \mathbf{x} auf einen Vektor \mathbf{y} kennengelernt. Wir wollen im folgenden beliebige Abbildungen von Vektoren auf Zahlen definieren und ihre Eigenschaften untersuchen. Solche Fragestellungen kommen in der Praxis häufig vor. Z.B. ergibt sich der Gesamtwiderstand eines Netzes ohmscher Widerstände als (nicht unbedingt lineare) Funktion aller im Netz auftretenden Widerstände. Dies wird schon an folgendem sehr einfachen Beispiel klar:



Bekanntlich gilt für den Gesamtwiderstand einer Parallelschaltung: $R_{\text{ges}} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}}$

Der Gesamtwiderstand ist also eine Funktion aller Einzelwiderstände oder anders gesprochen, eine Funktion des Vektors der Einzelwiderstände, d.h.:

$$R_{\text{ges}} = f(R_1, R_2, R_3) = f(\vec{r}) \text{ mit } \vec{r} = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$$

Jede konkrete Ausprägung der 3 Einzelwiderstände bildet also einen Vektor im 3-dimensionalen Raum. Die Koordinaten dieser Widerstandsvektoren sind dabei keine Längen, wie im Ortsraum, sondern eben Widerstände. Ansonsten verhalten sich diese Widerstandsvektoren mathematisch genau wie Ortsvektoren. Da Widerstandsnetze aus beliebig vielen Komponenten bestehen können, ist klar, dass in solchen Fällen Vektoren mit beliebig vielen Komponenten auftreten können.

Ein anderes Beispiel ist der Gewinn eines Unternehmens, der vom Umsatz sämtlicher n Produkte, die das Unternehmen vermarktet, abhängig ist. Die Umsatzzahlen aller Produkte des letzten Monats bilden dabei einen Vektor im n -dimensionalen kartesischen Raum, dessen i -te Achse für den monatlichen Umsatz des i -ten Produktes steht.

Es macht also mathematisch durchaus Sinn, sich mit Vektoren mit beliebig vielen Komponenten zu beschäftigen, auch wenn unsere Anschauung auf 3-dimensionale Räume beschränkt ist. Wir definieren den n -dimensionalen Raum \mathbf{R}^n wie in Mathe 1 (Kap. 7.4 „Vektoren“):

Def D 8-1 n -dimensionaler Raum

Jedes Element der Menge \mathbf{R}^n wird als Punkt eines n -dimensionalen Vektorraumes \mathbf{R}^n bezeichnet. In der Regel wird ein solcher Punkt durch den Vektor \vec{x} bezeichnet.

Def D 8-2 reellwertige Funktion mehrerer Veränderlicher

Eine reellwertige Funktion f ordnet jedem Punkt x_1, \dots, x_n (bzw. Vektor $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$) aus einer zusammenhängenden Teilmenge D des \mathbf{R}^n eindeutig einen reellen Wert $y \in \mathbf{R}$ zu, und man schreibt:

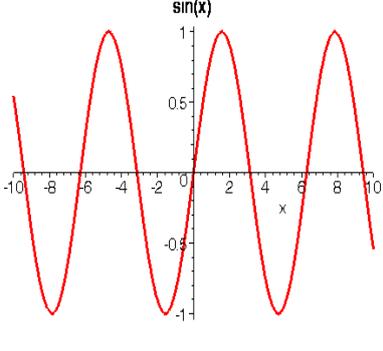
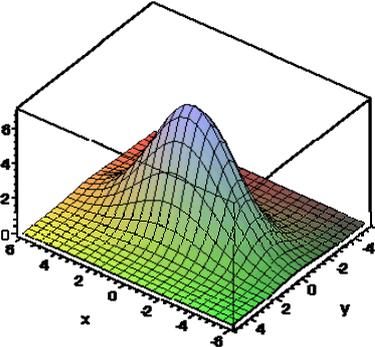
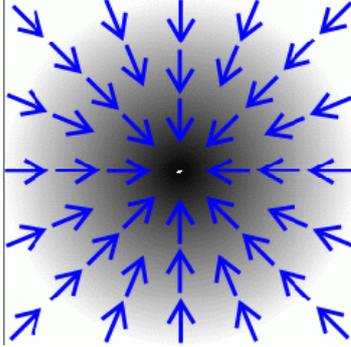
$$f : D \subseteq \mathbf{R}^n \rightarrow \mathbf{R} \quad \text{mit} \quad y = f(x_1, x_2, \dots, x_n)$$

Beispiel:

Die Temperatur auf der Erde ist eine Funktion der Längen- und Breitenkoordinate sowie der Höhe über dem Erdboden.

ANMERKUNG: Wir beschäftigen uns hier also mit **reellwertigen** Funktionen $f : \mathbf{R}^n \rightarrow \mathbf{R}$. In

Kapitel 8.6 werden wir noch kurz auf **vektorwertige** Funktionen $\vec{f} : \mathbf{R}^n \rightarrow \mathbf{R}^m$ eingehen, die einen n -dim. Vektor auf einen m -dim. Vektor abbilden. Beispiele:

„normale“ Funktion	reellwertige Funktion	vektorwertige Funktion
$f : \mathbb{R} \rightarrow \mathbb{R}$	$f : \mathbb{R}^2 \rightarrow \mathbb{R}$	$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$
		
„Kurve“	(Hyper-) „Fläche“	„Pfeile“ (Windkarte)

8.3. Darstellung einer Funktion mehrerer Veränderlicher

[Papula, Bd. 2, S. 272-286]

Dieses anschauliche Kapitel wird im 1. Maple-Praktikumsblock (V Th. Bartz-Beielstein) genauer behandelt, da sich hier mit Maple vieles anschaulich darstellen lässt.

Wir fragen uns hier nur, welche Darstellungsformen grundsätzlich in Frage kommen und gehen auf „**Fläche im Raum**“ kurz ein.

Mehr zu diesem Gebiet, der sog. **Visualisierung** (von Funktionen), können Sie auch im WPF „Computergrafik und Visualistik“ von Horst Stenzel erfahren.

[nur Kap. 8.3.3 in Vorlesung, Rest in V TBB >> weiter bei Kap. 8.4]

8.3.1. Analytische Darstellung

Darstellung in Form einer Gleichung

- explizite Form: $z = f(x,y)$
- implizite Form: $F(x,y,z) = 0$

Beispiele in Vorlesung.

Man verwendet die implizite Form, wenn eine Auflösung nach einer Variablen nicht möglich ist,¹ oder, wenn sie zwar prinzipiell möglich, aber zu aufwendig oder mit unnötigen Schwierigkeiten verbunden ist.

Anmerkung: Jede explizite Form lässt sich mit

$$F(x,y,z) = f(x,y) - z$$

in die "kanonische" implizite Form bringen. Die umgekehrte Richtung kann dagegen schwierig sein.

Zum Spielen und für „schöne Forme(l)n“ ist der ZEIT.de-Skulpturenwettbewerb **wärmstens** empfohlen !!

[Programme – Surfer zeigen, z.B. mit $(x^2+y^2+z^2-1)*(x^3+y^3+z^3-1)$]

¹ Die implizite Form kann komplizierte Flächen im \mathbb{R}^3 darstellen, die explizite Form „kann“ nur solche Flächen, die jedem (x,y) höchstens ein z zuordnen („Funktionsgebirge ohne überhängende Klippen“).

8.3.2. Tabellarische Darstellung

Bevorzugte Darstellung für Tabellenkalkulationsprogramme

$$z = f(x,y)$$

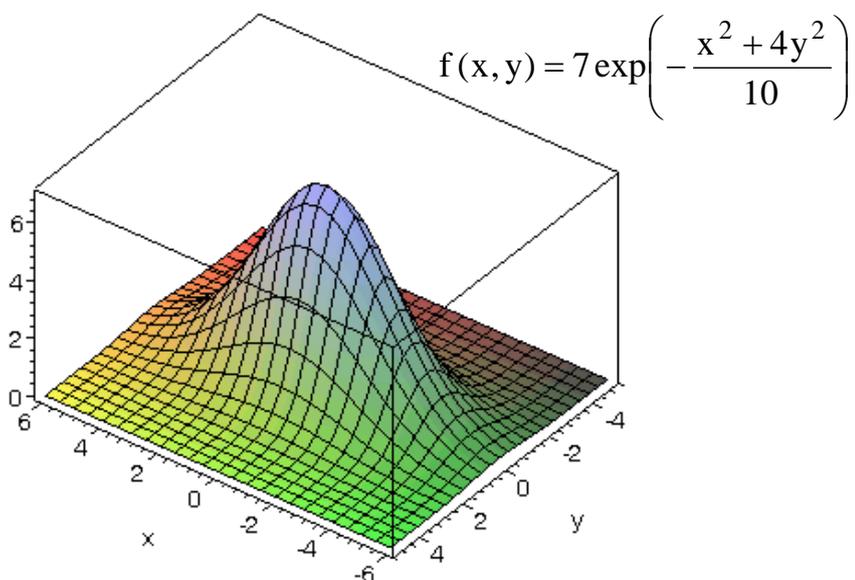
	y_1	y_2	...	y_k	...	y_n
x_1	z_{11}	z_{12}	...	z_{1k}	...	z_{1n}
...
x_m	z_{m1}	z_{m2}	...	z_{mk}	...	z_{mn}

Darstellung von Funktionen mit mehr als zwei Veränderlichen möglich über verschiedene Seiten mit Tabellen (z.B. Blätter in Excel).

