

Skript Analysis und Lineare Algebra A

Steffen Habermalz, PhD

22. Oktober 2022

Inhaltsverzeichnis

1	Einführung	4
1.1	Mathematische Aussagen	4
1.2	Logische Operatoren	4
1.3	Quantoren	8
1.4	Regeln	9
2	Mengen und Funktionen	11
3	Funktionen mehrerer Variablen	17
3.1	Funktionale Grenzwerte	17
3.2	Stetigkeit	21
3.3	Differenzierbarkeit	25
3.4	Partielle Ableitungen	27
3.5	Partielle Ableitung Höherer Ordnung	30
4	Mathematisches Handwerkszeug	33
4.1	Kettenregeln	33
4.2	Höhenlinien und deren Steigung	35
4.3	Implizites Differenzieren	40
4.4	Homogene Funktionen	42
4.5	Approximationen	44
5	Multivariate Optimierung	49
5.1	Generelles	49
5.2	Multivariate Optimierung Ohne Nebenbedingungen	52
5.3	Parameter, Wertfunktionen und das Envelope Theorem für Optimierungsprobleme ohne Nebenbedingungen	68
5.4	Multivariate Optimierung unter Nebenbedingungen	71
5.5	Exkurs: Nicht-Lineare Gleichungssysteme	72
6	Lineare Algebra	87
6.1	Gauß'sche Elimination	88
6.2	Die Inverse	97
6.3	Lineare Unabhängigkeit	98
6.4	Spanne von Vektoren	98

6.5	Lineare Projektion	99
6.6	Eigenwerte und Eigenvektoren	101
7	Integralrechnung	105
7.1	Unbestimmte Integrale	105
7.2	Bestimmte Integrale	108

1 Einführung

1.1 Mathematische Aussagen

Die meisten mathematischen Aussagen fallen in eine der folgenden Kategorien

1. Objekt a hat die Eigenschaft P .
2. Alle Objekte des Typs T haben die Eigenschaft P .
3. Es existiert ein Objekt des Typs T , das die Eigenschaft P hat.
4. Wenn *Aussage 1*, dann *Aussage 2*

Beispiele:

- A: 5 ist eine Primzahl
- B: Alle Primzahlen sind ungerade
- C: Es existiert eine Primzahl, die gerade ist
- D: Wenn $x > 7$, dann $x^2 > 47$
- Bonn ist die Hauptstadt von Deutschland

Aussagen

- sind entweder WAHR oder FALSCH
- sind entweder einzelne Aussagen oder können mit den logischen Operatoren UND (\wedge), ODER (\vee), NICHT (\neg), IMPLIZIERT (\implies), IST ÄQUIVALENT (\iff) und den Quantoren (ALLQUANTOR(\forall), EXISTENZQUANTOR (\exists)) beschrieben werden

1.2 Logische Operatoren

1.2.1 UND (\wedge , Konjunktion)

Beispiele:

- $(\text{MA liegt in BW}) \wedge (5 \text{ ist eine Primzahl})$
- $(\text{James Brown ist tot}) \wedge (\text{Donald Trump lebt})$

Der Wahrheitsgehalt einer Konjunktion wird durch die folgende Wahrheitstabelle definiert

A	B	$A \wedge B$
w	w	w
w	f	f
f	w	f
f	f	f

1.2.2 ODER (\vee , Disjunktion)

Beispiele:

- (MA liegt in BW) \vee (5 ist eine Primzahl)
- (James Brown ist tot) \vee (Donald Trump lebt)

Der Wahrheitsgehalt einer Disjunktion wird durch die folgende Wahrheitstabelle definiert

A	B	$A \vee B$
w	w	w
w	f	w
f	w	w
f	f	f

1.2.3 NICHT (\neg , Negation)

Beispiele:

- \neg (5 ist eine Primzahl)
- \neg (Donald Trump ist tot)

Der Wahrheitsgehalt einer Aussage wird durch die Negation umgekehrt

A	$\neg A$	B	$\neg(A \wedge B)$
w	f	w	f
w	f	f	w
f	w	w	w
f	w	f	w

1.2.4 IMPLIZIERT (\implies)

Beispiele:

- Wenn (A) sie mehr als 90 Punkte erreichen, dann (B) bekommen sie die Note 1,0.
- Wenn (A) sie mehr als 90 Punkte erreichen, dann (C) ist $\pi > 3$.

Der Wahrheitsgehalt einer Implikation wird durch die folgende Wahrheitstabelle definiert:

A	B	$A \implies B$
w	w	w
w	f	f
f	w	w
f	f	w

Warum ist die Implikation wahr, wenn die erste Aussage FALSCH ist? Nehmen wir als Beispiel die obige Implikation

mehr als 90 Punkte \implies Note=1,0

Die ersten zwei Fälle sind (hoffentlich) klar. Was passiert wenn der Student weniger als 90 Punkte erreicht? Er bekommt entweder eine 1,0 oder nicht! Wichtig ist, dass keine dieser beiden Situationen die Implikation unwahr macht (Aber auch nicht wahr). Mathematiker haben sich dafür entschieden, dass eine Implikation so lange wahr ist, bis das Gegenteil bewiesen ist. Deshalb ist die Implikation auch in den Fällen wahr, in denen die erste Aussage falsch ist.

Regel zum Merken: Aus einer falschen Aussage lässt sich alles ableiten!

Beispiel 1.1. Sind die folgenden Implikation wahr oder falsch?

1. $2 = 3 \implies 2 + 1 = 3 + 1$

2. Wenn der März 31 Tage hat, dann sind Hunde Säugetiere

Das zweite Beispiel zeigt, dass Implikation nicht mit einem Kausalzusammenhang gleichzusetzen ist.

Die Kontraposition Die Kontraposition stellt die Negation der Implikation dar.

$$\neg B \implies \neg A$$

Die Kontraposition ist Grundlage vieler Beweise und basiert auf der logischen Äquivalenz von Implikation und Kontraposition

$$(A \implies B) \iff (\neg B \implies \neg A)$$

Es gibt Situationen, in denen sich die Kontraposition einfacher beweisen lässt.

A	B	$\neg B$	$\neg A$	$A \implies B$	$\neg B \implies \neg A$
w	w	f	f	w	w
w	f	w	f	f	f
f	w	f	w	w	w
f	f	w	w	w	w

1.2.5 Notwendige und Hinreichende Bedingungen

Gegeben sei die Implikation

$$A \implies B$$

Hier ist

- B eine notwendige Bedingung für A
,d.h. wenn Aussage A wahr ist, ist *notwendigerweise* auch Aussage B wahr. A kann ohne B nicht wahr sein.
- A eine hinreichende Bedingung für B
,d.h. um zu zeigen, dass Aussage B wahr ist, ist es *hinreichend*, dass Aussage A wahr ist.

Beispiele

- Wassertemperatur $< 0 \implies$ Wasser gefroren
 - Notwendig: Ohne gefrorenes Wasser kann die Temperatur nicht negativ sein.
 - Hinreichend: Wenn ich weiß, dass die Temperatur negativ ist, MUSS das Wasser gefroren sein.
- Mit dem Kopf durch die Wand \implies Kopfschmerzen
 - Notwendig: Ohne Kopfschmerzen kann ich nicht (wortwörtlich) durch die Wand gehen.
 - Hinreichend: Wenn ich mit dem Kopf durch die Wand gehe, sind Kopfschmerzen unausweichlich.
- Lewis Hamilton gewinnt \implies Mercedes gewinnt
 - Notwendig: Ohne dass Mercedes gewinnt kann Lewis Hamilton nicht gewinnen
 - Hinreichend: Wenn Lewis Hamilton gewinnt, gewinnt automatisch auch Mercedes.
- Höre Lieblingssong \implies Habe Glücksgefühle
 - Notwendig: Ohne Glücksgefühle kann es nicht mein Lieblingssong sein.
 - Hinreichend: Wenn ich meinen Lieblingssong höre, folgen Glücksgefühle automatisch.
- x, y sind Primzahlen $\implies \text{ggT}(x, y) = 1$
 - Notwendig: Ohne dass der größte gemeinsame Teiler 1 ist, können x, y keine Primzahlen sein.
 - Hinreichend: Da Primzahlen nur durch sich selbst und 1 teilbar sind, muss der größte gemeinsame Teiler 1 sein.
- x ist gerade $\implies x^2$ ist gerade
 - Notwendig: Ohne dass x^2 gerade ist, kann x nicht gerade sein.
 - Hinreichend: Wenn x gerade ist folgt daraus, dass auch x^2 gerade ist.

1.2.6 IST ÄQUIVALENT, (\iff)

Zwei Aussagen A und B sind äquivalent, wenn sie sich gegenseitig implizieren, d.h.

$$(A \implies B) \wedge (B \implies A)$$

- A und B sind äquivalent, wenn sie sich gegenseitig implizieren

$$(A \implies B) \wedge (B \implies A)$$

- Man muss also die Implikation in beide Richtungen beweisen um die Äquivalenz zu etablieren
- Man schreibt dann $A \iff B$
- Anders ausgedrückt müssen A und B entweder gemeinsam wahr oder falsch sein
- Zwei Aussagen A und B sind logisch äquivalent, wenn A und B die gleiche Wahrheitstabelle besitzen
- Englisch: „iff“, oder „if and only if“. Deutsch bezeichnet: „dann und nur dann“ oder „wenn und nur wenn“

Beispiel 1.2. Wahrheitstabelle für $A \iff B$

A	B	$A \implies B$	$B \implies A$	$(A \implies B) \wedge (B \implies A)$	$A \iff B$
w	w	w	w	w	w
w	f	f	w	f	f
f	w	w	f	f	f
f	f	w	w	w	w

Beispiel 1.3. Logische Äquivalenz von $(A \implies B)$ und $(A \wedge B) \vee \neg A$

A	B	$A \wedge B$	$\neg A$	$(A \wedge B) \vee \neg A$	$A \implies B$
w	w	w	f	w	w
w	f	f	f	f	f
f	w	f	w	w	w
f	f	f	w	w	w

1.3 Quantoren

1.3.1 EXISTENZQUANTOR (\exists)

Der Existenzquantor impliziert, dass es ein Element einer Menge gibt, dass eine bestimmte Eigenschaft besitzt

- Es existiert eine reelle Zahl x , so dass $x^2 + 2x + 1 = 0$

$$\exists(x \in \mathbb{R})(x^2 + 2x + 1 = 0)$$

- Es existieren ganze Zahlen p, q , so dass $\frac{p}{q} = \sqrt{2}$

$$\exists(p \in \mathbb{Z})\exists(q \in \mathbb{Z})\left(\frac{p}{q} = \sqrt{2}\right)$$

1.3.2 ALLQUANTOR (\forall)

Der Allquantor impliziert, dass eine Aussage für alle Elemente einer Menge Gültigkeit besitzt.

- alle Quadrate reeller Zahlen sind positiv

$$\forall(x \in \mathbb{R})(x^2 > 0)$$

- Es gibt keine größte natürliche Zahl

$$\forall(n \in \mathbb{N})\exists(m \in \mathbb{N})(m > n)$$

Achtung: die Reihenfolge der Quantoren ist wichtig. Die Aussage

$$\exists(n \in \mathbb{N})(\forall m \in \mathbb{N})(n > m)$$

ist falsch!

1.4 Regeln

1.4.1 Kommutativität

$$A \wedge B \iff B \wedge A$$

$$A \vee B \iff B \vee A$$

1.4.2 Assoziativität

$$A \wedge (B \wedge C) \iff (A \wedge B) \wedge C$$

$$A \vee (B \vee C) \iff (A \vee B) \vee C$$

1.4.3 Distributivität

$$A \wedge (B \vee C) \iff (A \wedge B) \vee (A \wedge C)$$

$$A \vee (B \wedge C) \iff (A \vee B) \wedge (A \vee C)$$

1.4.4 Doppelte Negation

$$\neg(\neg A) \iff A$$

1.4.5 de Morgan'sche Regeln

$$\neg(A \wedge B) \iff (\neg A \vee \neg B)$$

$$\neg(A \vee B) \iff (\neg A \wedge \neg B)$$

1.4.6 Kontraposition

$$(A \implies B) \iff (\neg B \implies \neg A)$$

2 Mengen und Funktionen

Wir wollen hier noch einmal formal das Konzept der Funktion darstellen. Dazu betrachten wir zwei Mengen A und B . Grundsätzlich kann zwischen den Elementen der beiden Mengen jede nur erdenkliche Beziehung (oder keine) bestehen. Z.B. könnte die Menge A aller Erwachsenen und die Menge B aller Kinder sein. Dann kann eine Relation „ist Kind von“ definiert werden. Der Funktionsbegriff limitiert die Beziehung zwischen den zwei Mengen in zweierlei Hinsicht. Bei einer Funktion muss *jedem* Element von A *genau ein* Element in B zugeordnet sein.

Hier heißt die Menge A *Definitionsbereich* und die Menge B *Wertebereich*. Nicht alle Elemente im Wertebereich müssen mit Elementen im Definitionsbereich verknüpft sein, zum Beispiel kann für die Funktion $y = x^2$ der Wertebereich \mathbb{R} angenommen werden (der auch negative Elemente enthält).

Definition 2.1. *Funktion: Eine Funktion ist eine Beziehung zwischen zwei Mengen X und Y , in der JEDEM Element des Definitionsbereichs (X) GENAU EIN Element des Wertebereichs (Y) zugeordnet wird. Geschrieben als $f : X \rightarrow Y$*

Das Konzept einer Funktion umfasst die folgenden vier Fälle:

Definition 2.2. *Funktionstypen*

1. Eine Funktion $f : X \rightarrow Y$ heißt **injektiv**, wenn jedem Element des Wertebereichs höchstens ein Element des Definitionsbereichs zugeordnet ist.

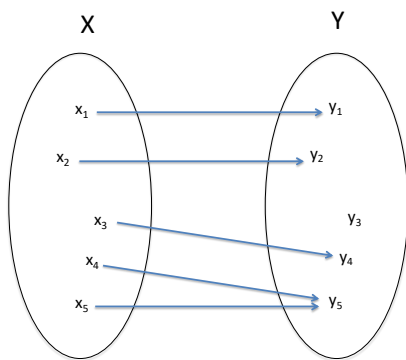
$$(\forall x_1, x_2 \in X)(f(x_1) = f(x_2) \implies x_1 = x_2)$$

2. Eine Funktion $f : X \rightarrow Y$ heißt **surjektiv**, wenn jedem Element des Wertebereichs mindestens ein Element des Definitionsbereichs zugeordnet ist.

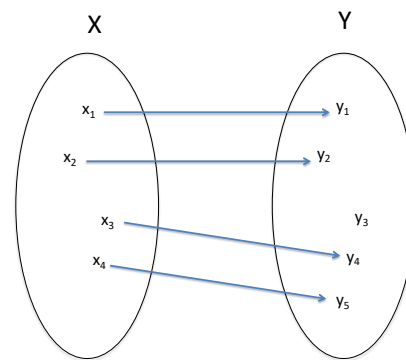
$$(\forall y \in Y)(\exists x \in X)(y = f(x))$$

3. Eine Funktion $f : X \rightarrow Y$ heißt **bijektiv**, wenn sie injektiv und surjektiv ist, d.h. für jedes y gibt es GENAU EIN x , so dass $f(x) = y$
4. Der Rest der Funktionen ist weder injektiv noch surjektiv.

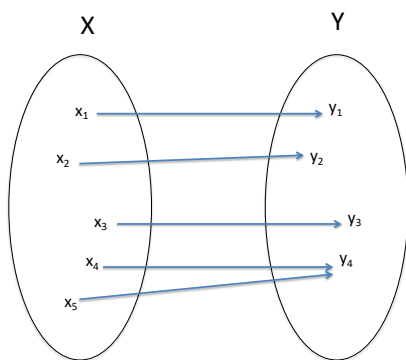
Die Abbildungen auf der nächsten Seite verdeutlichen dies.



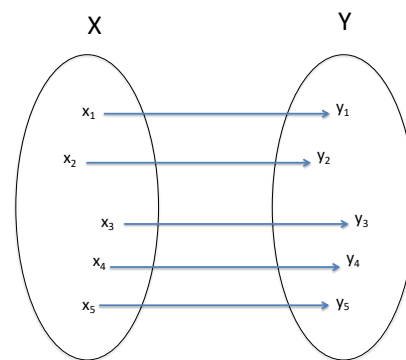
Funktion



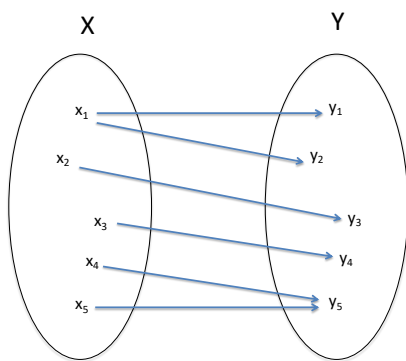
Injektive Funktion



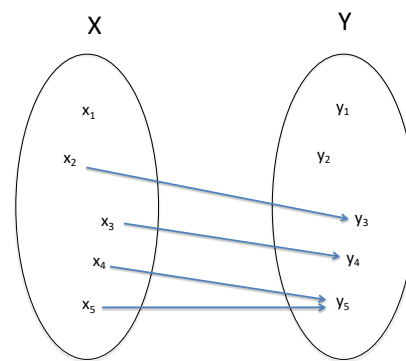
Surjektive Funktion



Injektiv und surjektiv (=bijektiv)



Keine Funktion



Keine Funktion

Abbildung 1: Funktionen

Beispiel 2.1. Zeigen Sie, dass eine lineare Funktion mit Definitionsbereich und Wertebereich \mathbb{R} bijektiv ist.

Beispiel 2.2. Gegeben sei die Funktion $f(x) = x^2$. Wählen sie Definitionsbereich und Wertebereich, so dass die Funktion

1. weder surjektiv noch injektiv ist.
2. injektiv aber nicht surjektiv ist.
3. surjektiv aber nicht injektiv ist.
4. bijektiv ist.

Oft werden Funktionen wie folgt geschrieben

$$f : \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto x^2 \end{cases}$$

Wir werden in diesem Kurs diese Funktion vereinfacht schreiben als

$$f(x) = x^2 \quad \text{oder} \quad y = x^2$$

Wenn nicht ausdrücklich aufgeführt wird der größtmögliche Definitionsbereich in den reellen Zahlen angenommen.

Wir führen jetzt die Umkehrfunktion ein. Umkehrfunktionen existieren für bijektive Funktionen.