8.3.3. Fläche im Raum

Bevorzugte Darstellung in Maple (**plot3d**)

Beispiel "Gaussglocke":



8.3.4. Schnittkurven: Höhenlinien, Kennlinienfeld

Alternative Darstellungen mit Maple:

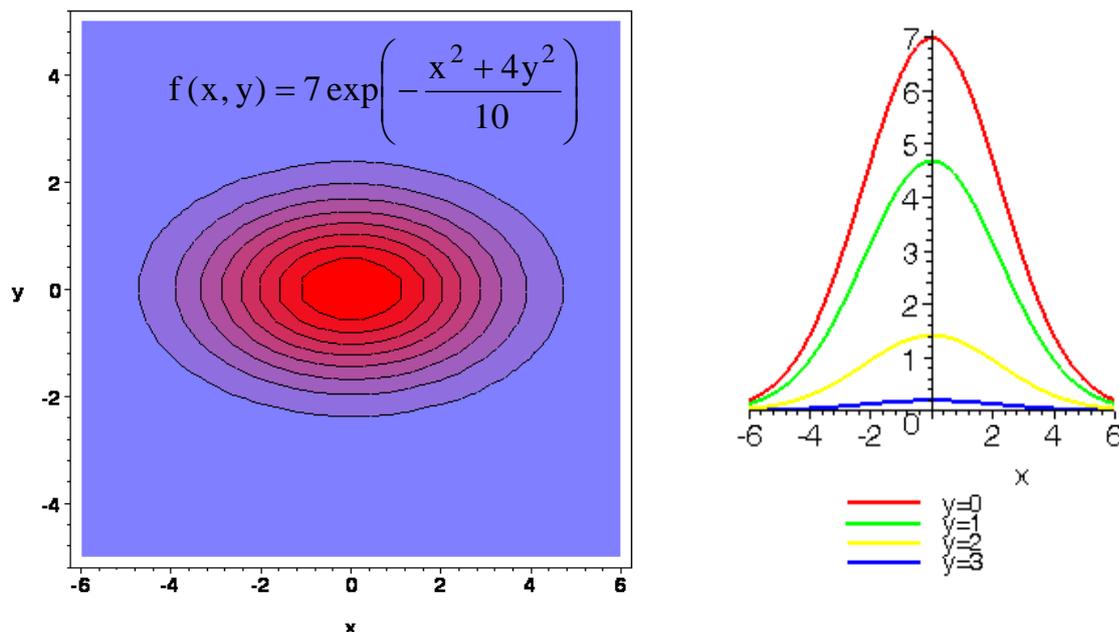


Abbildung 8-1: (a) Höhenliniendiagramm, (b) Kennlinienfeld²

Die Verallgemeinerung des Begriffs der Höhenlinie für mehr als zwei Dimensionen ist die **Äquipotentialfläche** (d.h. die Fläche³ mit $f(\mathbf{x}) = c = \text{const}$).

Wie findet man die Höhenlinien für eine explizite Form? – Indem man die linke Seite als konstant festsetzt und nach y auflöst. Im Beispiel:⁴

$$f(x, y) = z = 7 \exp\left(-\frac{x^2 + 4y^2}{10}\right) \Leftrightarrow \ln \frac{z}{7} = -\frac{x^2 + 4y^2}{10} \Leftrightarrow y = \pm \frac{1}{2} \sqrt{-10 \ln \frac{z}{7} - x^2}$$

Wenn sich die Gleichung nicht analytisch nach y auflösen lässt, geht es nur mühsamer: Numerisch ein Raster vieler Funktionswerte bestimmen und Punkte mit gleichen Werten verbinden. Oder durch numerische Nullstellenbestimmung.

Ein Kennlinienfeld lässt sich dagegen für die explizite Form immer leicht zeichnen: einfach verschiedene feste Werte für y einsetzen.

Übung: Leider ist gerade Ihr Laptop kaputt und Sie haben kein Maple zur Hand. Machen Sie sich trotzdem ein Bild von der Funktion $f(x, y) = x^2 e^y$, indem Sie handschriftlich ein Höhenliniendiagramm im Bereich 1,2,4,8 und ein Kennlinienfeld für $y=0.5, 1, 2$ erstellen.

² Erzeugt durch folgende Maple-Befehle:

```
(a) g := (x, y) -> 7 * exp(-(x^2 + 4*y^2) / 10);
contourplot(g(x, y), x = -6..6, y = -5..5, filled = true, axes = boxed,
coloring = [COLOR(RGB, 0.5, 0.5, 1), red], font = [HELVETICA, BOLD, 12]);
(b) plot([seq(g(x, y), y = 0..3)], x = -6..6, legend = ["y=0", "y=1", "y=2",
"y=3"], font = [HELVETICA, 12], thickness = 2);
```

³ bzw. Hyperfläche für mehr als 3 Veränderliche

⁴ Unter der Wurzel steht tatsächlich nichts Negatives: $\ln(z/7) < 0 \Rightarrow -10 \ln(z/7) > 0$. Weiter $x^2 < -10 \ln(z/7)$.

Weitere Beispiele in Übungen!

8.3.5. Mehr als zwei Veränderliche

Die Anschauung versagt, die Funktion läßt sich nicht mehr als Ganzes zu erfassen. Zahlreiche Techniken sind entwickelt worden, um sich dennoch ein Bild von der Lage zu machen; Stichwort "Visualisierung von Daten". Basis-Methoden:

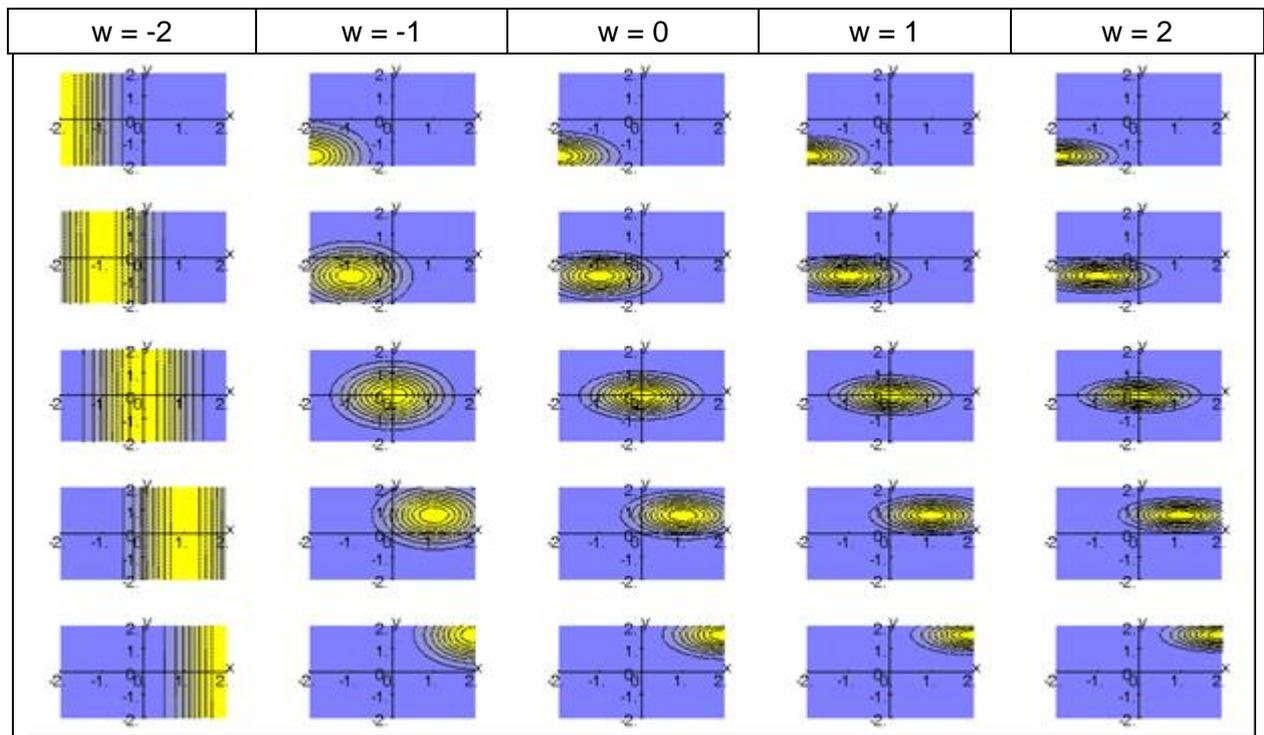
- Festhalten von $n-2$ Parametern und Betrachtung eines Schnitts, z.B. Höhenliniendiagramm in den restlichen beiden Parametern
- Anordnung vieler solcher Schnitte in rechteckigem Plot-Feld
- Animation, d.h. einer oder mehreren Variablen wird ein zeitlicher Verlauf zugeordnet, und man beobachtet die Änderung, die sich im Bild der anderen Variablen als Funktion der Zeit ergibt.
- u.v.a.m.