Definition 2.3. Es sei $f(x)$ eine bijektive Funktion. Die Umkehrfunktion von $f(x)$ ordnet jedem Element des Wertebereichs genau ein Element des Definitionsbereichs zu. Es gilt dann für jedes Element des Definitionsbereichs

$$f^{-1}(f(x)) = x$$

Anmerkung:

1. Die Umkehrfunktion ist graphisch das Spiegelbild von $f(x)$ mit der Winkelhalbierenden (45 Grad Linie) als Spiegelachse.

2. $f^{-1} \neq 1/f$

Beispiel 2.3. Finden sie die Umkehrfunktion zu den folgenden Funktionen und zeichnen sie den Graphen der beiden Funktionen.

1. $f(x) = 5 - 3x$

2. $f(x) = x^3$

3. Hat $f(x) = x^2$ eine Umkehrfunktion?



Bisher wurden (meist) nur Funktionen einer Variablen behandelt. In der Volkswirtschaftslehre begegnen uns aber ständig Funktionen, in denen eine Zielvariable von mehreren unabhängigen Variablen bestimmt wird. Beispiele sind:

1. Nutzenfunktionen. Hier ist der Gesamtnutzen einer Person abhängig vom Konsum mehrerer Güter
2. Produktionsfunktionen. Hier ist die Produktionsmenge abhängig von der Menge der eingesetzten Produktionsfaktor (z.B. Kapital und Arbeit)

Wir definieren zunächst den n-dimensionalen Euklidischen Raum mittels des kartesischen Produkts und dann die reelle Funktion mehrerer Variablen.

Definition 2.4. (Kartesisches Produkt) . Das kartesische Produkt zweier Mengen A und B wird $A \times B$ geschrieben und bezeichnet die Menge von 2-Tupeln

$$[(a, b) : a \in A \wedge b \in B]$$

Wenn immer die gleiche Menge \mathbb{R} verwendet wird erhält man $\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n \times}$, den Euklidischen Raum, in dem jeder n-dimensionale Punkt mit einem n-Tupel repräsentiert wird

$$(x_1, x_2, x_3, \dots, x_n)$$

Beispiel 2.4. Finden sie das Kartesische Produkt von zwei 6-seitigen Würfeln.



Definition 2.5. Eine reelle Funktion mehrerer Variablen ist eine Funktion mit Definitionsbereich $A \subseteq \mathbb{R}^n$ (mit $n > 1$) und Wertebereich $B \subseteq \mathbb{R}$, z.B. $f : \mathbb{R}^n \rightarrow \mathbb{R}$, die jedem n -dimensionalen Punkt im Definitionsbereich genau einen Punkt im Wertebereich zuordnet.

Beispiel 2.5. :

1. $f : \mathbb{R}^2 \rightarrow \mathbb{R} \quad f(x, y) = x^2 + y^2$
2. Eine Nutzenfunktion über n Güter $U(x_1, x_2, \dots, x_n) \rightarrow \mathbb{R}$
3. $y = f(\mathbf{x})$

Meist werden Funktionen mit zwei Variablen behandelt um den Sachverhalt einfacher zu gestalten. Wir werden das zunächst auch tun. Die Verallgemeinerung auf n Variablen ist im Anschluss nicht so schwer.

Eines unserer Hauptziele ist es Funktionen von zwei oder mehr Variablen mit und ohne Nebenbedingungen zu optimieren. Dazu müssen wir viele Konzepte der Optimierung von Funktionen einer Variablen auf Funktionen mehrerer Variablen übertragen (Stetigkeit und Differenzierbarkeit) und einige neue Konzepte (partielle Ableitung, Höhenlinien, Differential) hinzufügen. Damit beginnen wir im nächsten Kapitel.

Zu guter Letzt wollen wir noch Vektoren einführen. Ein Punkt im Euklidischen Raum kann nämlich entweder durch ein n -Tupel oder durch einen n -dimensionalen Vektor dargestellt werden.

Definition 2.6. (Vektor in \mathbb{R}^n) Ein n -dimensionaler Vektor in \mathbb{R}^n ist ein n -Tupel reeller Zahlen. Wir schreiben den (Spaltenvektor)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \in \mathbb{R}^n \quad \text{mit} \quad x_i \in \mathbb{R} \quad \text{mit} \quad 1 \leq i \leq n$$

1. Die Transponierte des Vektors \mathbf{x} wird als \mathbf{x}' geschrieben und ist der korrespondierende Zeilenvektor

$$\mathbf{x}' = [x_1 \quad x_2 \quad \dots \quad x_n]$$

2. Vektoren (gleicher Dimension) können addiert (subtrahiert) oder mit einer reellen Zahl λ multipliziert werden.

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \cdot \\ \cdot \\ x_n + y_n \end{bmatrix} \quad \lambda \mathbf{x} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \cdot \\ \cdot \\ \lambda x_n \end{bmatrix}$$

3. Ein Punkt im Euklidischen Raum kann also auch als Vektor dargestellt werden. Der Vektor ist ein Pfeil vom Ursprung zum Punkt und hat somit auch eine Länge.

4. Der Betrag (auch Norm genannt) (Länge vom Ursprung) eines Vektors \mathbf{x} ist definiert als

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

5. Vektoren mit $\|\mathbf{x}\| = 1$ werden als Einheitsvektoren bezeichnet. $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ ist ein Einheitsvektor, da

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = 1$$

Beispiel 2.6. Gegeben seien folgende Skalare und Vektoren

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix} \quad e = 3 \quad f = -4$$

Berechnen Sie

1. $\mathbf{a} + \mathbf{b}$

3. $e\mathbf{a}$

5. $\|\mathbf{a}\|$

2. $\mathbf{b} - \mathbf{c}$

4. $f\mathbf{c}$

6. $\left\| \frac{\mathbf{a}}{\|\mathbf{a}\|} \right\|$

3 Funktionen mehrerer Variablen

In diesem Kapitel geht es darum bekannte Konzepte aus der Analyse von Funktionen einer Variablen (wie z.B. Stetigkeit) auf Funktionen von mehreren Variablen zu übertragen. Wir werden zunächst noch einmal die "einfache" Version wiederholen.

3.1 Funktionale Grenzwerte

Hier interessieren wir uns dafür was mit einem Funktionswert $f(x)$ passiert, wenn der Wert der unabhängigen Variablen x einem Wert x_0 immer näher kommt (aber nie gleich x_0 wird). Einfach ist dies bei einer Funktion wie

$$f(x) = 3x - 2$$

Was passiert mit dem Funktionswert wenn x der Zahl $x_0 = 2$ immer näher kommt? Man kann hier einfach $x = 2$ in die Funktion einsetzen um zu zeigen, dass

$$\lim_{x \rightarrow 2} (3x - 2) = 3 \cdot 2 - 2 = 4$$

Wir lesen "Limes (oder Grenzwert) von $(3x - 2)$ mit x gegen 2 gleich 4".

Was aber, wenn wir den Grenzwert von $f(x) = \frac{e^x - 1}{x}$ an der Stelle $x_0 = 0$ berechnen wollen?

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x}$$

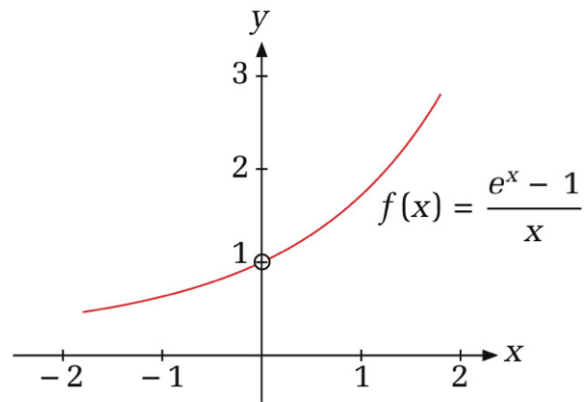
Einsetzen liefert $\frac{0}{0}$. Das Problem ist, dass die Funktion an dieser Stelle nicht definiert ist. Wir können den Funktionswert für Werte von x nahe 0 berechnen um eine Idee über den Grenzwert zu bekommen.

x	-1	-0.1	-0.001	-0.0001	0.0	0.0001	0.001	0.1	1
$f(x)$	0.632	0.956	0.999	1.000	*	1.000	1.001	1.052	1.718

* nicht definiert

Tabelle 1: Werte von $f(x) = (e^x - 1)/x$, wenn x nahe bei 0 ist

Wir sehen, dass die Funktion gegen den Grenzwert 1 konvergiert wenn x gegen 0 geht. Dies ist zwar kein formaler Beweis aber richtig, wie der folgende Graph zeigt.



Der Kreis auf dem Graph zeigt, dass die Funktion an der Stelle $x = 0$ nicht definiert ist. Wir wollen hier ein intuitives Verständnis für den Grenzwert erlangen und verzichten deshalb auf eine rigorose Formulierung.

Definition 3.1. : Grenzwert. Der Ausdruck

$$\lim_{x \rightarrow x_0} f(x) = A$$

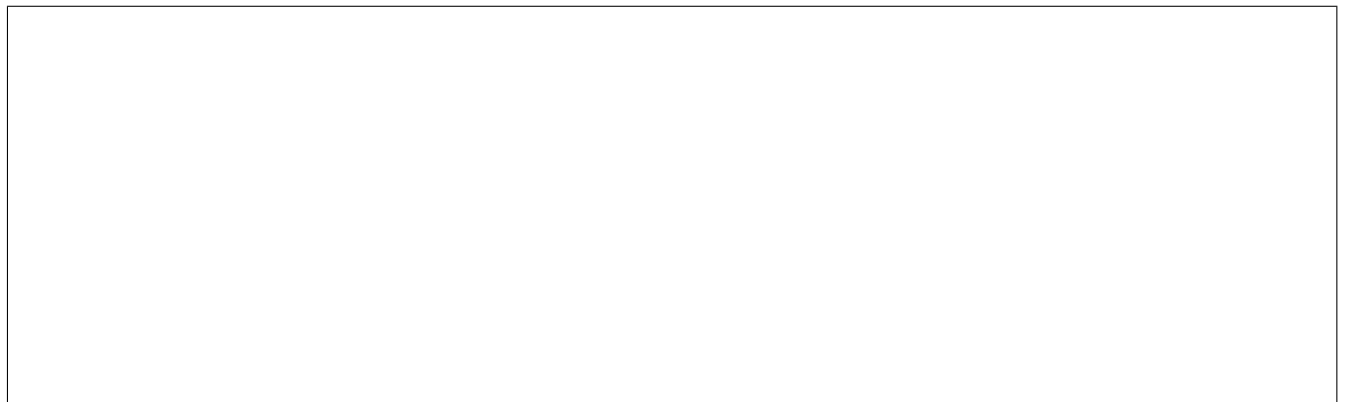
bedeutet, dass $f(x)$ beliebig nah am Grenzwert A gewählt werden kann, wenn man x hinreichend nah (aber nicht gleich) an x_0 wählt.

Zu beachten:

1. Der Punkt, an dem wir den Grenzwert untersuchen, muss nicht teil des Definitionsbereichs sein. Man muss sich diesem Punkt nur beliebig nähern können.
2. Wir sind nicht an $f(x_0)$ interessiert, sondern an $f(x)$ wenn x gegen x_0 geht.
3. Wir müssen x -Werte auf beiden Seiten von x_0 betrachten.
4. Es ist möglich, dass der Grenzwert nicht existiert.

Beispiel 3.1. Wir zeigen hier, dass sich der Grenzwert einer Funktion vom Funktionswert unterscheiden kann. Gegeben sei die Funktion

$$f(x) = \begin{cases} 5 & \text{für } x \neq 0 \\ 1 & \text{für } x = 0 \end{cases}$$



Definition 3.2. : Links- und rechtsseitiger Grenzwert

Um den linksseitigen Grenzwert zu bekommen (sollte er existieren) nähert man sich x_0 mit x -Werten an, die kleiner als x_0 sind. Wir schreiben

$$\lim_{x^- \rightarrow x_0} f(x)$$

Um den rechtsseitigen Grenzwert zu bekommen (sollte er existieren) nähert man sich x_0 mit x -Werten an, die größer als x_0 sind. Wir schreiben

$$\lim_{x^+ \rightarrow x_0} f(x)$$

Wir üben jetzt ein wenig Aussagenlogik, denn wir müssen das Öfteren zwischen notwendigen und hinreichenden Bedingungen unterscheiden. In diesem Fall ist beides gegeben.

Theorem 3.1. Die Gleichheit des links- und rechtsseitigen Grenzwertes ist notwendige und hinreichende Bedingung für die Existenz des Grenzwertes.

1. Notwendig: Wenn der Grenzwert existiert, müssen rechts- und linksseitiger Grenzwert übereinstimmen

$$(\text{Grenzwert von } f(x) \text{ an der Stelle } x_0 \text{ existiert}) \implies \left(\lim_{x^- \rightarrow x_0} f(x) = \lim_{x^+ \rightarrow x_0} f(x) \right)$$

2. Hinreichend: Wenn links- und rechtsseitiger Grenzwert übereinstimmen existiert der Grenzwert

$$\left(\lim_{x^- \rightarrow x_0} f(x) = \lim_{x^+ \rightarrow x_0} f(x) \right) \implies (\text{Grenzwert von } f(x) \text{ an der Stelle } x_0 \text{ existiert})$$

3. Notwendig und hinreichend (\iff):

$$\left(\lim_{x^- \rightarrow x_0} f(x) = \lim_{x^+ \rightarrow x_0} f(x) \right) \iff (\text{Grenzwert von } f(x) \text{ an der Stelle } x_0 \text{ existiert})$$

Notiz: Die Implikation (\implies) wird im Englischen in Beweisen oft mit "if" und die Äquivalenz (\iff) oft mit "iff" (= if and only if) abgekürzt.

Definition 3.3. : Nicht-Existenz von Grenzwerten

1. Der Grenzwert einer Funktion existiert nicht, wenn der linksseitige Grenzwert nicht gleich dem rechtsseitigen ist. Formal existiert der Grenzwert nicht, wenn

$$\lim_{x^- \rightarrow x_0} f(x) \neq \lim_{x^+ \rightarrow x_0} f(x)$$

2. Da ∞ oder $-\infty$ keine Zahlen sind, existiert der Grenzwert nicht, wenn $f(x)$ bei der Annäherung von x an x_0 immer größer (oder kleiner) wird. Formal existiert der Grenzwert nicht, wenn

$$\lim_{x \rightarrow x_0} f(x) = \infty \quad \text{oder} \quad \lim_{x \rightarrow x_0} f(x) = -\infty$$

Abbildung 2 verdeutlicht diese zwei Fälle.

Oft können wir die betreffende Funktion umformen, um den Grenzwert zu bestimmen. Diese Fähigkeit basiert auf dem folgenden Theorem

Theorem 3.2. Wenn die Funktionen f und g gleich sind für alle x in der Nähe von x_0 , aber nicht notwendigerweise für $x = x_0$, dann gilt

$$\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x)$$

immer dann, wenn einer der beiden Grenzwerte existiert.

Beispiel 3.2. : Berechnen sie die Grenzwerte oder zeigen sie, dass der Grenzwert nicht existiert

1. $\lim_{x \rightarrow 2} \frac{3x^2 + 3x - 18}{x - 2}$

2. $\lim_{x \rightarrow 2} \frac{|x - 2|}{x - 2}$

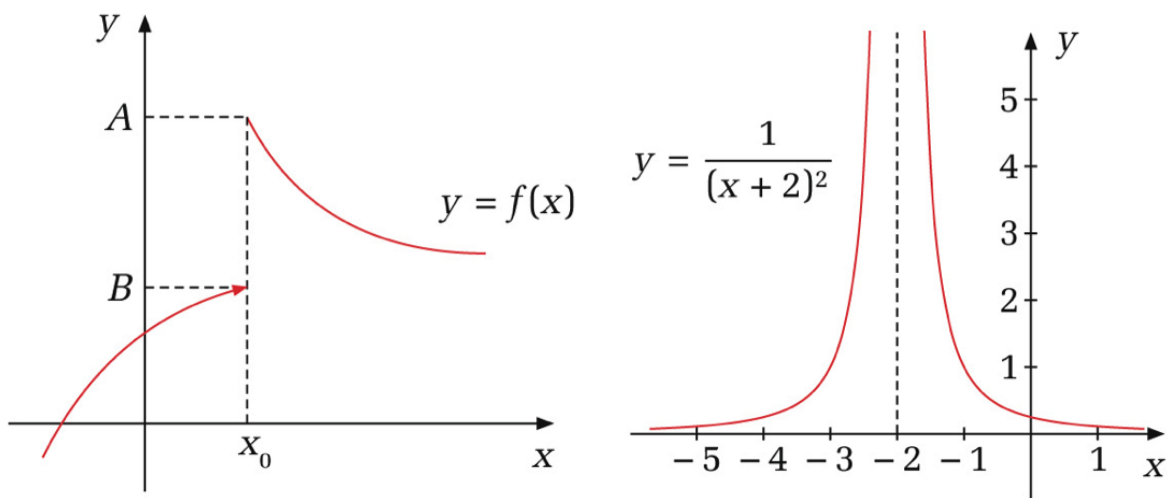
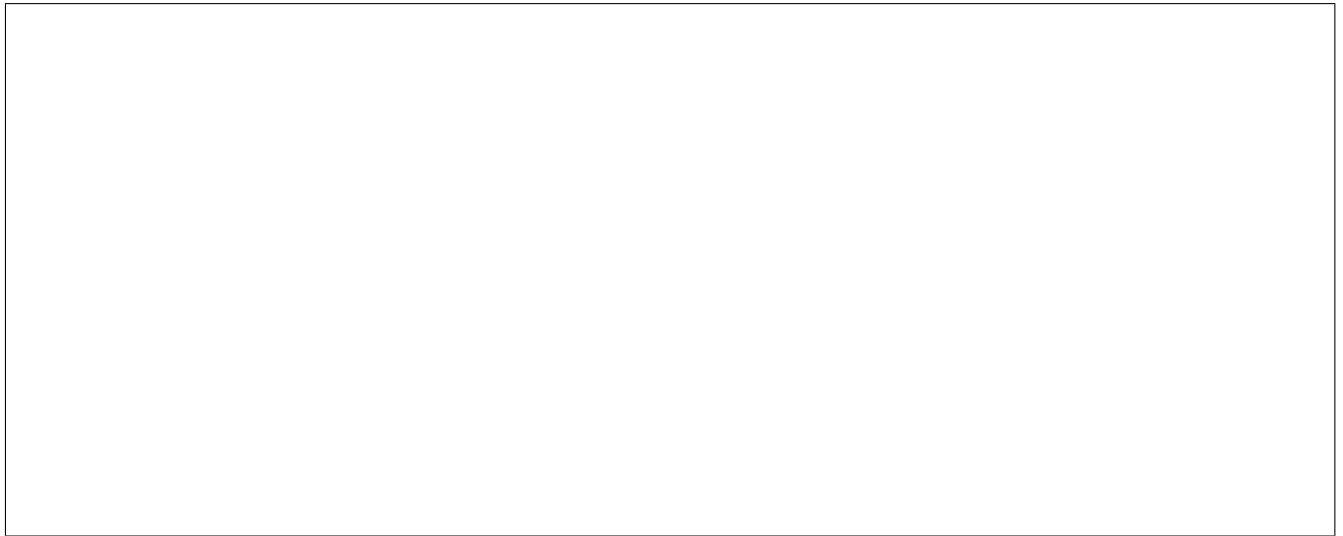


Abbildung 2: Grenzwerte existieren nicht

Grenzwert für reelle Funktionen mehrerer Variablen Wir werden den Fall des funktionalen Grenzwertes für Funktionen mit mehreren Variablen intuitiv behandeln. Eine reelle Funktion mehrerer Variablen $y = f(\mathbf{x})$ ordnet jedem n -dimensionalen Punkt (oder Vektor) in \mathbb{R}^n einen Wert im eindimensionalen Raum \mathbb{R} zu.