Beispiel 1: Anordnung in rechteckigem Plot-Feld:

Sei $f: \mathbf{R}^4 \rightarrow \mathbf{R}$ eine Funktion von 4 Veränderlichen X, Y, V, W :

$$f(x, y, v, w) = \exp\left(- (x - v)^2 - (w + 2)(y - 0.8v)^2\right)$$

Wir stellen f durch ein Array von X - Y -Höhenliniendiagrammen dar, in den Reihen läuft V von -2 bis 2, in den Spalten läuft W von -2 bis 2:



Welche Wirkung hat also der Parameter w , welche der Parameter v ?

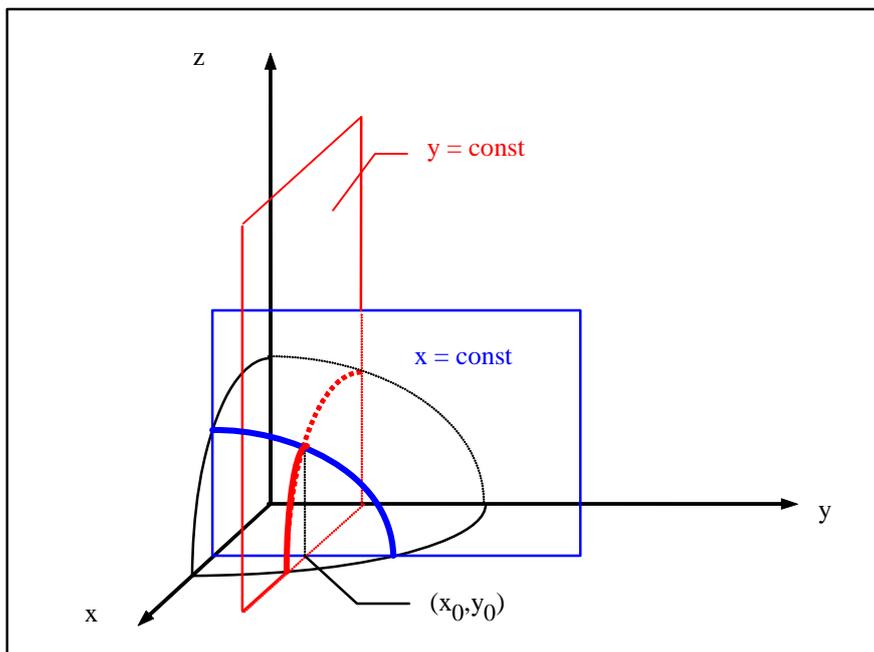
Beispiel 2: Wir stellen die gleiche Funktion $f(x,y,v,w)$ als Animation dar, wobei der Animationspfad längs der Diagonalen im V - W -Raum läuft, also von $v=w=-2$ bis $v=w=0.5$.

Überlegen Sie: Wie wird die Animation in etwa aussehen? **[Fragend entwickeln]**

Lösung: s. plot3d.mws, Animation in Abschnitt "Mehr als zwei Veränderliche".

8.4. Partielle Ableitungen

Wie schon bei Funktionen einer Veränderlichen liefert der Begriff der Ableitung auch bei Funktionen mehrerer Veränderlichen den Schlüssel zur Analyse von Zusammenhängen. Die Ableitung einer Funktion mehrerer Veränderlicher wird mittels partieller Ableitungen auf den Fall eindimensionaler Funktionen zurückgeführt. Betrachten wir die Situation zunächst bei Funktionen zweier Veränderlicher (Skizze).



Im Punkt (x_0, y_0) sind die Schnittebenen $x = \text{const}$ und $y = \text{const}$ eingezeichnet. Innerhalb der jeweiligen Schnittebene liegt dann nur noch eine Funktion $z = f(x)$ (für $y = \text{const}$) bzw. $z = g(y)$ (für $x = \text{const}$) vor. Insbesondere bereitet die Bildung der Ableitung in diesen Fällen keine Schwierigkeiten. Dies führt uns zum Begriff der partiellen Ableitung.

Def D 8-3 Partielle Ableitung

Die partielle Ableitung 1. Ordnung der Funktion

$$y = f(x_1, x_2, \dots, x_n)$$

nach der Variablen x_i ist durch den folgenden Grenzwert definiert:

$$\frac{\partial y}{\partial x_i}(\bar{x}) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)}{h}$$

Umgangssprachlich bedeutet dieser Grenzwert: Betrachte alle Variablen mit Ausnahme von x_i als Konstanten und bilde die übliche Ableitung nach der Variablen x_i .

Anschaulich: Setze $n-1$ Variablen fest, dann passt die verbleibende Variable in eine „Schau-
tafel“ (rotes oder blaues Rechteck in obiger Zeichnung), d.h. einen Graphen für eine „normale“
Funktion, den wir wie üblich ableiten können).

Weitere, allgemein übliche Symbole für partielle Ableitungen sind

$$\frac{\partial y}{\partial x_i}(\vec{x}) = y_{x_i}(\vec{x}) = \frac{\partial f}{\partial x_i}(\vec{x}) = f_{x_i}(\vec{x})$$

Wir werden im Folgenden meist die Schreibweise $f_{x_i}(\mathbf{x})$ benutzen, wenn keine Verwechslung mit dem Index (einer Vektorfunktion) zu befürchten ist.

Beispiel:

Die Zustandsgleichung eines idealen Gases lautet:

$$p(V, T) = \frac{RT}{V}$$

$$\frac{\partial p}{\partial V} = p_V = -\frac{RT}{V^2}$$

$$\frac{\partial p}{\partial T} = p_T = \frac{R}{V}$$

Anschaulich: Wenn ich das Volumen um einen kleinen Wert ∂V ändere, dann ändert sich

der Druck um $\partial p = -\frac{RT}{V^2} \partial V$. D.h. bei Volumenvergrößerung sinkt der Druck, weil

$-\frac{RT}{V^2} < 0$ (wenn man bei einer geschlossenen Luftpumpe den Kolben nach aussen zieht,

gibt es eine rückziehende Kraft nach innen, weil der Druck innen niedriger ist als aussen),
bei Temperaturerhöhung steigt der Druck.

Übung: Für $z(x, y) = 5xy - 2y + 3$ bestimme man z_x und z_y

Für $y(x_1, x_2, x_3) = x_1^2 x_2 \ln x_3 + \sqrt{x_1} \sin x_2 + \frac{e^{x_3}}{x_1}$ bestimme man y_{x_1}, y_{x_2} und y_{x_3}

Wie diese Beispiele zeigen, sind die partiellen Ableitungen im Allgemeinen selbst wieder Funktionen sämtlicher, in der Ausgangsfunktion auftretender, Veränderlicher.

Sind alle partiellen Ableitungen stetig, so heißt die Funktion stetig differenzierbar.

Def D 8-4 Stetig differenzierbar

Ist eine Funktion an allen Stellen eines Gebietes G (einmal) differenzierbar und sind die partiellen Ableitungen stetig, so heißt die Funktion im Gebiet (einmal) stetig differenzierbar. Analog: n -mal stetig differenzierbare Funktionen.

Die besondere Bedeutung dieser Definition liegt darin, dass stetig differenzierbare Funktionen in einer (kleinen) Umgebung eines Punktes durch den Funktionswert in diesem Punkt

und sämtliche partiellen Ableitungen angenähert (approximiert) werden können (s. Kap. 8.8.3 "Linearisierung einer Funktion").

Def D 8-5 Partielle Ableitungen 2. Ordnung

Ist eine Funktion 2mal stetig differenzierbar, so kann jede partielle Ableitung 1. Ordnung selbst wieder nach allen Variablen differenziert werden. Hierdurch entstehen partielle Ableitungen 2. Ordnung.

Beispiel: Zu $y(x_1, x_2, \dots)$ ist eine Ableitung 2. Ordnung $y_{x_1 x_2} = (y_{x_1})_{x_2}$

Analog: Partielle Ableitungen n. Ordnung.

Ü

Übung: Bilden Sie $y(x_1, x_2, x_3) = x_1^2 x_2 \ln x_3 + \sqrt{x_1} \sin x_2 + \frac{e^{x_3}}{x_1}$ (unter Verwendung

der Ergebnisse y_{x_1}, y_{x_2} und y_{x_3} aus voriger Übung) die 2. Ableitungen

$y_{x_1 x_2}$ und $y_{x_2 x_1}$

Satz S 8-1 Satz von Schwarz

Ist eine Funktion von mehreren Veränderlichen k-mal stetig differenzierbar, so sind die gemischten Ableitungen k-ter Ordnung unabhängig von der Reihenfolge des Differenzierens.

Wie wir gerade gesehen haben, gilt für $k = 2$ für die Funktion $u(x, y, \dots)$:

$$u_{xy} = (u_x)_y = (u_y)_x = u_{yx}$$

Ü

Übung: Überprüfen Sie an der Funktion $f(x, y, z) = \frac{e^{ax} \cos by}{zx}$ durch explizites Nach-

rechnen, dass gilt: $f_{xz} = f_{zx}$. Ist eine der Reihenfolgen ökonomischer?

8.5. Extremwerte

8.5.1. Lokale und globale Extremwerte

[Stingl, S. 361]

Analog zur Situation bei Funktionen mit einer Veränderlichen, lassen sich auch bei Funktionen mehrerer Veränderlicher die Begriffe lokales Minimum oder Maximum definieren. Notwendige Bedingungen ergeben sich aus den partiellen Ableitungen.

Def D 8-6 Relatives Minimum, relatives Maximum

Eine Funktion $y = f(x_1, x_2, \dots, x_n)$ besitzt im Punkt

$\vec{x}_0 = (x_{01}, x_{02}, \dots, x_{0n})$ ein relatives Minimum, wenn in einer Umgebung von \vec{x}_0 stets:

$$f(x_1, \dots, x_n) > f(x_{01}, \dots, x_{0n})$$

für alle $\vec{x} \neq \vec{x}_0$

gilt. Ein relatives Maximum liegt vor, falls in einer Umgebung stets:

$$f(x_1, \dots, x_n) < f(x_{01}, \dots, x_{0n})$$

für alle $\vec{x} \neq \vec{x}_0$ gilt.

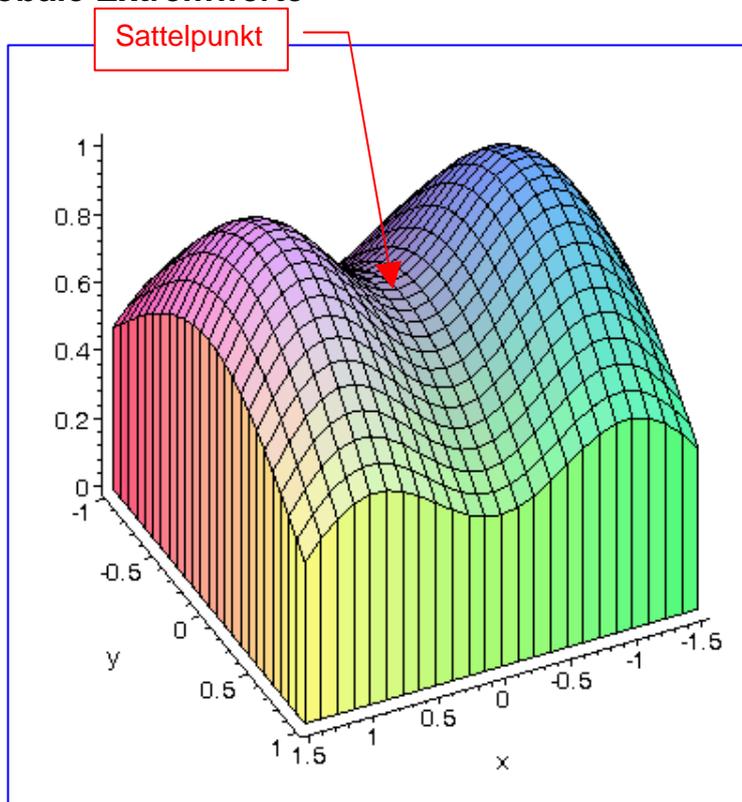
Ein Kriterium für das Vorliegen von Extremwerten liefert der nächste Satz:

Satz S 8-2 Stationärer Punkt

Ein Punkt \vec{x}_0 in dem sämtliche partiellen Ableitungen 1. Ordnung zu Null werden,

$f_{x_1}(\vec{x}_0) = f_{x_2}(\vec{x}_0) = \dots = f_{x_n}(\vec{x}_0) = 0$ heißt **stationärer Punkt**. Eine notwendige, aber im allgemeinen nicht hinreichende Bedingung für einen Extremstelle ist, dass sie ein stationärer Punkt ist.

Bemerkungen:



1. Bei zwei Veränderlichen folgt der Satz aus der Forderung, dass ein Extremwert eine waagerechte Tangentialebene haben muß.
2. Wie bei Funktionen einer Veränderlichen ist die Bedingung aus **Satz S 8-2** nicht hinreichend, auch **Sattelpunkte** können waagerechte Tangentialebenen haben. (Wie jeder weiß, der schon mal Bergsteigen war, muss es zwischen zwei Gipfeln eines stetigen Gebirges sogar Sattelpunkte geben.)

$$z = f(x, y) =$$

Beispiel (s. nebenstehendes Bild): $e^{-(x-1)^2 - y^2} / 2 + e^{-(x+1)^2 - y^2} / 2$

3. Die Angabe hinreichender Kriterien ist bei mehr als zwei Variablen schwierig. Für zwei Variablen erhält man als hinreichendes Kriterium:

Satz S 8-3 Hinreichendes Kriterium für lokale Extrema (2 Veränderliche)

Es sei $\Delta(x, y) = f_{xx}(x, y)f_{yy}(x, y) - [f_{xy}(x, y)]^2$ die Determinante der sog. Hesse-Matrix.

Eine Funktion $f(x, y) : D \rightarrow \mathbb{R}$ besitzt an der Stelle (x_0, y_0) mit Sicherheit ein lokales Extremum, wenn die folgenden Bedingungen zugleich erfüllt sind:

1. $f_x(x_0, y_0) = 0$ und $f_y(x_0, y_0) = 0$ stationärer Punkt, notwendige Bedingung

und

2. $\Delta(x_0, y_0) > 0$

Im Fall $f_{xx}(x_0, y_0) < 0$ liegt ein lokales Maximum, im Fall $f_{xx}(x_0, y_0) > 0$ ein lokales Minimum vor.

Ist $\Delta(x_0, y_0) < 0$, so liegt kein Extremwert, sondern ein **Sattelpunkt** vor.

Satz S 8-4 Hinreichendes Kriterium für **globale** Extrema (2 Veränderliche)

Eine Funktion $f(x, y) : D \rightarrow \mathbb{R}$ besitzt an einem **stationären Punkt** (x_0, y_0) mit Sicherheit ein globales Extremum, wenn gilt

1. $\Delta(x, y) > 0$ und $f_{xx}(x, y) < 0$ für **alle** $(x, y) \in D$ (globales Maximum)

– oder –

2. $\Delta(x, y) > 0$ und $f_{xx}(x, y) > 0$ für **alle** $(x, y) \in D$ (globales Minimum)

Beispiele und Übungen in Vorlesung!

--	--	--	--	--

Ü Übung 1: Bestimmen Sie die lokalen Extrema von $W(x, y) = 6x + 3y^2 - 0.1x^2 - \frac{1}{4}y^4$

Ü Übung 2: Gegeben sind n Punkte im zweidimensionalen Raum mit den Koordinaten $P_i = (x_i, y_i)$. Für welchen Punkt $P = (x, y)$ ist die Summe der Abstandsquadrate zu den gegebenen Punkten P_i minimal?

8.5.2. LS-Methode (Methode der kleinsten Quadrate)

Die LS-Methode ist eine der wichtigsten und gebräuchlichsten Methoden der Optimierung. LS steht für "least square"; oft findet man auch die dt. Abkürzung **KQ-Methode**.

Anwendung: Praktikum Physik bei Prof. Koch, z.B. Messungen zu Hall-Effekt oder Kondensator.

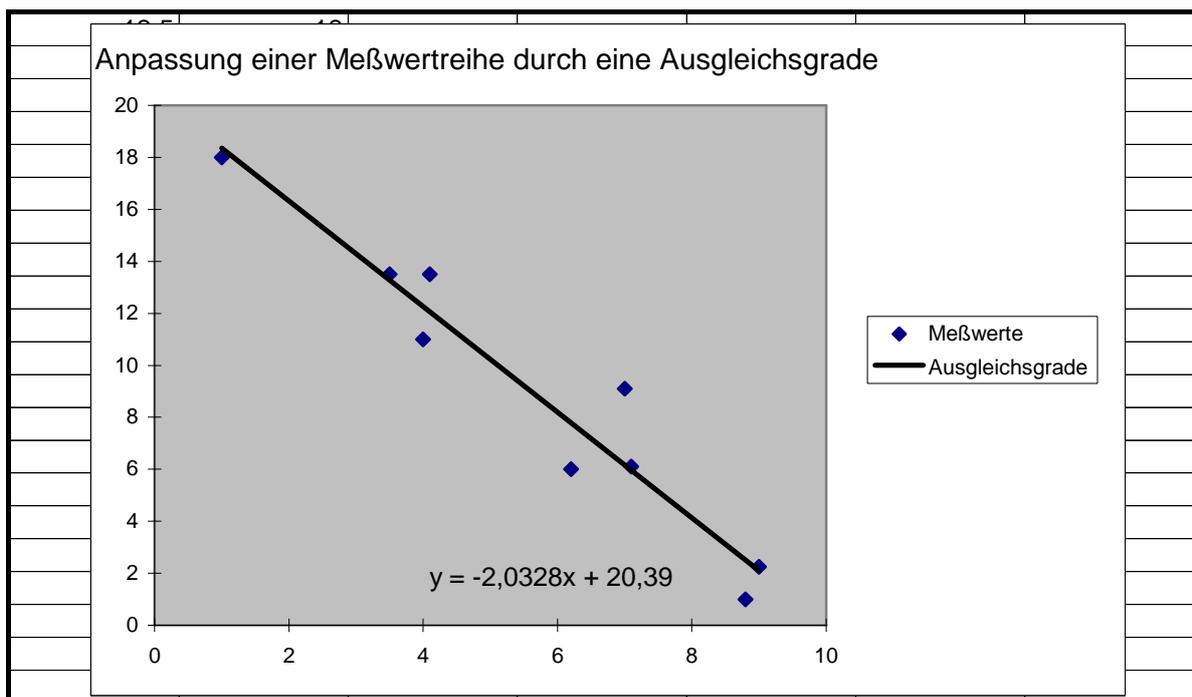
Gegeben sind n Meßpunkte (x_i, y_i) , die eigentlich auf einer Geraden liegen sollten, aber aufgrund von Meßfehlern etwas streuen. Gesucht sind nun die Koeffizienten **a** und **b** der Ausgleichsgeraden, d.h. der Geraden für die die Summe der quadratischen Abweichungen minimal wird:

Ausgleichsgerade (Regressionsgerade): $y = a + bx$

Abweichung der Ausgleichsgeraden beim i -ten Datenpunkt: $d_i = a + bx_i - y_i$

Wir setzen voraus, dass **nicht alle x_i identisch** sind, denn dann hätten wir eine senkrechte Gerade, die wir nicht als Funktion beschreiben können.

Zu minimierende Funktion: $Z(a, b) = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (a + bx_i - y_i)^2$ (Summe der Abweichungsquadrate, daher der Name "least squares" oder Methode der kleinsten Quadrate)



Wir setzen die partiellen Ableitungen gleich Null:

$$Z_a = 2 \sum_{i=1}^n (a + bx_i - y_i) = 0$$

$$Z_b = 2 \sum_{i=1}^n (a + bx_i - y_i)x_i = 0$$

Es ergibt sich ein lineares Gleichungssystem von zwei Gleichungen für die beiden Unbekannten a und b :

$$\begin{aligned} na + b \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i & \Leftrightarrow & \quad an + bS_x = S_y \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i & & \quad aS_x + bS_{xx} = S_{xy} \end{aligned}$$

wobei S_x, S_y, \dots einfach geeignete Abkürzungen für die Summen sind. Man multipliziert nun die 1. Gleichung mit S_x und die 2. Gleichung mit n durch, zieht voneinander ab und erhält:

$$\boxed{\begin{aligned} b &= \frac{nS_{xy} - S_x S_y}{nS_{xx} - (S_x)^2} \\ a &= \frac{S_{xx} S_y - S_x S_{xy}}{nS_{xx} - (S_x)^2} \end{aligned}}$$

Ü

Übung: (a) Theoretisch könnte ja der Nenner in den obigen Formeln für "pathologische" Kombinationen der x_i auch mal Null werden. Können Sie zeigen, dass der Nenner immer ungleich Null ist? Hinweis: Es gilt die nützliche Identität

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \quad \text{mit Mittelwert} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

(b) Weisen Sie nach, dass es sich bei der Lösung $\{a, b\}$ tatsächlich um ein Minimum handelt (s. **Satz S 8-3**)

Ü

Übung: Es muss nicht immer eine Gerade sein! Kombinationen von anderen "Basisfunktionen" gehen genauso gut.⁵ Beispiel: In einem Behälter sind radioaktive Stoffe vom Typ A, der

⁵ Den allgemeinen Fall beliebiger Basisfunktionen nennt man **GLS = "generalized least square"**.

proportional e^{-x} zerfällt und vom Typ B, der proportional e^{-2x} zerfällt. Durch Messungen soll ermittelt werden, wieviel vom Typ A, wieviel vom Typ B. Gegeben seien die Messpunkte:

x_i	0	1	2	3
y_i	4.1	1.3	0.4	0.3

Welches Modell $y = f(a,b) = ae^{-x} + be^{-2x}$ passt am besten zu diesen Daten? D.h. welche Parameter a , b minimieren die Summe der Abweichungsquadrate? Zeichnen Sie Ihr Modell und die Messpunkte in ein Diagramm!

8.6. Vektorfunktionen

Die Königsetappe: Synthese von Linearer Algebra und Analysis: Wie kann ich einen Vektor ableiten?

Def D 8-7 Vektorfunktion

Sind die Koordinaten eines Vektors \vec{x} als Funktionen einer skalaren Größe t (z.B. Zeit) gegeben, so liegt eine Vektorfunktion $\vec{x} : \mathbf{R} \rightarrow \mathbf{R}^3$ vor. In den Komponenten erhält man:

$$\vec{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

Bezeichnet t die Zeit und x_1, x_2, x_3 die Raumkoordinaten, so heißt \vec{x} der Ortsvektor des Punktes $P(x_1, x_2, x_3)$.

Ist zusätzlich für den Parameter t ein Intervall $t_1 \leq t \leq t_2$ vorgegeben, so beschreibt die Menge aller Punkte $\{\vec{x}(t) | t_1 \leq t \leq t_2\}$ eine räumliche Kurve.

In Vorlesung: Raumkurve, mittlere Geschwindigkeit, Momentangeschwindigkeit.

Def D 8-8 Ableitung einer Vektorfunktion

Die 1. Ableitung der Vektorfunktion $\mathbf{x}(t)$ ist der Grenzwert:

$$\lim_{\Delta t \rightarrow 0} \frac{\vec{x}(t + \Delta t) - \vec{x}(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{x}}{\Delta t} = \frac{d\vec{x}}{dt} \equiv \dot{\vec{x}}$$

Der Vektor $\dot{\vec{x}}(t_0)$ ist der Tangentenvektor der Bahnkurve an der Stelle t_0 .

Satz S 8-5

Die Koordinaten der Ableitung eines Vektors erhält man durch Differenzieren der Koordinaten des Vektors.

ANMERKUNGEN:

1. Die Definitionen gelten sinngemäß auch für m statt für 3 Koordinaten.

2. Die Koordinatenfunktionen eines Vektors können genauso gut Funktionen von n Veränderlichen sein (statt nur Funktionen von t). Dann haben wir die allgemeine vektorwertige Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ vor uns. Jede einzelne Koordinate ist eine Funktion von n Veränderlichen.

Wie man Funktionen von n Veränderlichen abzuleiten hat, ist Gegenstand des nächsten Kapitels.

8.7. Der Gradient: Wo bitte geht's nach oben?

Stellen Sie sich vor, Sie stehen an einer Stelle $P_0 = (x_0, y_0)$ im Funktionengebirge $f(x, y)$ und wollen wissen, wo geht es nach oben? Genauer: Wo geht's möglichst steil nach oben?

Mathematischer: Wenn ich einen (kleinen) Schritt der Länge ds mache, welche Richtung wähle ich? Das Problem: Es gibt unendlich viele Richtungen! Alle ausprobieren??

Zum Glück gibt es ein wesentlich einfacheres Rezept, das mit nur zwei (!) Messungen auskommt:

Rezept:

- Bilde die partiellen Ableitungen an der Stelle (x_0, y_0) . Nehmen wir an, es sei $f_x(x_0, y_0) = 1$ und $f_y(x_0, y_0) = 2$. (Die Ableitungen sind die Steigungen, d.h. in der Nähe von (x_0, y_0) ist der Zuwachs in f je waagerechter Kästchenkante 1, der Zuwachs je senkrechter Kästchenkante ist 2.)
- Stecke die Zahlen in einen Vektor und marschiere in die Richtung, die der Vektor angibt. Also hier: 1 mm in x-Richtung und 2 mm in y-Richtung.

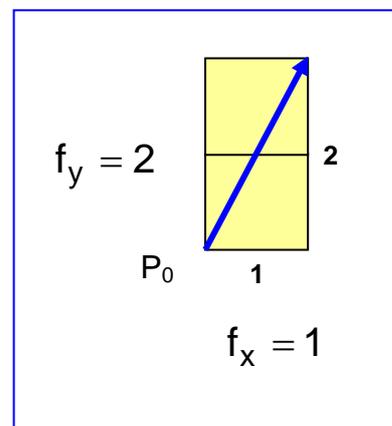
- Vektor $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, Strecke: $\sqrt{1^2 + 2^2} = \sqrt{5}$ mm.

- Zuwachs: $1 + 2 + 2 = 5$, also

- Zuwachs/mm = $\frac{5}{\sqrt{5}} = \sqrt{5} = 2.23$

- Das ist ein höherer Zuwachs als in x-Richtung alleine (1) oder in y-Richtung alleine (2)

- Keine andere Richtung bringt einen höheren Zuwachs/mm. Probieren Sie's aus!



Wer's genauer verstehen will: Totales Differential, Gradient.