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

Der Grenzwert der Funktion (wenn existent) ist also weiterhin eine reelle Zahl. Was sich ändert ist, dass nun ein n -dimensionaler Vektor als Input dient und dass man somit den Funktionswert $f(\mathbf{x})$ betrachtet, wenn sich \mathbf{x} dem Vektor \mathbf{x}_0 nähert.

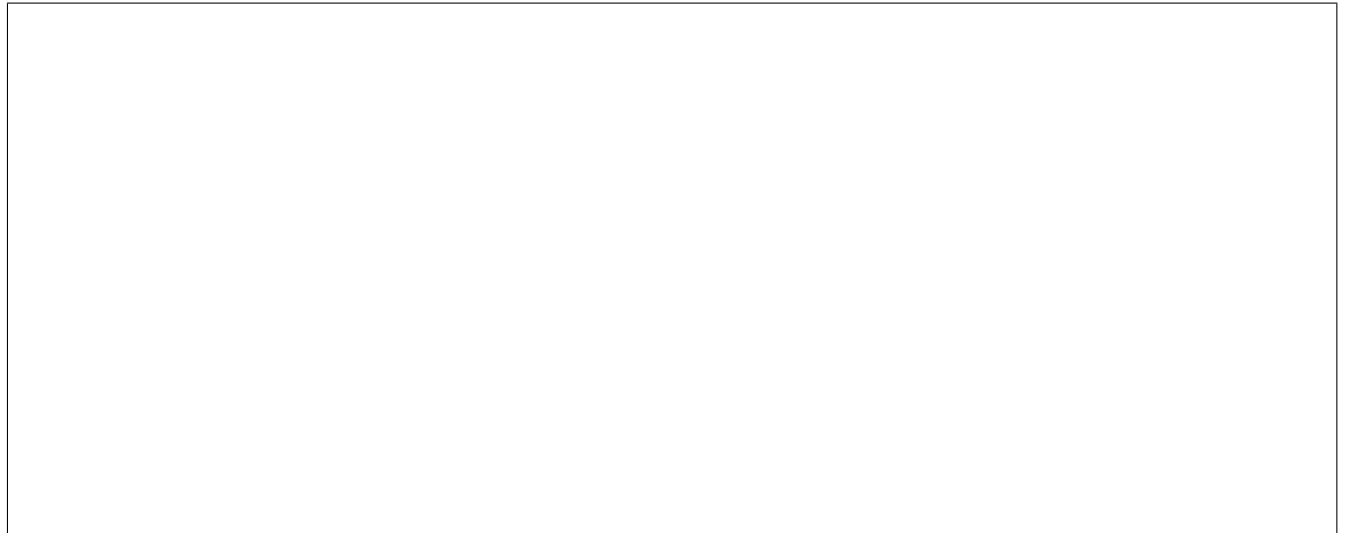
$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x})$$

Definition 3.4. Regeln für Grenzwerte: Wenn $\lim_{x \rightarrow x_0} f(x) = A$ und $\lim_{x \rightarrow x_0} g(x) = B$, dann gilt

1. $\lim_{x \rightarrow x_0} (f(x) \pm g(x)) = A \pm B$
2. $\lim_{x \rightarrow x_0} (f(x) \cdot g(x)) = A \cdot B$
3. $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{A}{B}$ für $B \neq 0$
4. $\lim_{x \rightarrow x_0} [f(x)]^r = A^r$ (A^r ist definiert und $r \in \mathbb{R}$)

Beispiel 3.3. Berechnen sie die folgenden Grenzwerte (c ist eine Konstante)

1. $\lim_{x \rightarrow x_0} c$
2. $\lim_{x \rightarrow 2} x$
3. $\lim_{x \rightarrow -2} (x^2 + 5x)$
4. $\lim_{x \rightarrow 4} \frac{2x^{3/2} - \sqrt{x}}{x^2 - 15}$



3.2 Stetigkeit

Wir werden zunächst Stetigkeit definieren und uns dann den Eigenschaften von stetigen Funktionen zuwenden

3.2.1 Definition

Vom funktionalen Grenzwert bis zur Definition von Stetigkeit ist der Weg nicht weit. Bei einer Funktion von nur einer Variablen haben wir Stetigkeit folgendermaßen definiert:

Definition 3.5. Stetigkeit einer Funktion von einer Variablen:

- *Wenig formal:* Eine Funktion ist stetig, wenn sie ohne absetzen mit einem Stift gezeichnet werden kann.
- *Formaler:* Eine Funktion ist an der Stelle x_0 stetig, wenn

1. Die Funktion an der Stelle x_0 definiert ist.
2. Wenn der Grenzwert der Funktion an der Stelle x_0 existiert.
3. Wenn der Grenzwert der Funktion an der Stelle x_0 gleich $f(x_0)$ ist

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

Eine Funktion heißt stetig wenn sie an allen Punkten ihres Definitionsbereichs stetig ist.

Mann sieht, dass es für Punkte im Definitionsbereich darum geht, ob der Grenzwert gleich dem Funktionswert ist.

Vorsicht. Wie eben beschrieben muss eine Funktion an einem Punkt definiert sein, um an diesem Punkt stetig sein zu können. Das heißt NICHT, dass eine Funktion an einer Stelle, an der die Funktion nicht definiert ist, unstetig ist. An Punkten, an denen die Funktion nicht definiert ist, stellt sich die Frage nach Stetigkeit erst gar nicht. Beispiel:

$$f(x) = \frac{1}{x}$$

Dies ist eine stetige Funktion, da sie an allen Punkten ihres Definitionsbereichs (maximal $D = \mathbb{R} \setminus \{0\}$) stetig ist. Die Tatsache, dass die Funktion an $x_0 = 0$ nicht definiert ist, macht hier keinen Unterschied.

Intuitiv bedeutet dies, dass eine Funktion $y = f(x)$ an einem Punkt x_0 stetig ist, wenn kleine Abweichungen in der unabhängigen Variablen (x) in der Nähe von x_0 auch nur kleine Abweichungen im Funktionswert in der Nähe von $f(x_0)$ verursachen. Die folgenden zwei Graphen verdeutlichen dies.

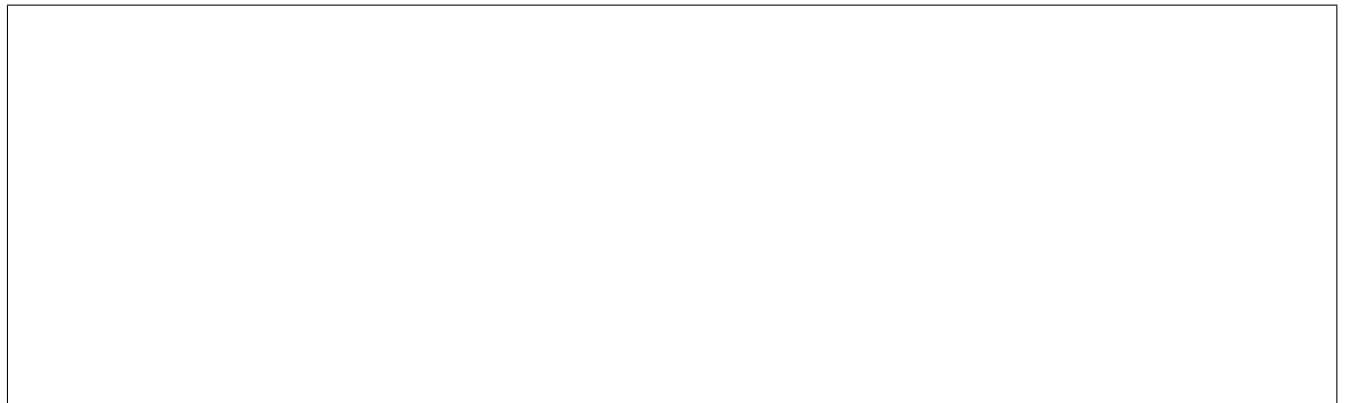
Beispiel 3.4. Sind die folgenden Funktionen an der angegebenen Stelle stetig?

1. $\frac{x-6}{x^3+1}$ an $x_0 = -1$

3. $f(x) = \begin{cases} 5 & x < 0 \\ 1 & x = 0 \\ 5 & x > 0 \end{cases}$ an $x_0 = 0$

2. $f(x) = \begin{cases} 5+3x & x \geq 0 \\ 4-2x & x < 0 \end{cases}$ an $x_0 = 0$

4. $f(x) = \begin{cases} x^2+4 & x \geq 0 \\ 4+x & x < 0 \end{cases}$ an $x_0 = 0$



Stetigkeit von Funktionen Mehrerer Variablen Die Definition von Stetigkeit für eine Funktion mehrerer Variablen kann mit kleinen Änderungen übernommen werden.

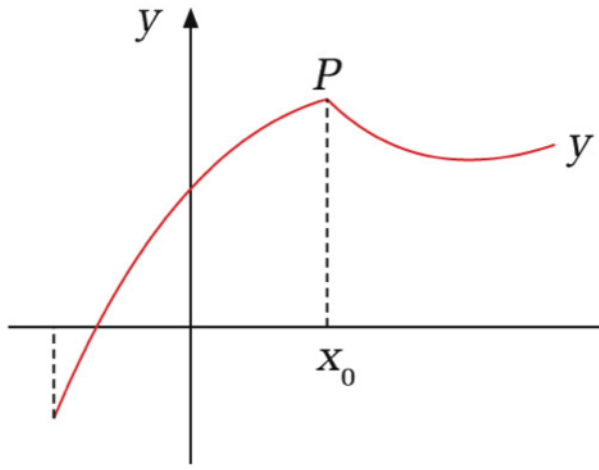


Abbildung 3: Stetige Funktion

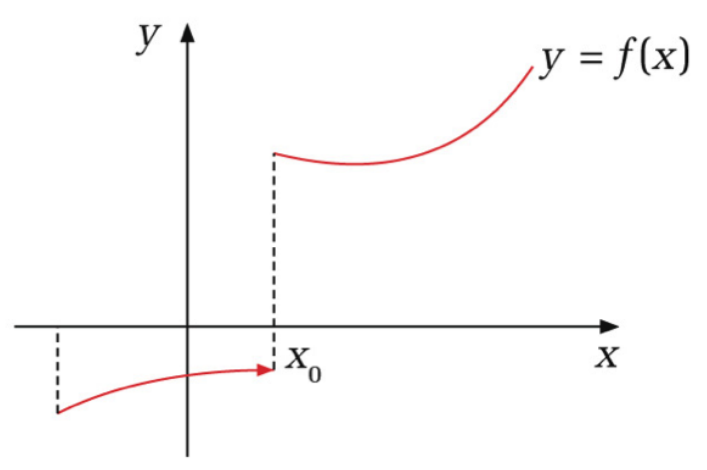


Abbildung 4: Unstetige Funktion

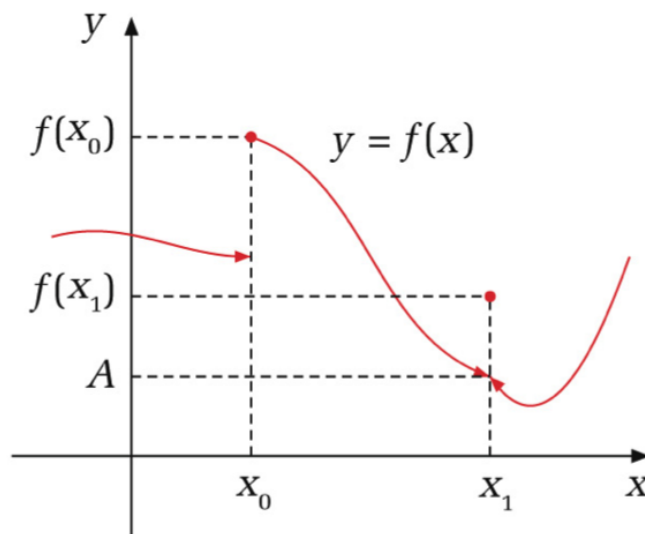


Abbildung 5: Unstetig in zwei Punkten

Definition 3.6. *Stetigkeit einer Funktion mehrerer Variablen: Eine Funktion $f(\mathbf{x})$ ist an der Stelle \mathbf{x}_0 stetig, wenn*

1. Die Funktion am Punkt \mathbf{x}_0 definiert ist.
2. Wenn der Grenzwert der Funktion am Punkt \mathbf{x}_0 existiert.
3. Wenn der Grenzwert der Funktion an der Stelle \mathbf{x}_0 gleich $f(\mathbf{x}_0)$ ist

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$$

Eine Funktion heißt stetig, wenn sie an allen Punkten ihres Definitionsbereichs stetig ist.

3.2.2 Eigenschaften von Stetigen Funktionen

Nachdem in Definition 3.4 die Eigenschaften der Grenzwerte gezeigt wurden und die Definition der Stetigkeit eng mit der Definition des Grenzwertes verbunden ist, sollten die folgenden Eigen-

schaften stetiger Funktionen keine allzu große Überraschung darstellen.

Definition 3.7. Wenn die Funktionen f und g in x_0 stetig sind, so gilt:

1. $f + g$ und $f - g$ sind stetig in x_0
2. $f \cdot g$ und f/g sind stetig in x_0 (für $g \neq 0$)

Beispiel 3.5. Zeigen sie, dass $f(x) = x$ und $g(x) = c$ (c =konstante) stetig sind.

Mit unserem neu gewonnenen Wissen aus Beispiel 3.5 können wir die Eigenschaften in Definition 3.7 nutzen um zu zeigen, dass viele andere Funktionen stetig sind.

Beispiel 3.6. Zeigen sie, dass die folgenden Funktionen stetig sind:

1. $f(x) = 5 + x$
2. $f(x) = x^2$
3. $f(x) = ax^2$
4. $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n$
5. $f(x) = \frac{3-x^2}{5x+6x^3}$

Wir können dieses (bemerkenswerte) Ergebnis wie folgt zusammenfassen:

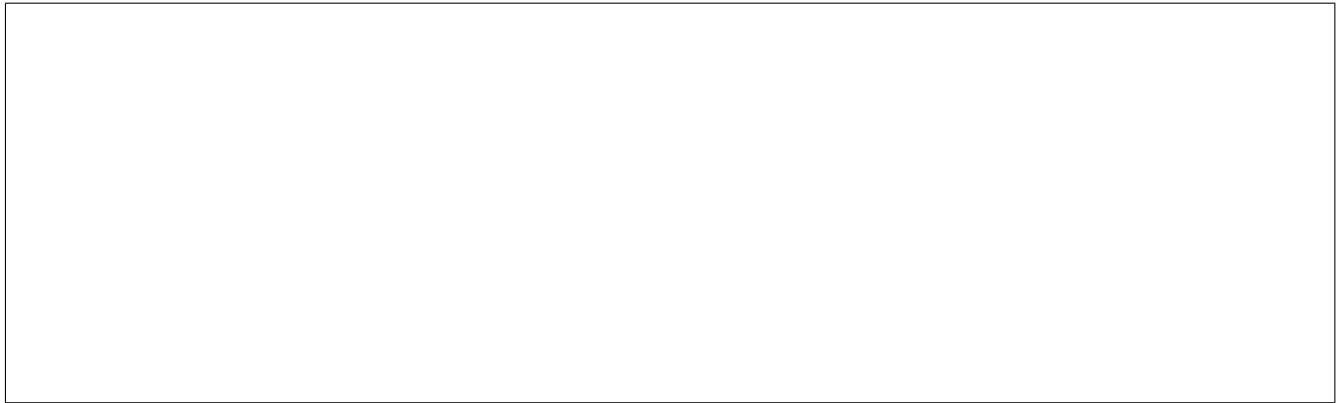
Definition 3.8. Jede Funktion, die aus stetigen Funktionen durch Kombination einer oder mehrerer der folgenden Operationen erzeugt werden kann, ist stetig in allen Punkten, in denen sie definiert ist.

- Addition
- Multiplikation
- Verkettung wie in $f[g(x)]$
- Subtraktion
- Division (wenn definiert)

Beispiel 3.7. Bestimmen sie, für welche Werte $f(x)$ und $g(x)$ stetig sind

1. $f(x) = \frac{x^4+3x^2-1}{(x-1)(x+2)}$

2. $g(x) = (x^2 + 2) \left(x^3 + \frac{1}{x}\right) + \frac{1}{\sqrt{x+1}}$



3.3 Differenzierbarkeit

Jeder Kursteilnehmer kann hoffentlich die erste Ableitung von $f(x) = 3x^2$ bilden¹, aber es ist wahrscheinlich weniger klar, dass auch die Ableitung ein Grenzwert ist. Genauer gesagt ist die Ableitung einer Funktion der Grenzwert des Differenzenquotienten

Definition 3.9. Die Ableitung der Funktion f an der Stelle x_0 wird mit $f'(x_0)$ bezeichnet und ist gegeben durch

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Notiz: $\Delta x = x - x_0$ Wenn dieser Grenzwert existiert, ist die Funktion an x_0 differenzierbar und gibt die Steigung der Tangente der Funktion $f(x)$ am Punkt x_0 an.

Theorem 3.3. Wenn eine Funktion in einem Punkt differenzierbar ist, dann ist sie dort auch stetig.

Abbildung 8 zeigt, dass der Differenzenquotient gleich der Sekantensteigung ist. Die restlichen beiden Abbildungen visualisieren die Konvergenz der Sekantensteigung zur Steigung der Tangente, wenn der Grenzwert existiert.