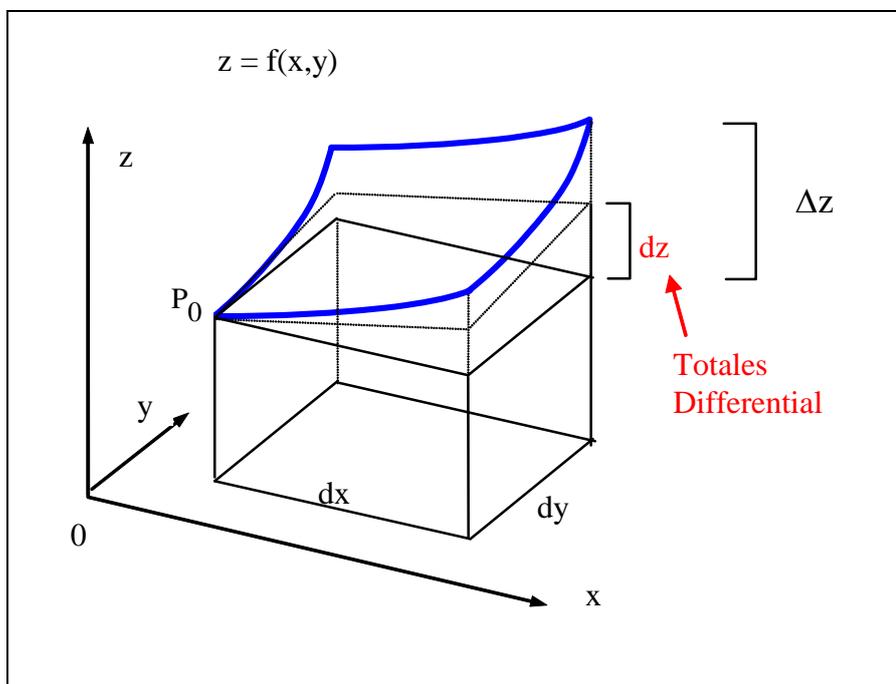
Große Bedeutung für die **praktische Optimierung**: Wenn ich ein Modell mit 5 oder 10 oder 50 Dimensionen habe (Parameter-Tuning für Simulationsmodell), dann bin ich in diesem hochdimensionalen Raum „blind wie ein Maulwurf“! Nur der Gradient gibt mir die Information, wie ich an den Steuerknöpfen drehen muss, um meinen Output zu verbessern.

Vorher noch eine kurze Auflockerung: Ein Applet zu PSO (Particle Swarm Optimization) von <http://gecco.org.chemie.uni-frankfurt.de/PsoVis/index.html> zeigt ein Beispiel für eine komplexere Optimierungsstrategie. „Ein Schwarm ist intelligenter als seine Individuen“ (→ WPF Spiele, Simulation u. Dynamische Systeme, Kapitel Partikel- und Schwarmssysteme).

8.7.1. Totales Differential

[evtl. nur Def. bringen, Rest im Selbststudium]

Betrachten wir eine Funktion $f(x,y)$ in zwei Veränderlichen an der Stelle $P_0=(x_0,y_0)$:



Wenn ich von P_0 ein Stück weitergehe, dann ist:

Totales Differential dz	=	Zuwachs der Tangentialebene in P_0, wenn in allen Koordinaten um $(dx,dy,...)$ weitergegangen wird
Funktionsänderung ΔZ	=	Zuwachs der Funktion, wenn man um denselben Vektor $(dx,dy,...)$ weitergeht

Als Formel:

$$dz = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy$$

$$\Delta z = f(x_0 + dx, y_0 + dy) - f(x_0, y_0)$$

Def D 8-9 Totales Differential (2 Veränderliche)

Das totale Differential dz einer Funktion $Z = f(x,y)$ im Punkt (x_0, y_0) ist definiert durch:

$$dz = f_x(x_0, y_0)dx + f_y(x_0, y_0)dy$$

Es gilt: $dz \approx \Delta z$ wenn dx, dy hinreichend klein sind (s. Zeichnung).

Die Tangentialebene im Punkt (x_0, y_0) ist gegeben durch:

$$Z(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Zum Beweis der Tangentialebenengleichung setzt man in allgemeiner Form $Z = a + b(x - x_0) + c(y - y_0)$ an und führt einen Koeffizientenvergleich durch.

Bei Funktionen von n Variablen erweitert man dies ganz analog:

Def D 8-10 Totales Differential (n Veränderliche)

Das totale Differential dz einer Funktion $Z = f(x_1, x_2, \dots, x_n) = f(\vec{x})$ wird definiert durch:

$$dz = f_{x_1} dx_1 + f_{x_2} dx_2 + \dots + f_{x_n} dx_n$$

dabei sind alle partiellen Ableitungen im betreffenden Punkt zu nehmen.

Es gilt auch hier: $dz \approx \Delta z = f(\vec{x} + d\vec{x}) - f(\vec{x})$, wenn $d\vec{x} = (dx_1, dx_2, \dots, dx_n)$ hinreichend klein ist.

Beispiele:

$$1) z = 2x + y^2 \quad x = 3, y = 5, dx = 0.3, dy = 0.2$$

$$z_1 = f(x, y) = f(3, 5) = 31$$

$$z_2 = f(x + dx, y + dy) = f(3.3, 5.2) = 33.64$$

$$\Delta z = 2.64$$

$$dz = 2dx + 2ydy = 2 \cdot 0.3 + 2 \cdot 5 \cdot 0.2 = 2.6$$

also gilt tatsächlich: $\Delta z \approx dz$

$$2) z = x \cdot y$$

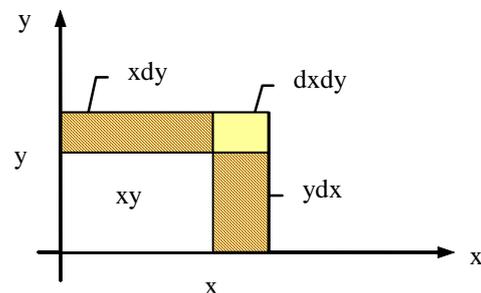
$$\Delta z = (x + dx)(y + dy) - xy$$

$$= ydx + xdy + dxdy$$

$$dz = ydx + xdy$$

$$dz = \Delta z \text{ bis auf Terme 2. Ordnung}$$

(dz : schraffiert, Δz : gelber Hintergrund)



3) Ein ideales Gas genügt der Zustandsgleichung $p(V, T) = \frac{RT}{V}$ (für 1 Mol). Das totale Differential dieser Funktion lautet somit:

$$dp = p_V dV + p_T dT = \frac{RT}{V^2} dV + \frac{R}{V} dT$$

und gibt näherungsweise die Änderung des Drucks bei kleinen Volumen- und Temperaturänderungen wieder.

8.8. Anwendungen Totales Differential

8.8.1. Kettenregel

Beispiel Zeitabhängigkeit: In vielen Anwendungen hat man Funktionen $z = f(x,y,t)$, in denen die Parameter $x = x(t)$ und $y = y(t)$ auch wieder von einem Parameter t (Zeitparameter) abhängen. Da die Ableitung nach der Zeit so häufig vorkommt, hat sich hierzu auch eine eigene "Punkt"-Schreibweise eingebürgert:

$$\dot{z} \equiv \frac{dz}{dt}$$

Wie genau verändert sich z , wenn sich die Zeit um ein hinreichend kleines dt verändert?

Satz S 8-6 Kettenregel (1 Parameter)

Ist $Z = f(X,y,t)$ eine Funktion von Veränderlichen $X = x(t)$ und $y = y(t)$, die selbst wieder von einem Parameter t (z. B. Zeit) abhängen, so hat die zusammengesetzte Funktion

$$z = z(t) = f(x(t), y(t), t)$$

die Ableitung (Differenzierbarkeit aller beteiligten Funktionen vorausgesetzt)

$$\dot{z} \equiv \frac{dz}{dt} = f_x \frac{dx}{dt} + f_y \frac{dy}{dt} + f_t$$



Übung: Erweitern Sie das Ergebnis aus Satz S 8-6 sinngemäß auf die Funktionen

$$f = f(x, y, z) = \cos z \exp(x^2 + 2y^2)$$

$$x(t) = t^2, \quad y = -t, \quad z = e^t \quad \text{indem Sie } \dot{f} = \frac{df}{dt} \text{ bilden!}$$

Beweis von Satz S 8-6 in Vorlesung (über totales Differential)

8.8.2. Der Gradient: Woher weht der Wind?

[Stingl, S. 343 und 353]

lat. Verb: **gradior**, gressus sum = **schreiten**

lat. Substantiv **gradus** = **Schritt**, Standpunkt, Stufe (vgl. graduell)

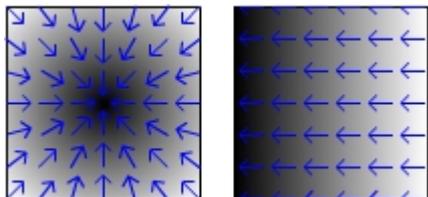
(hängt also eng mit unserem Bild vom Ausschreiten im Funktionengebirge zusammen)

Def D 8-11 Gradient

Der **Gradient grad f** einer Funktion $z = f(x_1, x_2, \dots, x_n)$ ist eine **Vektorfunktion** (s. Def D 8-7), die aus den partiellen Ableitungen besteht. Wertet man den Gradient an einer bestimmten Stelle $P_0 = (x_{10}, x_{20}, \dots, x_{n0})$ aus, so entsteht **(grad f)(P₀)**, ein einfacher **Vektor**:

$$\text{grad } f = \begin{pmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{pmatrix} \quad (\text{grad } f)(P_0) = \begin{pmatrix} f_{x_1}(P_0) \\ \vdots \\ f_{x_n}(P_0) \end{pmatrix}$$

In den beiden folgenden Bildern stellen die Grauschattierungen die Funktion f dar, wobei schwarz den höchsten Funktionswert darstellt, und die Pfeile symbolisieren den zugehörigen Gradienten:



[[http://de.wikipedia.org/wiki/Gradient_\(Mathematik\)](http://de.wikipedia.org/wiki/Gradient_(Mathematik))]

Man beachte: Der Gradient "lebt" im Raum (x,y) , in dem die Funktion f definiert ist, NICHT im Raum (x,y,z) , den man braucht, um sich die Funktion vorzustellen.

[in Vorlesung: wieso der Gradient die Windrichtung angibt]

Beispiel: Der Gradient der Funktion $f(x,y) = 3xy + y^2$ lautet $\text{grad } f = \begin{pmatrix} 3y \\ 3x + 2y \end{pmatrix}$, an

der Stelle $(x,y)=(2,1)$ wird er zum Vektor $(\text{grad } f)(2,1) = \begin{pmatrix} 3 \cdot 1 \\ 3 \cdot 2 + 2 \cdot 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$, an der Stelle

$(x,y)=(2,0)$ wird er zum Vektor $(\text{grad } f)(2,0) = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$.

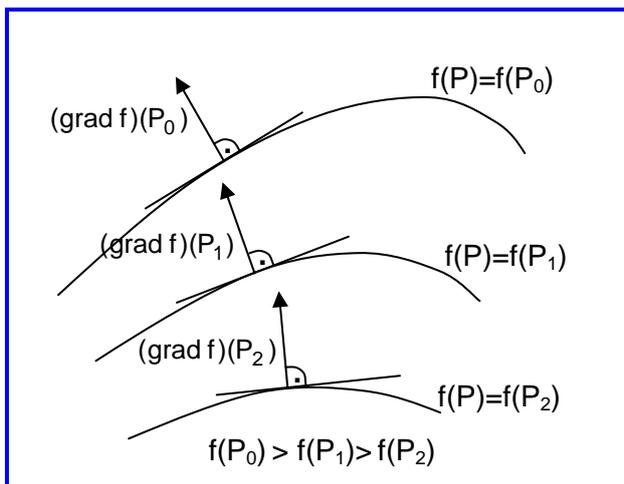
Die Ableitung einer Funktion mehrerer Veränderlicher $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ nach der Zeit lässt sich mit dem Gradienten sehr kompakt schreiben:

$$\frac{df(\vec{\mathbf{x}})}{dt} = \frac{df(x_1, \dots, x_n)}{dt} = \text{grad } f \cdot \frac{d\vec{\mathbf{x}}}{dt}$$

Satz S 8-7 Eigenschaften des Gradienten

1. Der Gradient $(\text{grad } f)(P_0)$ steht senkrecht auf der durch P_0 verlaufenden Äquipotentiallinie- oder fläche, also der Punktmenge $\{ P \in \mathbf{R}^n \mid f(P) = f(P_0) \}$.
2. Der Gradient weist in die Richtung des steilsten Anstiegs. D. h. die Änderung von f an der Stelle P_0 hat in Richtung von $(\text{grad } f)(P_0)$ ihren Maximalwert, nämlich den Betrag $|(\text{grad } f)(P_0)|$.

Der Gradient hat also eine sehr anschauliche Bedeutung im "Funktionengebirge".



Beispiele und Beweis von **Satz S 8-7** in Vorlesung



Übung: Wir befinden uns im Punkt $P=(x,y,z)=(1,2,-1)$. In welcher Richtung hat die Funktion

$$f = f(x,y,z) = \exp(x^2 + y^2 - 2z^2)$$

ihren steilsten Anstieg?

Der Gradient spielt eine große Rolle in der Optimierung, bei der man oft ein bestimmtes Fehlersignal zu minimieren hat. Statt unzählige (unendlich viele) Funktionsdifferenzen auszuprobieren, reicht es für „glatte“ Funktionen, an der Stelle P_0 den Gradienten auszurechnen (einen Vektor aus lauter Zahlen!) und ein Stückchen in die Gegenrichtung zu marschieren. Man spricht vom **Gradienten-Abstiegsverfahren** (*engl. gradient descent*), einer wichtigen Methode der Optimierung.

8.8.3. Linearisierung einer Funktion

[evtl. im Selbststudium, wenn Zeit knapp]

Beispiel: Nehmen wir die kompliziert aussehende Funktion

$$f(x, y) = \frac{x}{2} + \int_0^y \exp(t^2) dt$$

Jemand verlangt von uns, dass wir die Funktion in der Nähe von $(x_0,y_0)=(1,0)$ tabellieren, d.h. wir sollen ausrechnen

f(x,y)		y		
		-0.1	0.0	0.1
x	0.9	?	?	?
	1.0	?	?	?
	1.1	?	?	?

Nun ist das Integral analytisch gar nicht berechenbar, und numerische Methoden sind uns viel zu viel Aufwand. Wie kommen wir da raus?

Aus dem totalen Differential

$$dz \approx \Delta z = f(\vec{x}_0 + d\vec{x}) - f(\vec{x}_0)$$

folgt nach **Def D 8-10** mit $d\vec{x} = \vec{x} - \vec{x}_0 = \vec{h} = (h_1, \dots, h_n)^T$ unmittelbar:

Satz S 8-8 Linearisierung

Für Vektoren \vec{h} mit kleinem Betrag $|\vec{h}|$ ist

$$f(\vec{x}_0 + \vec{h}) \approx P(\vec{h}) = f(\vec{x}_0) + \sum_{i=1}^n f_{x_i}(\vec{x}_0) \cdot h_i$$

eine gute Näherung für $f(\vec{x}_0 + \vec{h})$. Man nennt $P(\vec{h})$ auch die **Linearisierung** der Funktion $f(\vec{x})$ um \vec{x}_0 .

Die Funktion $P(\vec{h})$ ist die **Taylor-Entwicklung** einer Funktion von n Veränderlichen bis zur 1. Ordnung.

Beachte: Die Linearisierung bis zur 1. Ordnung liefert gerade die Gleichung für die **Tangentialebene** durch den Punkt \vec{x} . Der Vektor \vec{h} enthält die freien Parameter.

Lösung zum Beispiel: Es ist $\vec{x}_0 = (1,0)^T$ und

$$f_x = \frac{1}{2}, \quad f_y = \exp(y^2), \text{ also ist } f_x(1,0) = \frac{1}{2}, \quad f_y(1,0) = 1. \text{ Weiterhin können wir}$$

$f(\vec{x}_0) = f(1,0) = 0.5$ ausrechnen (zum Glück!), denn das Integral wird zu 0. Wir setzen in die Formel aus **Satz S 8-8**

$$f(\vec{x}_0 + \vec{h}) \approx f(\vec{x}_0) + f_x(\vec{x}_0) \cdot h_x + f_y(\vec{x}_0) \cdot h_y = 0.5 + \frac{1}{2}h_x + 1 \cdot h_y \text{ die verschiedenen}$$

Werte für $h_x \in \{-0.1, 0.0, 0.1\}$ und $h_y \in \{-0.1, 0.0, 0.1\}$ ein und erhalten schließlich.

Näherung f. f(x,y)		y		
		-0.1	0.0	0.1
x	0.9	0.35	0.45	0.55
	1.0	0.40	0.50	0.60
	1.1	0.45	0.55	0.65

Die exakte Lösung kann mit Maple numerisch ausgerechnet und im Vergleich mit der Linearisierung visualisiert werden (s. [plot3d.mws](#), Abschnitt Linearisierung)

f(x,y) exakt		y		
		-0.1	0.0	0.1
x	0.9	0.35033	0.45	0.54967
	1.0	0.40033	0.50	0.59967
	1.1	0.45033	0.55	0.64967

Die Linearisierung ist also eine sehr gute Näherung. Und sie ist viel weniger Aufwand!

8.8.4. Optimierung mit Lagrange-Multiplikatoren

[Papula, Bd. 2, S. 333-340],

<http://www.slimy.com/~steuard/teaching/tutorials/Lagrange.html>

Die meisten realen Optimierungsprobleme haben Nebenbedingungen:

- Maximiere den Gewinn, wobei die Summe der Maschinen-Stunden konstant ist
- Minimiere die Freistunden in einem Stundenplan, wobei jeder Raum in jeder Stunde nur durch eine Klasse belegt sein darf
- usw.