¹ $f'(x)=6x$

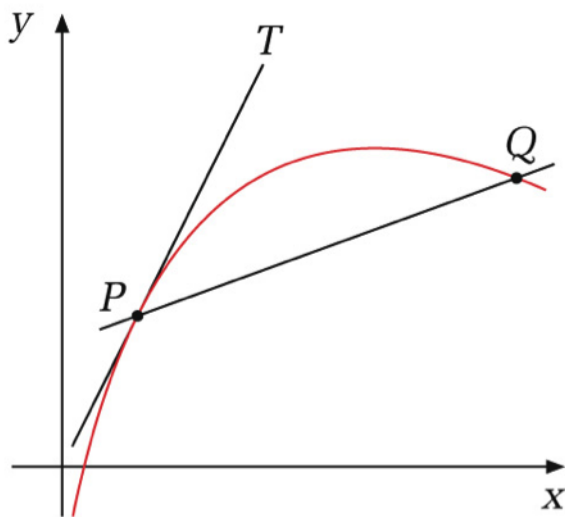


Abbildung 6: Sekante (Q) und Tangente (T)

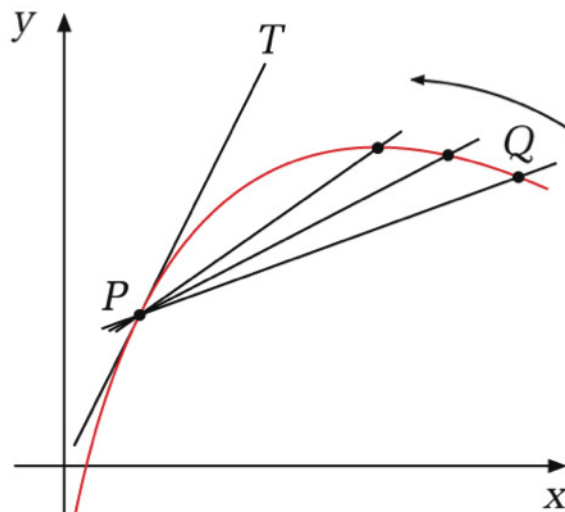


Abbildung 7: Sekanten mit $x \rightarrow x_0$ oder $\Delta x \rightarrow 0$

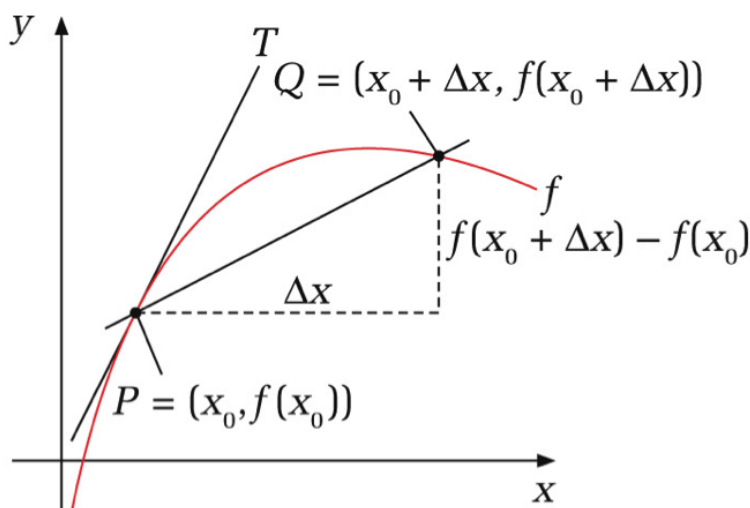


Abbildung 8: Sekantensteigung

Beispiel 3.8. Finden sie folgende Ableitungen (wenn existent)

1. $f(x) = c$ an x_0

3. $f(x) = x^2$ an x_0

2. $f(x) = x$ an x_0

4. $f(x) = |x|$ an $x_0 = 0$

Um den Sprung zur Differenzierung von Funktionen mehrerer Variablen zu machen, nehmen wir an die Ableitung existiert. Somit können wir den Differenzenquotienten schreiben als

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0) + R(\Delta x)$$

Hier repräsentiert $R(\Delta x)$ die Differenz zwischen der Steigung der Sekante und der Ableitung (=Steigung der Tangente) der eine Funktion von Δx ist. Wir formen um zu

$$\frac{f(x_0 + \Delta x) - f(x_0) - f'(x_0)\Delta x}{\Delta x} = R(\Delta x)$$

Wenn die Ableitung existiert, können wir also schreiben

$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0) - f'(x_0)\Delta x}{\Delta x} = 0 \quad (1)$$

Differenzierbarkeit von Funktionen mehrerer Variablen Wir können Gleichung (1) direkt für reelle Funktionen von mehreren Variablen anpassen.

Definition 3.10. Eine reelle Funktion mehrerer Variablen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ist an der Stelle $\mathbf{x}_0 \in \mathbb{R}^n$ differenzierbar, wenn folgendes gegeben ist:

$$\lim_{\Delta \mathbf{x} \rightarrow 0} \frac{f(\mathbf{x}_0 + \Delta \mathbf{x}) - f(\mathbf{x}_0) - \mathbf{a}'\Delta \mathbf{x}}{\|\Delta \mathbf{x}\|} = 0 \quad (2)$$

\mathbf{a} wird als Gradient Vektor bezeichnet und oft mit $\nabla f(\mathbf{x}_0)$ dargestellt.

Wir sehen drei Unterschiede zur Bedingung mit einer Variablen

1. \mathbf{x} und \mathbf{x}_0 sind jetzt Vektoren in \mathbb{R}^n
2. Die "Ableitung" ist nun ebenfalls ein Vektor in \mathbb{R}^n
3. Es wird durch die Norm des Vektors $\Delta \mathbf{x}$ geteilt.

Wir sehen, dass in beiden Fällen (univariat/multivariat) der Zähler der Differenz zwischen dem Funktionswert $f(x_0 + \Delta x)$ (bzw. $f(\mathbf{x}_0 + \Delta \mathbf{x})$) und einer linearen Funktion gleicht.

1. Im univariaten Fall ist diese $f(x_0) + f'(x_0)\Delta x$
2. Im multivariaten Fall $f(\mathbf{x}_0) + a_1\Delta x_1 + a_2\Delta x_2 + \dots + a_n\Delta x_n$

Der nächste Abschnitt behandelt die nun brennende Frage nach der Bedeutung der Konstanten a_1, a_2, \dots, a_n .

3.4 Partielle Ableitungen

Beim Ableiten einer Funktion $f(x)$ waren wir daran interessiert wie sich der Funktionswert verändert, wenn sich die einzige Variable x verändert. Nun nehmen wir an, wir haben eine Funktion von zwei Variablen $f(x_1, x_2)$. Um eine ähnliche Interpretation einer Ableitung dieser Funktion zu bekommen, fragen wir wie sich der Funktionswert verändert, wenn x_1 variiert wird, während x_2 konstant gehalten wird (und umgekehrt). Dies sind die partiellen Ableitungen der Funktion.

Definition 3.11 (Partielle Ableitungen). Gegeben sei die Funktion $y = f(x_1, x_2)$. Dann ist

1. $\frac{\partial y}{\partial x_1}$ (oder f_{x_1} oder $f_{x_1}(x_1, x_2)$ oder $f'_1(x_1, x_2)$ oder....) die partielle Ableitung von $y = f(x_1, x_2)$ nach x_1 , wenn x_2 konstant gehalten wird. Formal

$$f_{x_1}(x_1, x_2) = \lim_{\Delta x \rightarrow 0} \frac{f(x_1 + \Delta x, x_2) - f(x_1, x_2)}{\Delta x}$$

2. $\frac{\partial y}{\partial x_2}$ die partielle Ableitung von $y = f(x_1, x_2)$ nach x_2 , wenn x_1 konstant gehalten wird. Formal

$$f_{x_2}(x_1, x_2) = \lim_{\Delta x \rightarrow 0} \frac{f(x_1, x_2 + \Delta x) - f(x_1, x_2)}{\Delta x}$$

Notiz 3.1. Um die partielle Ableitung nach, zum Beispiel, x_1 zu bilden behandelt man in der Funktionsgleichung alle anderen Variablen als Konstanten und wendet die schon bekannten Regeln der Ableitungen von Funktion mit einer Variablen an. Zum Beispiel für $f(x, y) = xy$ sind die partiellen Ableitungen gleich

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

Beispiel 3.9. Finden sie die partiellen Ableitungen der folgenden Funktionen:

1. $f(x, y) = x^3y + x^2y^2 + x + y^2$

2. $f(x, y) = \frac{xy}{x^2+y^2}$

Partielle Ableitungen haben auch eine geometrische Interpretation die zeigt, dass sie eine direkte Weiterentwicklung der einfachen Ableitung sind. Dazu führen wir die Tangentialebene ein. Zunächst jedoch ein bisserl Wiederholung.

3.4.1 Lineare Funktionen

Wir wissen (hoffentlich), dass eine lineare Funktion einer Variablen die folgende Form hat

$$f(x) = a + bx$$

Hier sind a und b Konstanten. Die Konstante

- a wird Achsenabschnitt genannt, weil sie den Punkt angibt, an dem die Funktion die y -Achse schneidet, also wenn $x = 0$
- b ist die Steigung ($= f'(x)$) der Funktion und ist nicht vom Wert der Variablen x abhängig. Sie kann mit zwei beliebigen Punkten wie folgt berechnet werden

$$f'(x) = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Neben dieser uns bekannten Darstellungsform gibt es noch zwei weitere, die Punkt-Steigungsform und die Zwei-Punkte-Form. Allen Formen lassen sich darauf zurückführen, dass eine lineare Funktion eindeutig bestimmt ist, wenn zwei Informationen vorliegen (2 Punkte, 1 Punkt und Steigung, Achsenabschnitt/Steigung)

Zwei-Punkte-Form Gegeben sind zwei Punkte $(x_1, y_1), (x_2, y_2)$ einer linearen Funktion, dann gilt

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$$

Diese Form resultiert aus der Eigenschaft linearer Funktionen, dass die Steigung unabhängig vom x-Wert ist und somit zwischen beliebigen Punktpaaren gemessen werden kann.

Punkt-Steigung-Form Gegeben sei die ein Punkt (x_1, y_1) und die Steigung b einer linearen Funktion, dann gilt

$$y - y_1 = b(x - x_1)$$

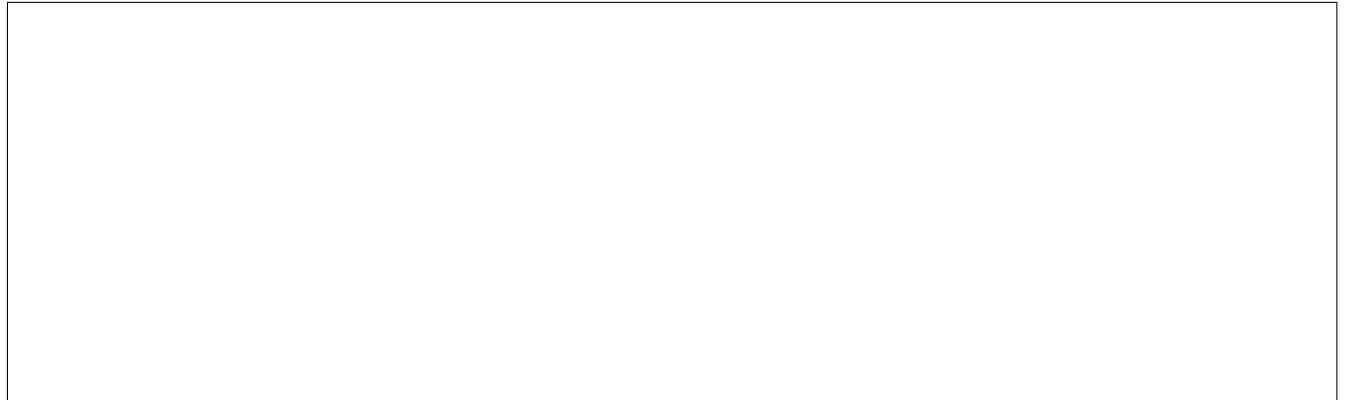
Diese Formulierung basiert auf der Definition der Steigung und der Tatsache, dass die Steigung nicht vom x-Wert abhängig ist.

Wir zeigen nun, dass alle Formulierungen gleich sind

Beispiel 3.10. Gegeben sei die lineare Funktion $y = 1 - 4x$. Nutzen sie die folgenden Informationen um zu zeigen, dass die aus ihnen resultierende linearen Funktionen mit der gegebenen Funktion identisch ist.

1. Punkte $(2, -7)$ und $(-3, 13)$

2. Steigung $b = -4$ und Punkt $(5, -19)$



Eine Ebene mit Steigungen a und b , die durch den Punkt (x_0, y_0, z_0) geht, ist eine lineare Funktion zweier Variablen und wird wie folgt geschrieben

$$z - z_0 = a(x - x_0) + b(y - y_0)$$

Alternativ können wir auch $f(x, y) = z = d + ax + by$ schreiben, wobei $d = z_0 - ax_0 - by_0$ der Achsenabschnitt ist. Ebenen in Räumen mit mehr als 3 Dimensionen werden als Hyperebenen bezeichnet.

3.4.2 Die Tangentialebene

Eine Tangentialebene ist die Verallgemeinerung der Tangenten an einem Punkt einer Funktion. In Abbildung 9 wird ein Punkt in \mathbb{R}^3 gezeigt. Beim Graph einer Funktion mit zwei Variablen wird in der dritten Dimension der Funktionswert abgetragen. Punkt P in Abbildung 10 ist also

$$(x_0, y_0, f(x_0, y_0))$$

Hier sind weiterhin mit K_x und K_y Funktionen eingezeichnet, bei denen eine der beiden Variablen (x bei K_x) konstant gehalten wird. Die partielle Ableitung der Funktion nach x am Punkt P ist somit

gleich der Steigung der Tangente I_y . Die partielle Ableitung der Funktion nach y am Punkt P ist gleich der Steigung der Tangente I_x . Abbildung 11 zeigt nun die Tangentialebene, die die Menge aller möglichen Tangenten am Punkt darstellt. Die Tangentialebene hat die Formel²

$$z = z_0 + f_x(x - x_0) + f_y(y - y_0)$$

Diese Formulierung sollte uns bekannt vorkommen. Im Zähler des Grenzwertes in Definition 3.10 sieht man die Differenz zwischen dem Funktionswert $f(\mathbf{x}_0 + \Delta\mathbf{x})$ und dem Funktionswert der Tangentialebene im Punkt (x_1, x_2, \dots, x_n) . Der Gradientenvektor \mathbf{a} ist also der Vektor der partiellen Ableitungen der Funktion.

$$\mathbf{a} = \nabla f(\mathbf{x}_0) = \begin{bmatrix} f_{x_1}(\mathbf{x}_0) \\ f_{x_2}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{bmatrix}$$

3.5 Partielle Ableitung Höherer Ordnung

Die partiellen Ableitung, insofern sie differenzierbar sind, können nun wieder abgeleitet werden. Interessanterweise kann sich im Falle von $f(x, y)$ zum Beispiel der Wert der partiellen Ableitung f_x ändern, wenn sich entweder x oder y (oder beide) verändern. Somit gibt es im Falle zweier unabhängigen Variablen vier partielle Ableitungen der Funktion $f(x, y)$.

- | | | | |
|-------------------|-------------------|-------------------|-------------------|
| 1. $f_{xx}(x, y)$ | 2. $f_{xy}(x, y)$ | 3. $f_{yy}(x, y)$ | 4. $f_{yx}(x, y)$ |
|-------------------|-------------------|-------------------|-------------------|

Beispiel 3.11. Berechnen sie die partiellen Ableitungen zweiter Ordnung der folgenden Funktionen:

- | | |
|--|--------------------------|
| 1. $f(x, y) = x^3y + x^2y^2 + x + y^2$ | 2. $f(x, y) = x^5 \ln y$ |
|--|--------------------------|

Uns fällt auf, dass im obigen Beispiel

$$f_{xy}(x, y) = f_{yx}(x, y)$$

Dies ist kein Zufall, sondern trifft immer zu wenn alle partiellen Ableitungen der Funktion stetig sind.