Beispiel: Wo liegen die Extrema von $Z(x,y) = x+2y$, wenn die Nebenbedingung $x^2+y^2=5^2$ einzuhalten ist?

[Lösung in den Übungen]

Der simple Ansatz: Nebenbedingung nach einer Variablen auflösen, z.B. $y=y(x)$, in $Z(x,y)$ einsetzen, dann Extrema von $F(x) = Z(x,y(x))$ suchen.

Dies geht jedoch nicht immer: Sei $Z(x,y)$ eine zu optimierende Zielfunktion und $\varphi(x,y)=0$ die Nebenbedingung. Die obige Methode funktioniert nicht (gut),

- wenn die Auflösung von $\varphi(x,y)=0$ nach x oder y nicht möglich oder aber zu aufwendig ist;
- wenn die Auflösung $y=y(x)$ zwar gelingt, aber $Z(x,y(x)) = F(x)$ zu unnötig komplizierten Ableitungen $F'(x)$ oder $F''(x)$ führt.

Die Methode der Lagrange-Multiplikatoren bietet hier einen guten Trick:

Satz S 8-9 Lagrange-Multiplikator

Gegeben eine zu optimierende Zielfunktion $Z(x,y)$ und eine Nebenbedingung $\varphi(x,y)=0$, die gleichzeitig einzuhalten ist. Dieses Problem wird in folgenden Schritten gelöst:

1. Bilde die Hilfsfunktion

$$F(x, y, \lambda) = Z(x, y) + \lambda\varphi(x, y)$$

Der (noch unbekannte) Parameter λ heißt Lagrange-Multiplikator

2. Setze die partiellen Ableitungen gleich Null:

$$F_x = Z_x(x, y) + \lambda\varphi_x(x, y) = 0$$

$$F_y = Z_y(x, y) + \lambda\varphi_y(x, y) = 0$$

$$F_\lambda = \varphi(x, y) = 0$$

Aus diesen 3 Gleichungen lassen sich die 3 Unbekannten x , y und λ bestimmen.

3. Gibt es mehrere Lösungen, so kann man durch Einsetzen in $Z(x,y)$ herausfinden, welche der Lösungen ein Maximum (bzw. Minimum) sein kann. (Einen hinreichenden Nachweis hat man damit allerdings nicht)

Die Sache mutet wie ein Taschenspielertrick an: Erst ergänzen wir ein $\lambda \cdot 0$, erhalten so eine neue Funktion $F(x, y, \lambda)$, eliminieren dann λ wieder und haben angeblich eine Lösung von $Z(x, y)$, die die Nebenbedingung einhält? Wieso?

In Vorlesung beweisen wir, wieso dieser Trick funktioniert.

Anmerkung:

- Das Verfahren der Lagrange-Multiplikatoren lässt sich ohne Schwierigkeiten auch auf Funktionen von n Variablen mit m Nebenbedingungen ($m < n$) verallgemeinern. Die Hilfsfunktion lautet dann:

$$F(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = Z(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i \varphi(x_1, \dots, x_n)$$

und die $(n+m)$ partiellen Ableitungen und damit Gleichungen ergeben sich analog.

- Die Nebenbedingungen müssen in **Gleichungsform** vorliegen. Bei Nebenbedingungen in Ungleichungsform helfen die Lagrange Multiplikatoren nicht weiter, hier braucht man andere Optimierungsmethoden (Simplex oder Interior Points). Das wollen wir aber hier nicht weiterverfolgen.

Beispiel: Auf einer ebenen Bühne soll der Punkt A optimal ausgeleuchtet werden. Es gilt das Lambertsche Gesetz

$$B(\alpha, r) = \frac{I_0 \cos \alpha}{r^2}$$

Die Lichtquelle in festem Abstand a ist auf einer senkrechten Stange verschiebbar.

Lösung: Nebenbedingung ist

$$\sin \alpha = \frac{a}{r} \Leftrightarrow r \sin \alpha - a = 0$$

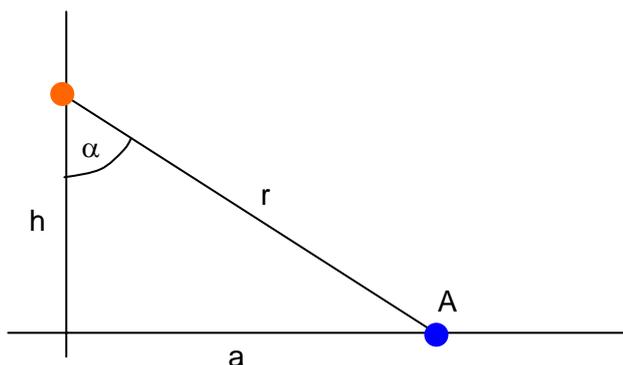
Zu minimieren:

$$F(\alpha, r, \lambda) = \frac{I_0 \cos \alpha}{r^2} + \lambda(r \sin \alpha - a) \Rightarrow$$

$$\left. \begin{aligned} F_\alpha &= -\frac{I_0 \sin \alpha}{r^2} + \lambda r \cos \alpha = 0 \Rightarrow \lambda = \frac{I_0 \tan \alpha}{r^3} \\ F_r &= -\frac{2I_0 \cos \alpha}{r^3} + \lambda \sin \alpha = 0 \Rightarrow \lambda = \frac{2I_0}{r^3 \tan \alpha} \end{aligned} \right\} \Rightarrow \tan^2 \alpha = 2 \Rightarrow \tan \alpha = \pm \sqrt{2}$$

Da die gesuchte Lösung im 1. Quadranten liegen muss ($0 \leq \alpha \leq 90^\circ$), kommt nur die +-Lösung in Frage: $\alpha = \arctan \sqrt{2} = 54.74^\circ$

Über die Nebenbedingung folgt $r = a/\sin \alpha = 1.225a$ sowie $h = a \tan \alpha = 0.707a$.





Übung: Ein Zufallsexperiment habe 4 mögliche Ergebnisse, die mit den Wahrscheinlichkeiten p_1, \dots, p_4 auftreten. Weil *eines* dieser Ergebnisse immer herauskommen muss, gilt offensichtlich $p_1 + p_2 + p_3 + p_4 = 1$. Bei welchen Wahrscheinlichkeiten wird das Produkt

$$Z(p_1, \dots, p_4) = p_1 p_2 p_3 p_4$$

maximal?

Zeigen Sie mit Lagrange-Multiplikatoren, dass die Lösung $p_1 = \dots = p_4 = 0.25$ ist!

8.9. Fazit

Wichtige Begriffe und Ergebnisse aus diesem Kapitel waren:

reelle Funktion mehrerer Veränderlicher	$f: \mathbf{R}^n \rightarrow \mathbf{R}$: n Veränderliche, 1 abhängige Größe
Vektorfunktion	$\vec{x}: \mathbf{R}^n \rightarrow \mathbf{R}^m$: n Veränderliche, m abhängige Größen
Tangentialebene	Ebene im Raum \mathbf{R}^{n+1} durch den Punkt $(\vec{x}, f(\vec{x}))$, die in allen Richtungen die Steigung der (stetigen) Funktion f in \vec{x} hat.
Äquipotentialflächen	Flächen mit $f(\vec{x}) = \text{const.}$ im \vec{x} -Raum. Für $\vec{x} \in \mathbf{R}^2$ werden die Flächen zu Linien, den <u>Höhenlinien</u> .
partielle Ableitung nach x_i	alle Veränderlichen außer x_i als konstant festsetzen, dann "normal" nach x_i ableiten
totales Differential	Zuwachs in der Tangentialebene bei Verrückung um $d\vec{x}$
Gradient von f	Vektorfunktion im Raum \mathbf{R}^n , die i. Komponente ist f_{x_i} .

Wichtige Ergebnisse:

- Funktionen mehrerer Veränderlicher lassen sich über Flächen im Raum, über Höhenliniendiagramme oder über Kennlinienfelder visualisieren.
 - Höhenlinien: $z = f(x, y)$ nach y auflösen
 - Kennlinien: alle Veränderliche bis auf eine konstant festsetzen.
- Die Differentialrechnung einer Veränderlichen lässt sich auf Funktionen mehrerer Veränderlicher übertragen.
 - **partielle Ableitung:** alle Veränderliche bis auf eine konstant, dann ableiten.
- Der **Gradient** ist der Vektor aller 1. partiellen Ableitungen (Kap. 8.8.1). Er steht an jeder Stelle senkrecht auf den Äquipotentialflächen und weist in Richtung des steilsten Anstiegs.
- **Extremwerte:** Hinreichende Kriterien sind für mehr als 2 Variablen schwierig, für 2 Variablen aber gut angebar (Satz S 8-3).
- Mit der Methode der **kleinsten Quadrate (LS-Methode)** (Kap. 8.5.2) lassen sich gut Ausgleichsgeraden bestimmen.
- Viele reale Optimierungsprobleme mit mehreren Veränderlichen haben neben einem Maximierungsziel auch weitere Nebenbedingungen zwischen den Veränderlichen in Gleichungsform. Hier hilft die Methode der **Lagrange-Multiplikatoren** (Kap. 8.8.4) entscheidend weiter.