Theorem 3.4. *Youngs Theorem: Angenommen wird, dass alle partiellen Ableitungen m -ter Ordnung der Funktion*

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

²Die Tangentialebene ist eine lineare Approximation und wird in Abschnitt 4.5 näher behandelt

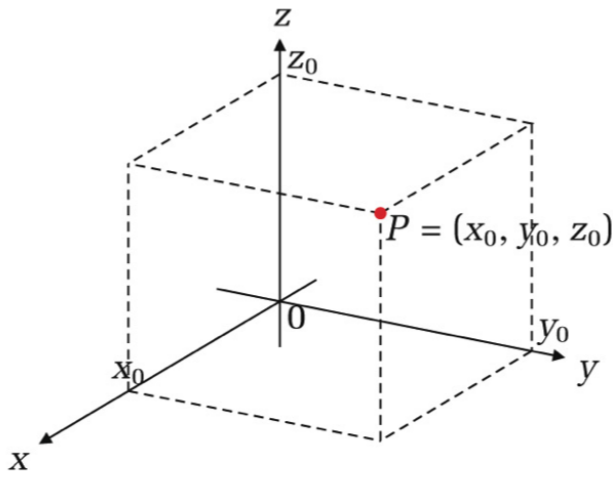


Abbildung 9: Punkt in \mathbb{R}^3

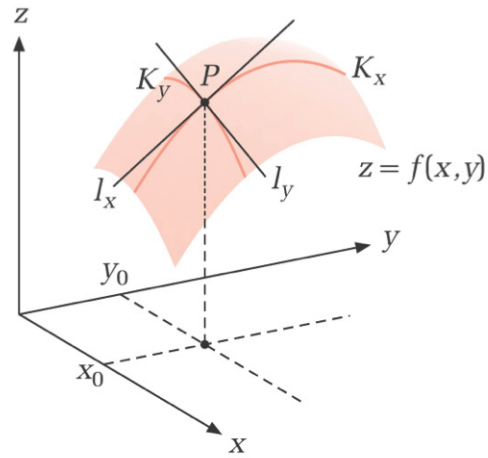


Abbildung 10: Tangenten in x- und y-Richtung

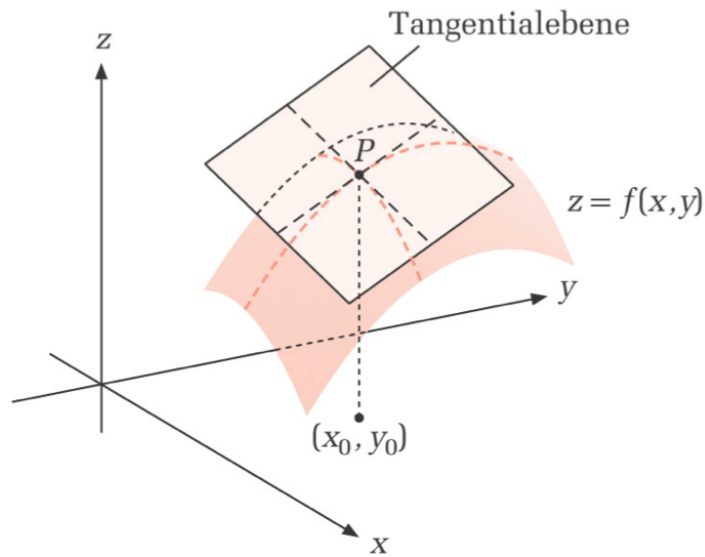


Abbildung 11: Tangentialebene

stetig sind. Wenn zwei von ihnen bezüglich jeder der Variablen die gleiche Anzahl von Differentiationen verlangen, dann sind sie notwendigerweise gleich. Sind für $f(x, y)$ die Bedingungen erfüllt gilt somit

$$f_{xy}(x, y) = \frac{\partial^2 f(x, y)}{\partial x \partial y} = \frac{\partial^2 f(x, y)}{\partial y \partial x} = f_{yx}(x, y)$$

Wir können nun die partiellen Ableitungen zweiter Ordnung in einer Matrix zusammen fassen

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

Definition 3.12 (Hesse Matrix). Gegeben sei die Funktion $f(\mathbf{x})$. Die Hesse Matrix $H(x, y)$ ist die Matrix aller partiellen Ableitungen zweiter Ordnung.

$$H(\mathbf{x}) = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & f_{x_1x_3} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & f_{x_2x_3} & \cdots & f_{x_2x_n} \\ f_{x_3x_1} & f_{x_3x_2} & f_{x_3x_3} & \cdots & f_{x_3x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & f_{x_nx_3} & \cdots & f_{x_nx_n} \end{bmatrix}$$

Wir beschließen den Abschnitt mit einer weiteren Definition

Definition 3.13 (Stetig differenzierbare Funktionen oder C^k Funktionen). *Wenn die Funktion $f(\mathbf{x})$ stetige partielle Ableitungen in ihrem Definitionsbereich hat, nennen wir sie stetig differenzierbar oder eine C^1 Funktion. Wenn alle partiellen Ableitungen bis zur Ordnung k existieren und stetig sind, wird f eine C^k Funktion genannt.*

4 Mathematisches Handwerkszeug

4.1 Kettenregeln

Wir erinnern uns an die Kettenregel bei Funktionen von einer Variablen.

Definition 4.1. Kettenregel bei einer Funktion von einer Variablen. Gegeben sei die Funktion $f[g(x)]$. Die Ableitung der Funktion nach x ist gleich

$$\frac{df[g(x)]}{dx} = f'[g(x)]g'(x)$$

Beispiel 4.1. Finden sie die erste Ableitung der folgenden Funktionen:

1. $f(x) = (3x + 1)^2$

2. $f(x) = \ln(17 - x)$

Wir erweitern nun die Kettenregel für den bivariaten Fall. Gegeben ist nun die Funktion $z = f(x, y)$, wobei x und y selbst Funktionen von einer Variablen t sind.

$$x = g(t) \quad y = h(t)$$

Intuitiv verändert sich der Funktionswert z , wenn sich x verändert oder wenn sich y verändert. Die Werte dieser Variablen hängen wiederum von t ab. Somit ist der Effekt von t auf den Funktionswert z gleich der Summe aus

1. dem Effekt von x auf z multipliziert mit dem Effekt von t auf x
2. dem Effekt von y auf z multipliziert mit dem Effekt von t auf y

Wir definieren

Definition 4.2. Kettenregel für Funktionen von zwei Variablen. Gegeben sei $z = f(x, y)$ mit $x = g(t)$ und $y = h(t)$, dann gilt

$$\frac{df(x, y)}{dt} = f_x(x, y) \frac{dx}{dt} + f_y(x, y) \frac{dy}{dt}$$

Diese Ableitung heißt totale Ableitung von z nach t .

Beispiel 4.2. Finden sie die totale Ableitung der folgenden Funktionen:

1. $f(x, y) = x^2 + y^3$ mit $x = t^2$ und $y = 2t$

2. $f(x, y) = \ln\left(\frac{x^2}{y}\right)$ mit $x = e^t$ und $y = t^2$

Wir können diese Definition auch auf allgemeinere Fälle ausweiten, wie die folgende Definition zeigt

Definition 4.3. Kettenregel. Gegeben sei die Funktion $z = f(x, y)$ mit

$$x = g(s, t) \quad \text{und} \quad y = h(s, t)$$

Die Ableitungen der Funktion nach s und t sind.

$$\frac{\partial f}{\partial s} = f_x(x, y) \frac{\partial x}{\partial s} + f_y(x, y) \frac{\partial y}{\partial s}$$

$$\frac{\partial f}{\partial t} = f_x(x, y) \frac{\partial x}{\partial t} + f_y(x, y) \frac{\partial y}{\partial t}$$

Notiz: Man beachte, dass es sich hier partielle Ableitungen handelt, da wir z.B. bei der ersten Ableitung die Variable t konstant halten.

Beispiel 4.3. Lösen sie die folgende Aufgabe:

1. Bestimmen sie $\frac{\partial z}{\partial t}$ und $\frac{\partial z}{\partial s}$, wenn $z = f(x, y)$ mit $x = t - s^2$ und $y = ts$

Zum Schluss dieses Abschnitts sei noch der Allgemeine Fall erwähnt, der aber eine direkte Erweiterung der bisher behandelten Fälle darstellt.

Definition 4.4. Wenn $z = f(\mathbf{x})$ stetig differenzierbar ist und $x_i = g(t_1, t_2, \dots, t_m)$ für jedes $i = 1, 2, \dots, n$ differenzierbar ist, dann gilt

$$\frac{\partial z}{\partial t_i} = \frac{\partial f(\mathbf{x})}{\partial x_1} \frac{\partial x_1}{\partial t_i} + \frac{\partial f(\mathbf{x})}{\partial x_2} \frac{\partial x_2}{\partial t_i} + \dots + \frac{\partial f(\mathbf{x})}{\partial x_n} \frac{\partial x_n}{\partial t_i}$$

4.2 Höhenlinien und deren Steigung

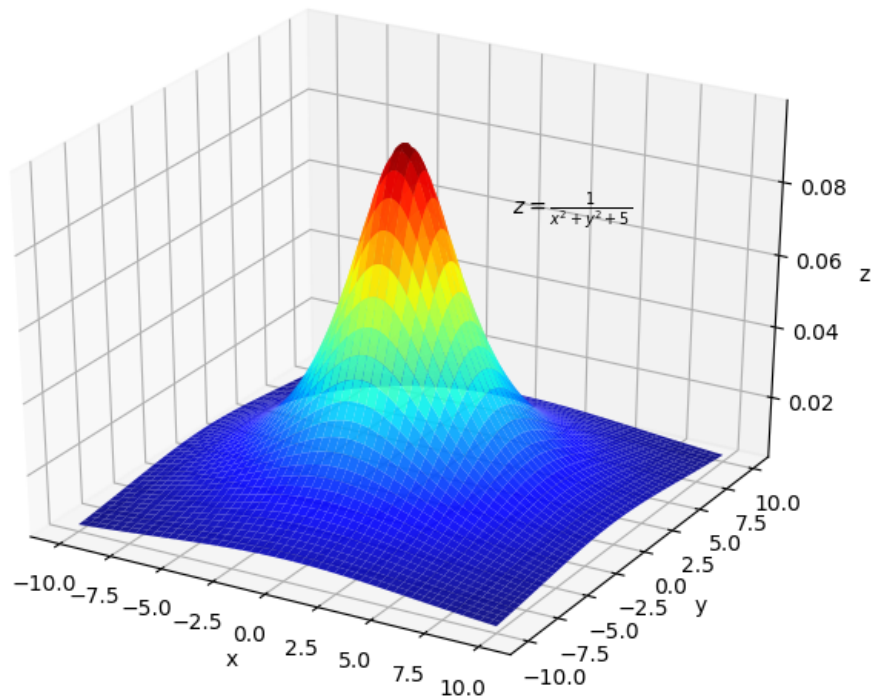
Wir führen zunächst den Begriff der Höhenlinie ein und leiten danach deren Steigung her.

4.2.1 Höhenlinien

Wenn wir eine Funktion mit zwei Variablen haben, können wir diese als Graph in drei Dimensionen darstellen. Zum Beispiel

$$z = \frac{1}{x^2 + y^2 + 5}$$

Abbildung 12



Weitere Beispiele sind

$$z = \frac{1}{x^2 + y^2 + 1} \cos(\pi x) \cos(\pi y) \quad \text{und} \quad z = x^2 - y^2$$

Abbildung 13

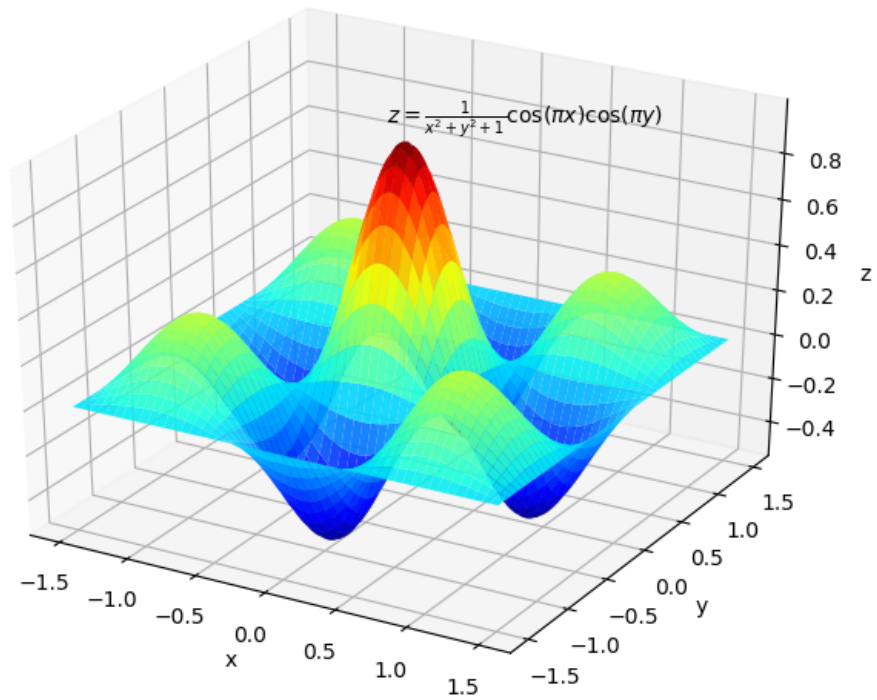
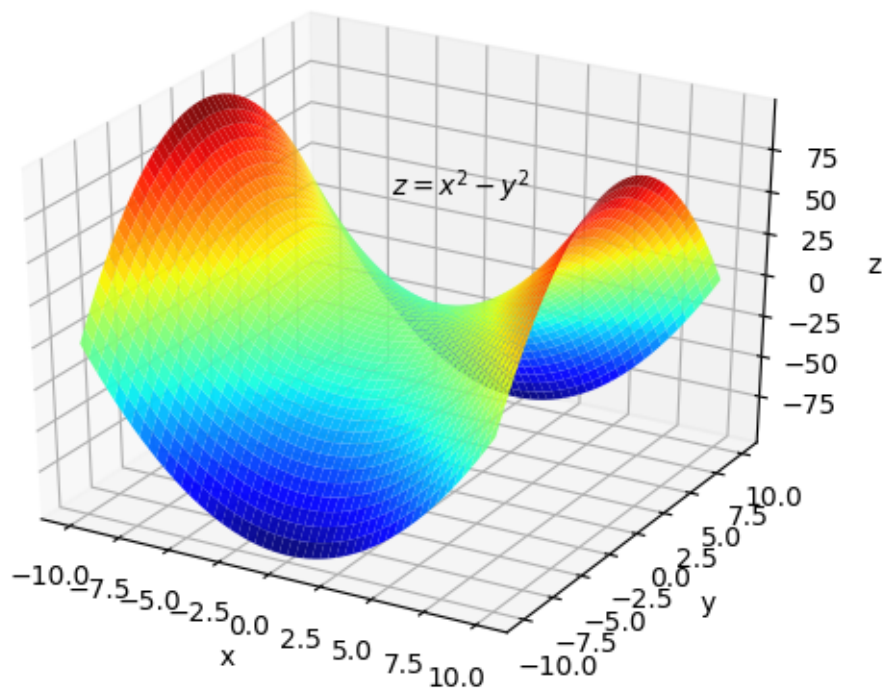


Abbildung 14: Ein Sattelpunkt bei $(0,0)$



Definition 4.5 (Höhenlinien). Eine Höhenlinie der Funktion $z = f(x, y)$ ist die Menge aller Kombinationen von x und y im Definitionsbereich, die den selben Funktionswert ergeben. Alle Elemente der Menge

lösen die Gleichung

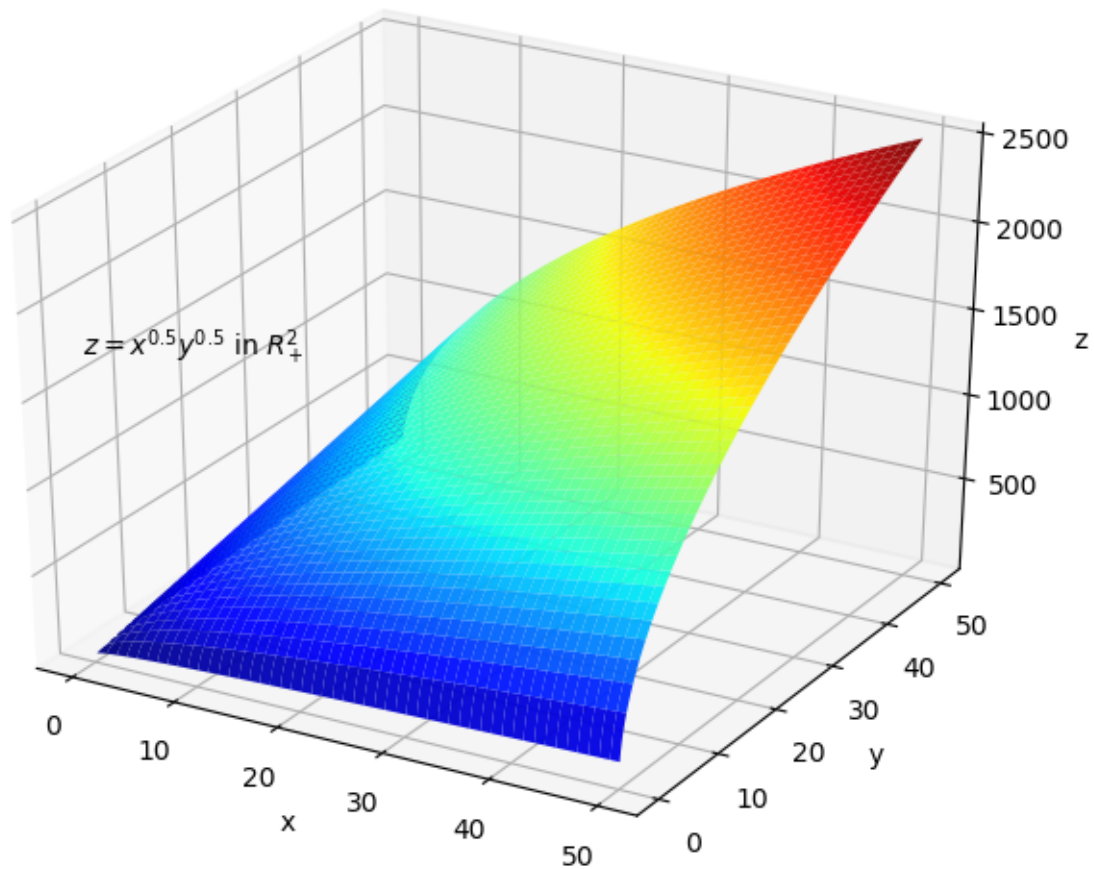
$$f(x, y) = c$$

Graphisch und intuitiv können wir Höhenlinien also als Schnittmenge einer Funktion im dreidimensionalen Raum und einer Ebene, die parallel den x - und y -Achsen verläuft, zeigen. Diese Ebene repräsentiert einen gegebenen Funktionswert. Wir benutzen die in der VWL oft verwendete Cobb-Douglas Produktionsfunktion

$$z = x^{0.5}y^{0.5}$$

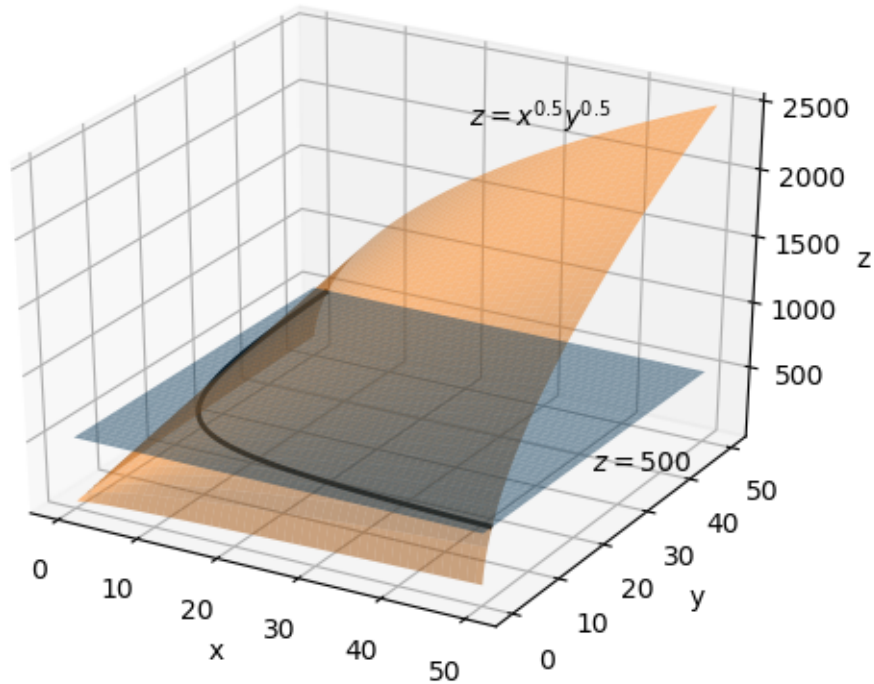
die für nicht-negative Werte von x und y folgenden Graphen hat.

Abbildung 15: Die Funktion $z = x^{0.5}y^{0.5}$



Wenn wir nun eine Ebene in den Graphen integrieren (hier mit $z = 500$), ergibt sich der nächste Graph und die Schnittmenge zwischen den beiden Graphen ist die Höhenlinie.

Abbildung 16: Höhenlinie der Funktion $z = x^{0.5}y^{0.5}$ für $z = 500$



Höhenlinien begegnen uns in der VWL sehr oft:

1. Indifferenzkurven sind die Kombinationen von Gütern, die zu einem gegebenen Nutzenniveau führen
2. Isoquanten sind die Mengen an Arbeit und Kapital, die in einer gegebenen Produktionsmenge resultieren.

Indifferenzkurven oder Isoquanten werden meist nur 2-dimensional dargestellt. Wie das geschieht zeigen die nächsten zwei Graphen. Im ersten werden die Höhenlinien der Cobb-Douglas Funktion gezeigt. Im zweiten Graphen werden diese Linien auf die x/y -Ebene projiziert (oder: man guckt von oben auf den 3-dimensionalen Graphen).

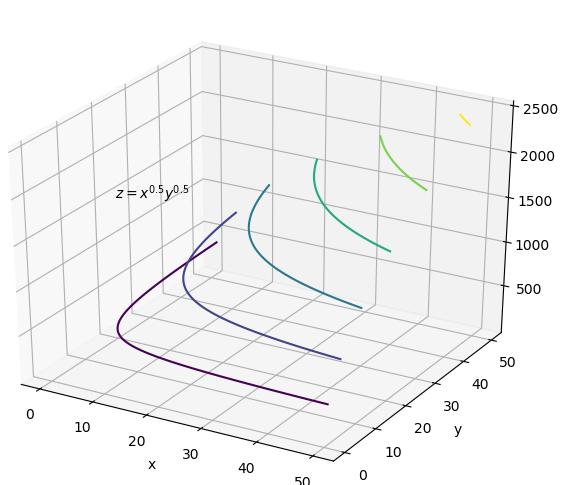


Abbildung 17: Höhenlinien in 3D

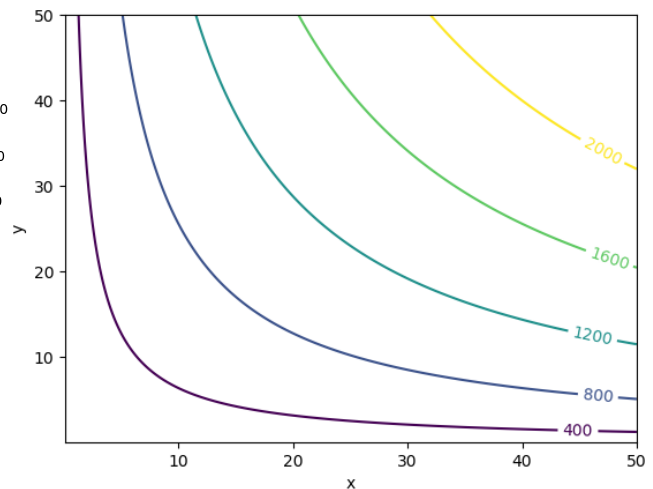


Abbildung 18: Höhenlinien in 2D

Zum Schluss noch ein Rätsel. Zu welchen Funktionen gehören die folgenden Höhenlinien?

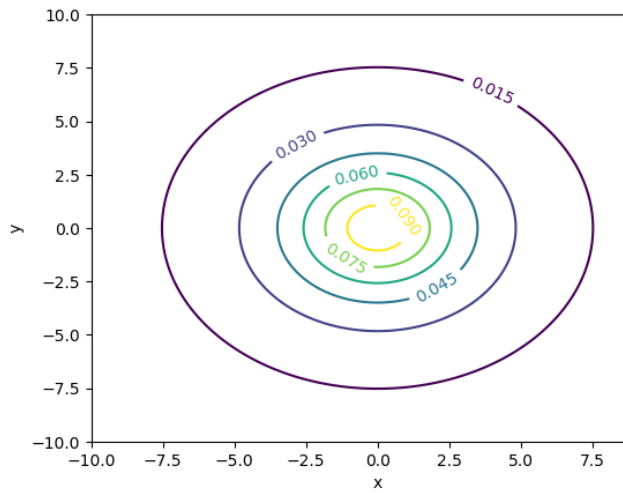


Abbildung 19: Höhenlinien in 3D

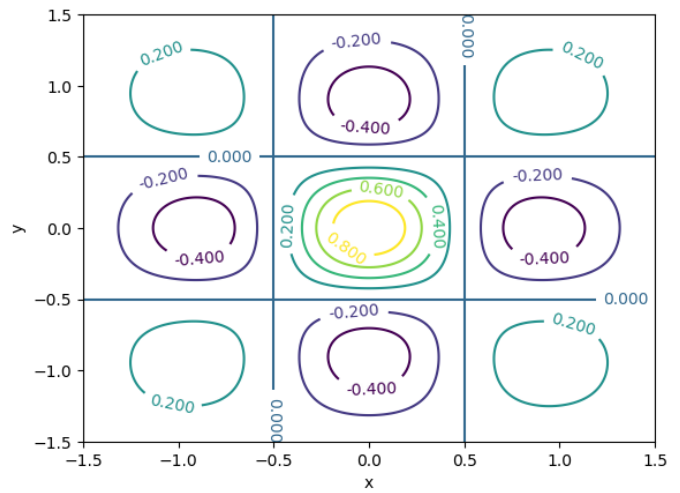


Abbildung 20: Höhenlinien in 2D

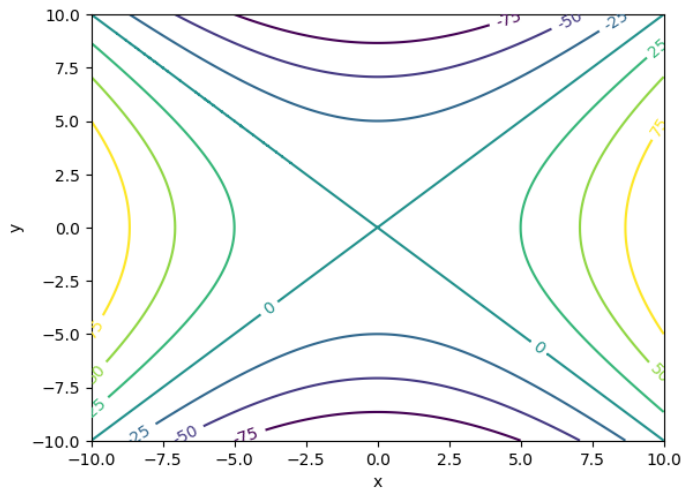
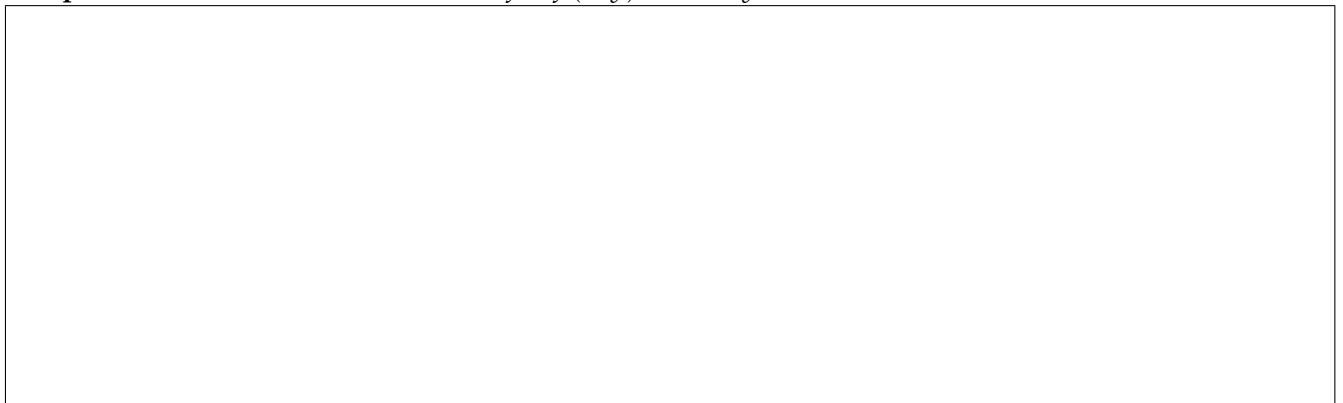


Abbildung 21: Höhenlinien in 2D

Beispiel 4.4. Zeichnen sie die Höhenlinie für $f(x, y) = x^2 + y^2 = 4$



4.3 Implizites Differenzieren

Wir werden zunächst das Beispiel einer Höhenlinie verwenden und dann den allgemeinen Fall behandeln.

4.3.1 Impliziertes Differenzieren entlang einer Höhenlinie

Wir erinnern uns, dass entlang einer Höhenlinie der Funktionswert konstant ist, also zum Beispiel für den Fall von zwei Variablen

$$F[x, y] = c$$

Dies bedeutet, dass entlang der Höhenlinie bei einer Veränderung von x immer ein y gefunden werden kann, so dass sich der Funktionswert nicht ändert. Das heißt, auch wenn es explizit nicht möglich sein sollte die Gleichung nach y aufzulösen, besteht doch *implizit* ein Funktion zwischen den beiden Variablen.

$$y = f(x) \quad \text{mit} \quad F[x, f(x)] = c$$

Wir können also auf der Höhenlinie F nur als Funktion von x schreiben. Was passiert nun, wenn wir die Funktion $F[x, f(x)]$ nach x ableiten?

$$\frac{dF[x, f(x)]}{dx} = F_x \frac{dx}{dx} + F_y \frac{dy}{dx} = 0$$

Wenn wir nun umformen erhalten wir ein wichtiges (und vielleicht auch überraschendes Ergebnis). Wenn F und f differenzierbar sind gilt:

$$\frac{dy}{dx} = -\frac{F_x}{F_y} \quad \text{mit} \quad F_y \neq 0$$

Obwohl keine explizite Funktion besteht, kann die Ableitung von y nach x entlang der Höhenlinie gefunden werden. Die Variablen x und y können natürlich auch vertauscht werden, so dass gilt

$$\frac{dx}{dy} = -\frac{F_y}{F_x} \quad \text{mit} \quad F_x \neq 0$$

Die Steigung der Höhenlinie in einem Punkt ist also gleich dem negativen Quotienten der partiellen Ableitungen! Dieses Resultat wird in der VWL viel benutzt, zum Beispiel bei der Berechnung der Steigungen von Indifferenzkurven und Isoquanten.

Beispiel 4.5. Berechnen sie die Steigung der Höhenlinie gegeben durch

1. $yx = 5$ im Punkt $(2,1)$

2. $x^3 + x^2y - 2y^2 - 10y = 0$ im Punkt $(2,1)$.

Notiz 4.1. In der Literatur wird oft der allgemeine Fall $F(x, y) = 0$ benutzt. Dies steht im Einklang des bisher behandelten, da man immer umformen kann

$$F(x, y) = c \implies G(x, y) = F(x, y) - c = 0$$

4.3.2 Implizites Differenzieren: Der allgemeine Fall

Man kann dieses Konzept um weitere Variablen erweitern. Bevor wir das allgemeine Resultat präsentieren diskutieren wir noch den Fall mit drei Variablen $F(x, y, z) = 0$. Hier kann nun z als Funktion der beiden anderen Variablen geschrieben werden und es gilt

$$F(x, y, f(x, y)) = 0$$

Wenn F und f differenzierbar sind erhalten wir die folgenden zwei partiellen Ableitungen

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z} \quad \text{mit } F_z \neq 0$$

$$\frac{\partial z}{\partial y} = -\frac{F_y}{F_z} \quad \text{mit } F_z \neq 0$$

Der allgemeine Fall folgt dann als direkte Erweiterung

Definition 4.6. Implizites Differenzieren. Gegeben sei die Funktion

$$F(x_1, x_2, \dots, x_n, z) = 0$$

. Wenn F und $z = f(x_1, x_2, \dots, x_n)$ differenzierbar sind, gilt

$$\frac{\partial z}{\partial x_i} = -\frac{F_{x_i}}{F_z} \quad \text{für alle } i = 1 \dots n \quad \text{mit } F_z \neq 0$$

Beispiel 4.6. Berechnen sie folgendes:

1. Gegeben ist die Gleichung $\sqrt{x} + y^2 = x^2 + 5 - y$. Finden sie die Ableitung $\frac{\partial y}{\partial x}$.

2. Gegeben ist die Gleichung

$$\frac{h}{m(\theta)} = \frac{(y - w)}{r + q}$$

, wobei $m(\theta)$ eine Funktion von θ ist. Finden sie die Ableitung $\frac{\partial w}{\partial \theta}$

4.4 Homogene Funktionen

Homogenität lässt sich am besten am Beispiel der ökonomischen Produktionsfunktion erklären.

In ökonomischen Modellen wird die Produktionsmenge (Y) meistens in Abhängigkeit von den eingesetzten Mengen an (Real)Kapital und Arbeit modelliert. Allgemein schreiben wir

$$Y = F(K, L)$$

wobei L für den Faktor Arbeit (Labor) steht. Man kann nun Fragen, inwiefern sich die Produktionsmenge verändert, wenn beide Produktionsfaktoren (K und L) mit der gleichen positiven Konstanten (t) multipliziert werden. Wie verändert sich zum Beispiel die Produktionsmenge, wenn beide Produktionsfaktoren verdoppelt werden?

Definition 4.7. Homogenität vom Grad k . Eine Funktion $f(x, y)$ von zwei Variablen x und y , mit Definitionsbereich D , heißt homogen vom Grad k , wenn für alle $(x, y) \in D$ gilt

$$f(tx, ty) = t^k f(x, y) \quad \text{für alle } t > 0$$

In Worten bedeutet dies, dass Multiplikation beider Variablen mit einem positiven Faktor t den Wert der Funktion mit dem Faktor t^k multiplizieren wird.

Als Beispiel führen wir die allgemeine Cobb-Douglas Produktionsfunktion ein

Definition 4.8. Eine Cobb-Douglas Produktionsfunktion zweier Variablen hat die Form

$$z = f(x, y) = Ax^\alpha y^\beta$$

, wobei α und β Konstanten sind.

Beispiel 4.7. Homogenität der Cobb-Douglas Produktionsfunktion. Um herauszufinden, ob eine Funktion homogen ist (und von welchem Grad) multiplizieren wir beide Variablen mit $t > 0$. In unserem Falle ergibt dies

$$\begin{aligned} f(tx, ty) &= A(tx)^\alpha (ty)^\beta \\ &= At^\alpha x^\alpha t^\beta y^\beta \\ &= t^{\alpha+\beta} Ax^\alpha y^\beta \\ &= t^{\alpha+\beta} f(x, y) \end{aligned}$$

Die Cobb-Douglas Produktionsfunktion ist also homogen vom Grade $\alpha + \beta$.

Beispiel 4.8. Entscheiden sie, ob die folgenden Funktionen homogen sind. Wenn ja, berechnen die den Grad der Homogenität.

1. $z = 5x^2y^2 + xy^3$

3. $y = K^2$

2. $z = x^2y + xy$

4. $y = K^\alpha + L^\beta$



4.5 Approximationen

Es ist oft der Fall, dass ein theoretisches Modell zu kompliziert ist und deshalb nur näherungsweise (approximativ) gelöst werden kann. Hochkomplexe nicht-lineare Gleichungen werden dabei durch approximative Gleichungen ersetzt, die nur lokal (= in einer kleinen Umgebung eines Punktes) vergleichbare Werte wie die Originalfunktion liefern. Wir beschäftigen uns hier mit zwei dieser Approximationen: der linearen und der polynomialen Approximation.

4.5.1 Lineare Approximation

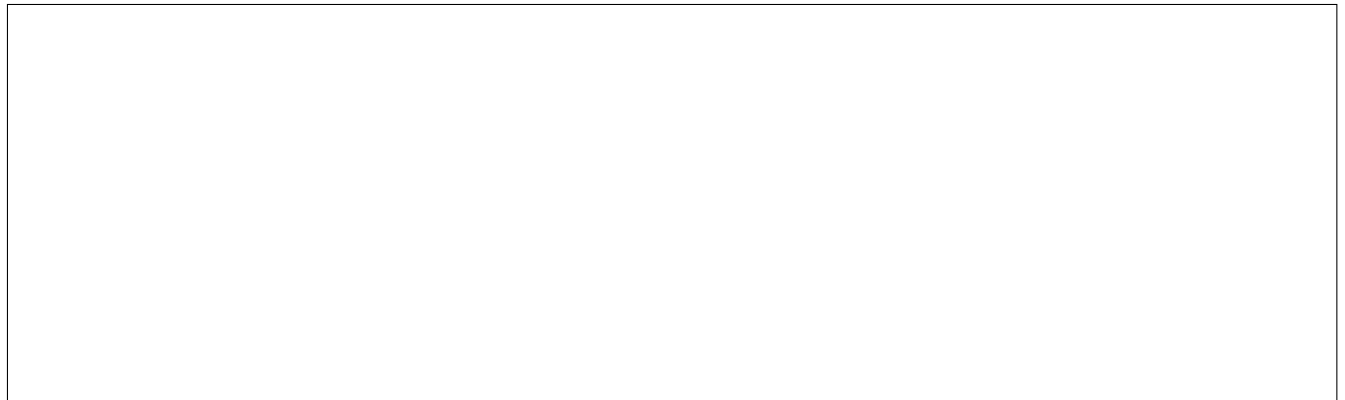
Bei der linearen Approximation einer Funktion an einem Punkt x_0 wird der Graph an der Stelle x_0 mit seiner Tangente approximiert. Die Tangente der (differenzierbaren) Funktion $f(x)$ am Punkt x_0 hat die Gleichung $p(x) = f(x_0) + f'(x_0)(x - x_0)$

Definition 4.9. Bei der Linearen Approximation einer Funktion $f(x)$ um den Punkt x_0 gilt für Werte von x in der Nähe von x_0

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

Beispiel 4.9. :

1. Approximieren sie die Funktion $y = \ln x$ um den Punkt $x_0 = 1$ und zeigen sie, dass der Approximationsfehler mit zunehmender Distanz von x_0 steigt. (Notiz: $\ln(1.05) = 0.04879$ und $\ln(1.1) = 0.09531$)
2. Zeigen sie mit der selben Approximation Zeigen, dass $\ln x_2 - \ln x_1 \approx \frac{x_2 - x_1}{x_1}$



Natürlich kann auch eine Funktion von mehreren Variablen linear approximiert werden. Gegeben sei eine Funktion von zwei Variablen $f(x, y)$. Wir approximieren diese Funktion um einen Punkt (x_0, y_0) nicht mit einer Tangente, sondern mit einer Tangentialebene im 3-dimensionalen Raum, die wir im Abschnitt 3.4.2 behandelt haben

Definition 4.10. Bei der linearen Approximation einer Funktion $z = f(x, y)$ um den Punkt (x_0, y_0) gilt für Werte von (x, y) in der Nähe von (x_0, y_0)

$$f(x, y) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Beispiel 4.10. Bestimmen sie die lineare Approximation von $f(x, y) = e^{x+y}(xy - 1)$ um $(0,0)$.

Wir beschließen diesen Abschnitt mit dem allgemeinen Fall im n-dimensionalen Raum, wenn auch nur um das Wort Hyperebene zu benutzen. Dem ist so weil eine Funktion $z = f(\mathbf{x})$ mit einer Hyperebene approximiert wird.

Definition 4.11. Bei der linearen Approximation einer Funktion $z = f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ um den Punkt $(\mathbf{x}^0) = (x_1^0, x_2^0, \dots, x_n^0)$ gilt für Werte von \mathbf{x} in der Nähe von \mathbf{x}^0

$$f(\mathbf{x}) \approx f(\mathbf{x}^0) + f_{x_1}(\mathbf{x}^0)(x_1 - x_1^0) + f_{x_2}(\mathbf{x}^0)(x_2 - x_2^0) + \dots + f_{x_n}(\mathbf{x}^0)(x_n - x_n^0)$$

Das Differential einer Funktion Während die lineare Approximation den Funktionswert um eine Stelle approximiert, approximiert das Differential die Veränderung des Funktionswerts. Es ist im univariaten Fall gegeben durch

$$dy = f'(x_0)dx$$

, wobei dx eine kleine Veränderung der Variable x ausgehend von x_0 darstellt. Der tatsächliche Zuwachs von y wenn sich x von x_0 um Δx verändert ist

$$\Delta y = f(x_0 + \Delta x) - f(x_0)$$

Die lineare Approximation and der Stelle x_0 liefert für den x-Wert $x_0 + \Delta x$

$$f(x_0 + \Delta x) \approx f(x_0) + f'(x_0)\Delta x$$

Wenn wir nun dx als sehr kleine Änderung von x betrachten und auf beiden Seiten $f(x_0)$ abziehen, sehen wir

$$\Delta y \approx dy = f'(x_0)dx$$

Wir sehen, dass das Differential die Veränderung des Funktionswertes approximiert.

Die folgenden Regeln für Differentiale sind denen der Ableitungen sehr ähnlich

Definition 4.12. Regeln für Differentiale: Seien f und g differenzierbare Funktionen von x (oder mehreren Variablen) und a und b seien Konstanten. Dann gelten folgende Regeln.

1. $d(af + bg) = a df + b dg$

2. $d(fg) = g df + f dg$

3. $d(f/g) = \frac{g df - f dg}{g^2}$ für $g \neq 0$

4. Kettenregel für Differentiale. Für $z = f[g(x)]$ ist das Differential gleich

$$dz = f'[g(x)]dg$$

Zum Abschluss weiten wir die Definition des Differentials auf Funktionen mehrerer Variablen aus.

Definition 4.13. *Differentiale: Gegeben sei die Funktion*

1. $z = f(x)$. Dann ist das Differential dz gleich

$$dz = f'(x)dx$$

2. $z = f(x, y)$. Dann ist das Differential dz gleich

$$dz = f_x(x, y)dx + f_y(x, y)dy$$

3. $z = f(\mathbf{x})$ mit $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Dann ist das Differential dz gleich

$$dz = f_{x_1}(\mathbf{x})dx_1 + f_{x_2}(\mathbf{x})dx_2 + \dots + f_{x_n}(\mathbf{x})dx_n$$

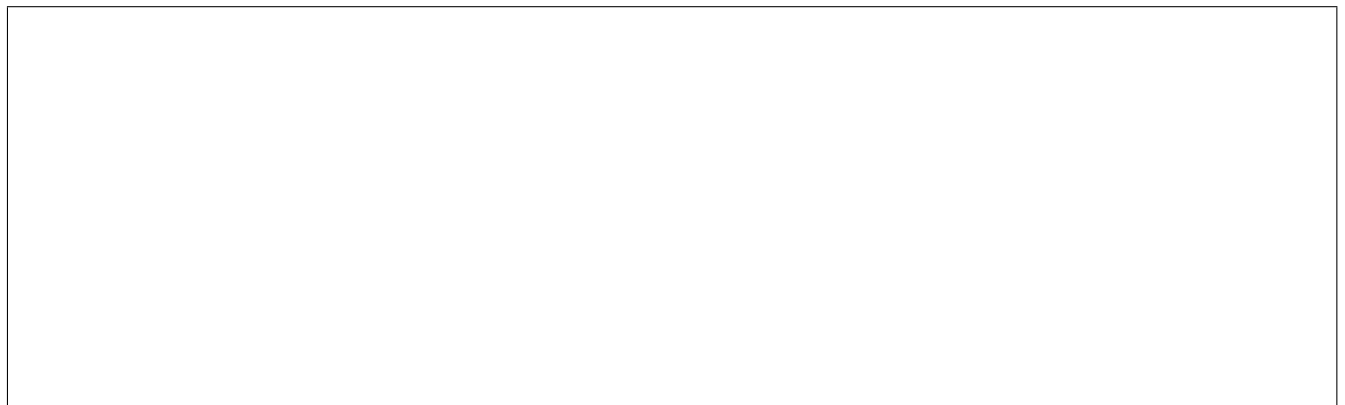
Beispiel 4.11. *Bestimmen sie die Differentiale der folgenden Funktionen:*

1. $z = f(K)$

3. $z = e^{xu}$ mit $u = f(x, y)$

2. $z = Ax^a + By^b$

4. $z = \ln(x^2 + y)$



4.5.2 Polynomiale Approximationen

Bisher haben wir eine nicht-lineare Funktion mit mit einer linearen Funktion approximiert. Natürlich hält uns nichts davon ab, die Approximation mit einer nicht-linearen Funktion durchzuführen. Dies führt zu komplexeren aber auch besseren Ergebnissen. Im Folgenden werden wir für die Approximation Polynome benutzen.

Definition 4.14. *Als Polynom n -ter Ordnung der Variablen x bezeichnen wir die folgende Funktion*

$$P(x) = \sum_{i=0}^n a_i x^i = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots + a_n x^n$$

Durch diese Definition wird klar, dass wir bisher die Approximation mit einem Polynom erster Ordnung durchgeführt haben, da dies die lineare Funktion $a_0 + a_1 x$ darstellt. Wir erweitern nun und führen die Approximation der Funktion $f(x)$ an der Stelle x_0 mit einer quadratischen Funktion (Polynom zweiten Grades) durch.

$$f(x) \approx a_0 + a_1(x - x_0) + a_2(x - x_0)^2$$

Wir müssen nun die Koeffizienten a_0, a_1 und a_2 bestimmen. Wir bewerkstelligen dies, indem wir verlangen, dass die Funktion und deren Approximation an der Stelle x_0 den selben

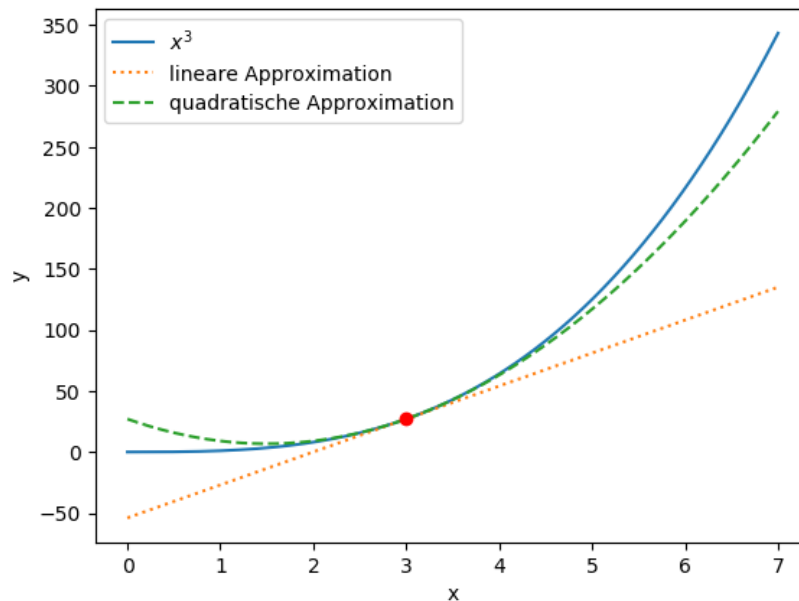
1. Funktionswert haben. $\implies a_0 = f(x_0)$
2. der ersten Ableitung haben. $\implies a_1 = f'(x_0)$
3. der zweiten Ableitung haben. $\implies a_2 = \frac{1}{2}f''(x_0)$

Definition 4.15. Quadratische Approximation von $f(x)$ um die Stelle x_0 . Für x in der Nähe von x_0 gilt

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

Die folgende Abbildung zeigt Approximationen der Funktion $y = x^3$ um die Stelle $x_0 = 3$. Wir sehen, dass die quadratische Approximation in der Nähe von $x_0 = 3$ besser ist.

Abbildung 22: Lineare und quadratische Approximation von $y = x^3$ an der Stelle $x = 3$



Beispiel 4.12. Bestimmen sie die quadratischen Approximationen:

1. $y = \sqrt[3]{x}$ um $x = 1$
2. $y = x^2$ um $x = 4$

Wir können die Approximation in zwei Richtungen erweitern

1. Approximation durch Polynome höheren Grades
2. Approximation von Funktionen mit mehreren Variablen

Definition 4.16. Taylor-Approximation n -ten Grades für ein Funktion $f(x)$ um x_0 . Für x nahe x_0 gilt

$$f(x) \approx f(x_0) + \frac{f^{(1)}(x_0)}{1!}(x - x_0) + \frac{f^{(2)}(x_0)}{2!}(x - x_0)^2 + \frac{f^{(3)}(x_0)}{3!}(x - x_0)^3 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

$f^{(n)}$ sind hier die Ableitungen von $f(x)$ n -ter Ordnung.

Notiz 4.2. Die vorher behandelten linearen und quadratischen Approximationen sind Spezialfälle der obigen Taylor-Approximation.

Beispiel 4.13. Bestimmen sie die Taylor-Approximation dritten Grades für $f(x) = \sqrt{1+x}$ um $x = 0$

Wir betrachten hier den Fall der Taylor-Approximation einer Funktion von mehreren Variablen nur für einen Spezialfall, der quadratischen Taylor-Approximation von $z = f(x, y)$.

Definition 4.17. Die Taylor-Approximation der Funktion $z = f(x, y)$ um den Punkt (x_0, y_0) ist gleich

$$f(x) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \frac{1}{2} [f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2]$$

oder in Matrixform mit $\mathbf{a}' = [x_0, y_0]$ und $\mathbf{x}' = [x, y]$

$$\approx f(\mathbf{a}) + (\mathbf{x} - \mathbf{a})\nabla f(\mathbf{a}) + (\mathbf{x} - \mathbf{a})H(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

5 Multivariate Optimierung

5.1 Generelles

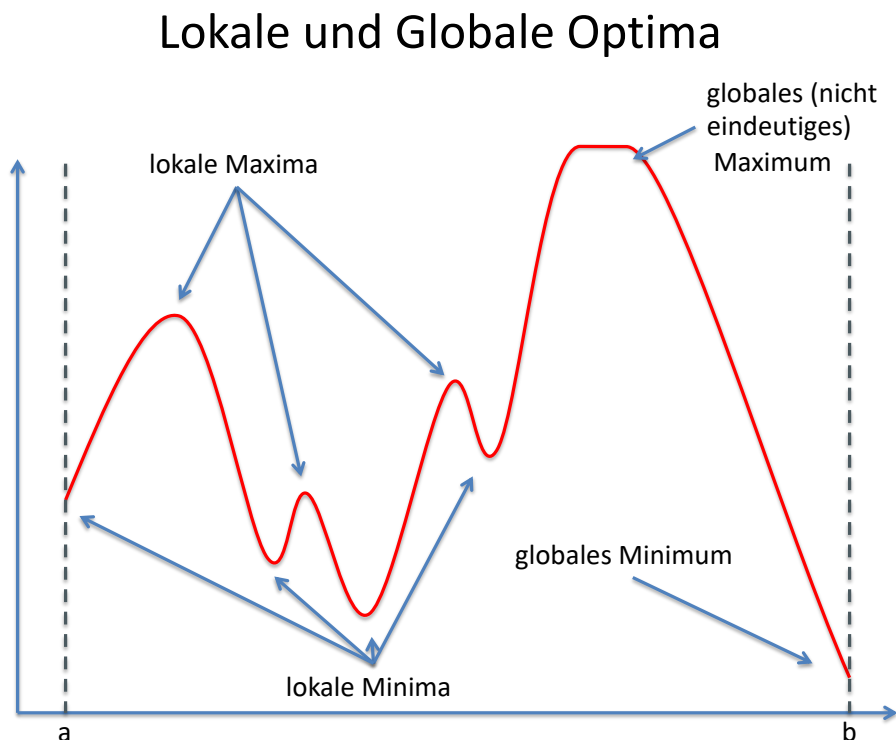
5.1.1 Globale und Lokale Extrema

Minima und Maxima sind Extrema einer Funktion. Diese können lokaler oder globaler Natur sein. Wir definieren diese Begriffe für Maxima und verlassen uns auf unsere Fähigkeiten dies auch für Minima zu tun.

Definition 5.1. *Globales vs. lokales Maximum einer Funktion $f(\mathbf{x})$ mit Definitionsbereich D . Ein Punkt \mathbf{x}^* ist ein*

1. *globales Maximum für $f(\mathbf{x})$ auf D , wenn gilt $f(\mathbf{x}^*) \geq f(\mathbf{x})$ für alle $\mathbf{x} \in D$.*
2. *striktes globales Maximum für $f(\mathbf{x})$ auf D , wenn gilt $f(\mathbf{x}^*) > f(\mathbf{x})$ für alle $\mathbf{x} \in D$.*
3. *lokales Maximum für $f(\mathbf{x})$ auf D , wenn gilt $f(\mathbf{x}^*) \geq f(\mathbf{x})$ für alle x , die nah an \mathbf{x}^* sind.*
4. *striktes lokales Maximum für $f(\mathbf{x})$ auf D , wenn gilt $f(\mathbf{x}^*) > f(\mathbf{x})$ für alle \mathbf{x} , die nah an \mathbf{x}^* sind.*

Die folgende Abbildung verdeutlicht diese Definitionen



Das Optimierung einer differenzierbaren Funktion von einer Variablen war in vielen Fällen (relativ) unkompliziert: Man bildet die erste Ableitung, setzt diese gleich Null und löst nach x auf. Wenn dann noch die zweite Ableitung positiv oder negativ ist, hat man ein (lokales) Maximum bzw. Minimum gefunden. Aber auch hier können Komplikationen auftreten. Zum Beispiel:

1. Das globale Minimum/Maximum kann am Rand des Definitionsbereichs auftreten, an dem die Funktion nicht differenzierbar ist.
Maximiere $y = x^2$ im Definitionsbereich des Intervalls $[2, 4]$
2. Das globale Minimum/Maximum kann nicht existieren.
Maximiere $y = x^2$ im Definitionsbereich des Intervalls $(2, 4)$
3. Das globale Minimum/Maximum ist an einer Stelle, an der die Funktion nicht differenzierbar oder nicht stetig ist
Minimiere $y = |x|$

Partielle Abhilfe schafft der Satz von Weierstrass, der eine hinreichende (nicht notwendige) Bedingung für die Existenz eines globalen Maximums und eines globalen Minimums formuliert. Bevor wir uns diesem Theorem zuwenden können, müssen wir aber noch einige Begriffe klären

5.1.2 Kompakte Mengen

Wir betrachten Mengen der reellen Zahlen, z.B ein Intervall $[a, b]$ in \mathbb{R} . Diese Mengen haben zwei verschiedene Arten von Punkten, *Randpunkte* und *innere Punkte*.

Definition 5.2. *Randpunkte und innere Punkte eines Intervalls $I \in \mathbb{R}$. Gegeben ist ein Punkt p und ein weiteres Intervall $A = (p + \delta, p - \delta)$ (mit $\delta > 0$, "kleine Umgebung" von p).*

1. Der Punkt p ist ein *Randpunkt* des Intervalls I , wenn jedes noch so kleine Intervall A ein Element von *innerhalb* und von *außerhalb* I enthält.

Notiz: Randpunkte müssen keine Elemente der betrachteten Menge sein

2. Der Punkt p ist ein *innerer Punkt*, wenn ein δ existiert, so dass A komplett in I liegt.

Notiz 5.1. Bei Mengen in \mathbb{R}^2 wird das Intervall A durch einen Kreis mit Radius δ ersetzt.

Beispiel 5.1. Finden sie die *Randpunkte* und *inneren Punkte* der folgenden Mengen

1. Das Intervall $(4, 5]$
2. Die Menge $B = (x, y \in \mathbb{R} : x^2 + y^2 \leq 1)$

Wir definieren nun offene und geschlossene Mengen.

Definition 5.3. *Offene und geschlossene Mengen: Eine Mengen B ist*

1. *offen*, wenn sie nur aus *inneren Punkten* besteht.

2. geschlossen, wenn sie alle ihre Randpunkte beinhaltet

Notiz 5.2. Mengen sind keine Türen, die entweder offen oder geschlossen sind. Sie können

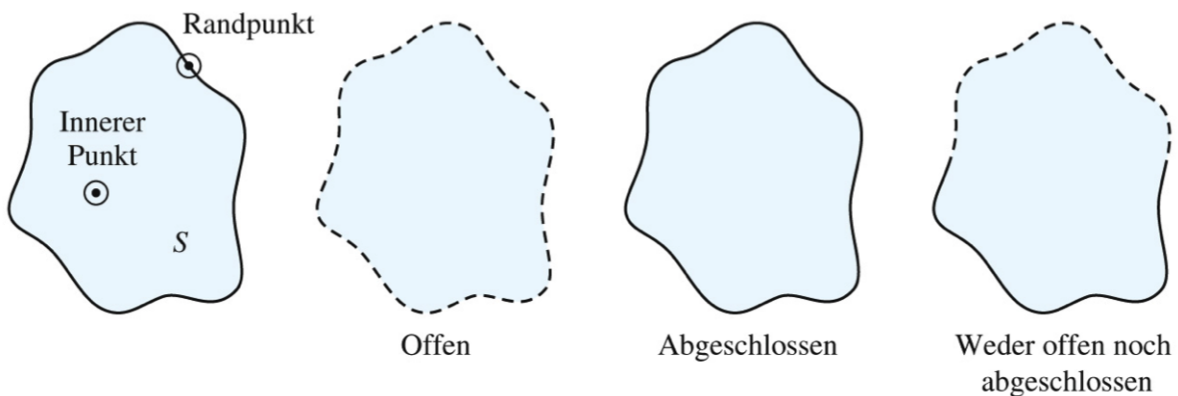
1. offen und geschlossen sein.
2. weder offen noch geschlossen sein.

Beispiel 5.2. Bestimmen sie, ob die folgenden Intervalle offen, geschlossen oder weder noch sind.

1. $(2,5)$
2. $[2,5]$
3. $(2,5]$



Die folgende Abbildung zeigt die Erweiterung der Begriffe auf Mengen in \mathbb{R}^2 .



Jetzt müssen wir den Unterschied zwischen beschränkten und unbeschränkten Mengen definieren. Intuitiv ist eine Menge beschränkt, wenn ihre Elemente nicht gegen unendlich oder minus unendlich streben.

Definition 5.4. Eine Menge ist beschränkt, wenn

1. es ein Intervall gibt, in der die Menge komplett enthalten ist (beschränkt in \mathbb{R})
2. es einen Kreis gibt, in der die Menge komplett enthalten ist (beschränkt in \mathbb{R}^2)

Zu guter Letzt können wir kompakte Mengen in \mathbb{R} oder \mathbb{R}^2 (oder \mathbb{R}^n) definieren.

Definition 5.5. Eine Menge $A \in \mathbb{R}^n$ ist kompakt, wenn sie abgeschlossen und beschränkt ist.

Beispiel 5.3. Bestimmen sie, ob die folgenden Mengen kompakt sind

1. $(-\infty,5)$
2. $[2,5]$
3. $A = \{x, y \in \mathbb{R} : x^2 + y^2 \leq 1\}$

5.1.3 Der Extremwertsatz von Weierstrass

Was nun folgt ist ein sehr wichtiger Satz, den es zu verinnerlichen gilt.

Theorem 5.1. *Extremwertsatz von Weierstrass. Wenn die Funktion $f(\mathbf{x})$ mit $\mathbf{x} = (x_1, x_2, \dots, x_n)$ innerhalb einer nicht-leeren, kompakten Menge K stetig ist, dann existiert sowohl eine Stelle \mathbf{x}_{\min} , an der die Funktion ein globales Minimum hat als auch eine Stelle \mathbf{x}_{\max} , an der die Funktion ein globales Maximum hat.*

$$f(\mathbf{x}_{\min}) \leq f(\mathbf{x}) \leq f(\mathbf{x}_{\max}) \quad \text{für alle } \mathbf{x} \in K$$

Notiz 5.3. :

1. Der Satz gilt natürlich auch für Funktionen von einer Variablen.
2. Der Satz gibt eine hinreichende Bedingung für die Existenz eines Extremums, die Bedingung ist **nicht** notwendig, d.h. es gibt Funktionen, deren Definitionsbereich nicht kompakt ist und/oder die nicht stetig sind und trotzdem globale Extrema haben.
3. Der Satz sagt uns nicht wie wir die Extrema finden.

5.2 Multivariate Optimierung Ohne Nebenbedingungen

Der Zusatz „ohne Nebenbedingungen“ bedeutet hier, dass innerhalb des Definitionsbereichs einer Funktion nach globalen Minima und globalen Maxima gesucht wird. Der Definitionsbereich (D) wird also nicht weiter eingeschränkt. Wir behandeln Problem wie

$$\max_{x \in D} f(x) \quad \text{oder} \quad \min_{x \in D} f(x)$$

Zum Beispiel wird die Funktion

$$y = \frac{1}{x}$$

über die Definitionsbereiche $D = \mathbb{R} \setminus \{0\}$ oder $D = \mathbb{R}^+$ oder $D = [a, b]$ optimiert.

Im Abschnitt 5.2 sehen wir, dass wir beim Versuch Extrema zu finden (analog zur Optimierung mit einer Variablen) nun die partiellen Ableitungen erster und zweiter Ordnung verwenden.

5.2.1 Bedingungen erster Ordnung

In diesem Abschnitt werden wir sehen, dass die partiellen Ableitungen erster Ordnung *notwendige* Bedingungen für Extrema liefern. Das Resultat ist aus der Analysis mit einer Variable bestens bekannt:

Theorem 5.2. Eine differenzierbare Funktion $z = f(x, y)$ kann nur dann ein Maximum oder Minimum an einem inneren Punkt (x_0, y_0) des Definitionsbereichs D annehmen, wenn dieser eine **stationäre Stelle** ist - d.h. wenn der Punkt (x_0, y_0) die zwei folgenden Gleichungen erfüllt.

$$f_x(x_0, y_0) = 0 \quad \text{und} \quad f_y(x_0, y_0) = 0$$

Diese Gleichungen werden Bedingungen erster Ordnung genannt (Englisch: First Order Conditions oder FOCs).

Die folgenden Abbildungen geben eine intuitive Erklärung für das obige Theorem. Abbildung 23 zeigt ein 'normales' Maximum am Punkt P. Es ist ersichtlich, dass am Punkt P beide partiellen Ableitungen gleich Null sein müssen. Abbildung 24 zeigt den Unterschied zwischen einem lokalen und einem globalen Maximum und zeigt auch einen Sattelpunkt, der weder Minimum noch Maximum ist. Abbildung 25 zeigt den Sattelpunkt im größeren Detail. Am Sattelpunkt erreicht die Funktion in x-Richtung ein Minimum und in y-Richtung ein Maximum.

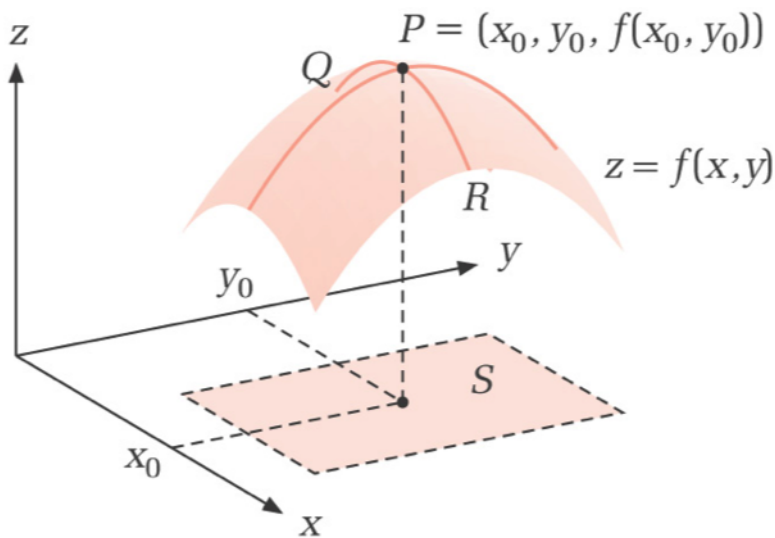


Abbildung 23: Globales Maximum an P

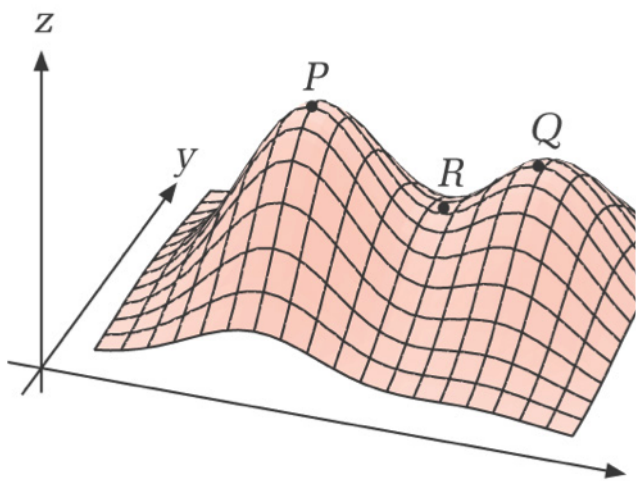


Abbildung 24: Globales Maximum P, lokales Maximum Q und Sattelpunkt R

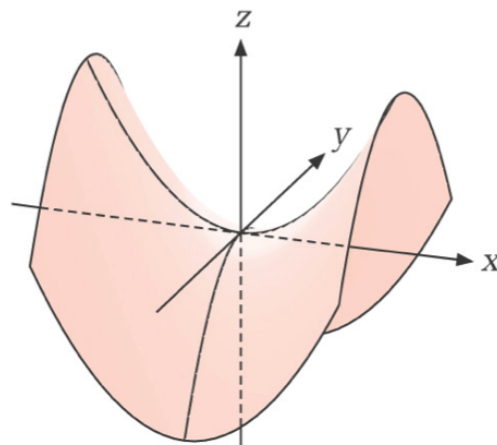
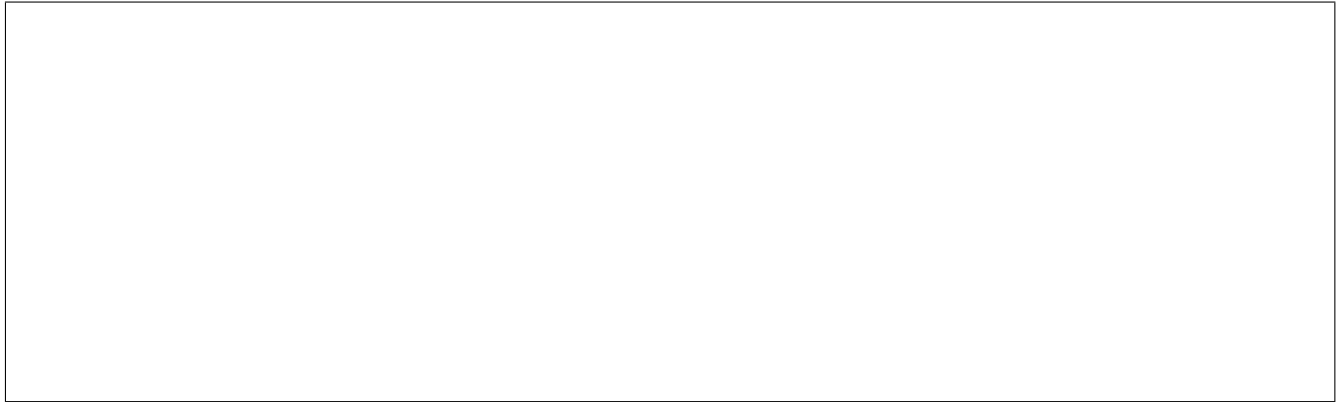


Abbildung 25: Sattelpunkt an $(0,0)$

Beispiel 5.4. Finden Sie die stationären Stellen der folgenden Funktionen

1. $f(x, y) = x^2 + y^2$
2. $f(x, y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$



Es ist möglich, dass eine Funktion an einer Stelle stetig, aber nicht differenzierbar ist. Auch an dieser Stelle kann eine Maximum oder Minimum vorliegen (siehe $f(x) = |x|$). Wir definieren deshalb *kritische Punkte* einer Funktion.

Definition 5.6 (Kritische Punkte). Gegeben sei eine stetige Funktion $f(\mathbf{x})$. Die kritischen Punkte dieser Funktion sind die stationären Punkte ($\nabla f(\mathbf{x}) = 0$) und die Punkte, an denen partielle Ableitungen nicht definiert sind. Als kritische Punkte kommen nur Punkte in Frage, für die $f(\mathbf{x})$ definiert ist.

Beispiel 5.5. Finden sie die kritischen Punkte der folgenden Funktionen

1. $f(x) = \sqrt[3]{x^2}(2x - 1)$
2. $f(x) = \frac{x^2+1}{x^2-x-6}$

Wir wollen jetzt auch noch die partiellen Ableitungen zweiter Ordnung der Funktion $y = f(x, y)$ betrachten und gehen im folgenden Abschnitt der Frage nach, wie wir die notwendigen Bedingungen in Theorem 5.2 bereichern können um Extrema zu identifizieren(ohne Gefahr zu laufen an einem Sattelpunkt) zu landen. Wir werden sehen, dass dies nur bedingt möglich ist.

Zum Abschluss geben wir noch die notwendigen Bedingungen erster Ordnung für den n-dimensionalen Fall.

Definition 5.7. Es sei \mathbf{x}^* ein lokales Maximum (oder Minimum) der differenzierbaren Funktion $f(\mathbf{x})$ mit $\mathbf{x} = (x_1, x_2, \dots, x_n)$ an einem innerer Punkt des Definitionsbereichs. Dann gilt

$$\nabla f(\mathbf{x}) = \begin{bmatrix} f_{x_1}(\mathbf{x}) \\ f_{x_2}(\mathbf{x}) \\ \vdots \\ f_{x_n}(\mathbf{x}) \end{bmatrix} = \mathbf{0}$$

5.2.2 Bedingungen zweiter Ordnung

Wir zeigen zunächst noch einmal den Fall einer reellen Funktion einer Variablen. Für reelle Funktionen einer Variablen gleicht die Taylor Approximation 2. Ordnung um einen stationären Punkt x_0

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

$$f(x) - f(x_0) \approx \frac{1}{2}f''(x_0)(x - x_0)^2$$

Die Umformung nutzt die Tatsache, dass $f(x_0) = 0$. Somit bestimmt die zweite Ableitung das Verhalten der Funktion um einen stationären Punkt

- Wenn $f''(x_0) > 0$ liegt ein striktes lokales Minimum vor.
- Wenn $f''(x_0) < 0$ liegt ein striktes lokales Maximum vor.
- In den anderen Fällen kann keine Aussage getroffen werden

Gegeben sei nun (x_0, y_0) ein stationärer Punkt der Funktion $f(x, y)$. Intuitiv ist (hoffentlich) klar, dass auch diese Funktion in einem kleinen Bereich um den stationären Punkt ansteigen muss, wenn (x_0, y_0) ein striktes lokales Minimum sein soll. Für ein (nicht-striktes) lokales Minimum darf sie in diesem Bereich zumindest nicht fallen. Eine Taylor-Approximierung zweiter Ordnung der Funktion $f(x, y)$ um (x_0, y_0) liefert uns die approximativen Funktionswerte um den stationären Punkt. Laut Definition 4.17 ist diese gegeben durch

$$f(x) \approx f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

$$+ \frac{1}{2} [f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2]$$

$$f(x) - f(x_0, y_0) \approx \frac{1}{2} [f_{xx}(x_0, y_0)(x - x_0)^2$$

$$+ 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2] \quad (3)$$

Die Umformung nutzt die Tatsache, dass die partiellen Ableitungen erster Ordnung am stationären Punkt gleich Null sind. Man sieht nun, dass die Differenz zwischen dem Funktionswert an der Stelle (x_0, y_0) und in der Umgebung dieser Stelle von einem Term bestimmt wird, dessen Wert von ALLEN partiellen Ableitungen zweiter Ordnung abhängt. Dieser Term ist eine sogenannte *Quadratische Form*

Definition 5.8 (Quadratische Form). *Es sei A eine $n \times n$ Matrix und \mathbf{v} ein $n \times 1$ Vektor mit den Elementen (v_1, v_2, \dots, v_n) Dann ist die Quadratische Form gleich*

$$z = \mathbf{v}' A \mathbf{v}$$

Notiz 5.4. *Wir beachten, dass es sich bei der Quadratischen Form (z) um einen Skalar (= eine reelle Zahl handelt).*

Zum Beispiel, für den Fall $n = 2$

$$z = [v_1 \quad v_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = a_{11}v_1^2 + v_1v_2(a_{12} + a_{21}) + a_{22}v_2^2$$

Wir sehen, dass der Term in der Taylor-Approximation zweiter Ordnung (Gleichung 3), der darüber entscheiden kann, ob es sich um ein lokales Minimum, lokales Maximum oder einen Sattelpunkt handelt, eine

quadratische Form der Hesse-Matrix darstellt.

$$A = \begin{bmatrix} (x - x_0) & (y - y_0) \end{bmatrix} \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix} \begin{bmatrix} (x - x_0) \\ (y - y_0) \end{bmatrix}$$

Die Aussagekraft der partiellen Ableitungen zweiter Ordnung wird also durch das Vorzeichen der Quadratischen Form um den stationären Punkt bestimmt. Um diese zu definieren führen wir den Begriff der Definitheit einer Quadratischen Form ein.

Definition 5.9. *Definitheit einer Quadratischen Form.* Es sei A eine $(n \times n)$ Matrix und \mathbf{v} ein n -dimensionaler Vektor. Eine Quadratische Form

$$z = \mathbf{v}' A \mathbf{v}$$

heißt

1. positiv definit, wenn $z = \mathbf{v}' A \mathbf{v} > 0$ für alle $\mathbf{v} \neq \mathbf{0}$.
2. negativ definit, wenn $z = \mathbf{v}' A \mathbf{v} < 0$ für alle $\mathbf{v} \neq \mathbf{0}$.
3. positiv semidefinit, wenn $z = \mathbf{v}' A \mathbf{v} \geq 0$ für alle \mathbf{v} .
4. negativ semidefinit, wenn $z = \mathbf{v}' A \mathbf{v} \leq 0$ für alle \mathbf{v} .
5. indefinit wenn keiner der obigen Fälle zutrifft.

Notiz 5.5. Da die Quadratische Form komplett von der Matrix A bestimmt wird, werden wir ab nun nur noch die Matrix A als positiv (semi)definit, negativ (semi)definit oder indefinit bezeichnen.

5.2.3 Exkurs: Determinante und Definitheit einer Matrix

Gegeben sei eine quadratische $(n \times n)$ Matrix A . Die Determinante einer Matrix ist eine aussagekräftige Zahl, die aus den Elementen der Matrix berechnet werden kann. Wir werden hier nur 2×2 und 3×3 Matrizen behandeln. Wir motivieren die Berechnung der Determinante mit dem Gleichungssystem

$$\begin{aligned} a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2 \end{aligned}$$

in Matrixschreibweise können wir schreiben

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Die Lösungen (z.B. durch auflösen und einsetzen) des Systems sind

$$x = \frac{b_1 a_{22} - b_2 a_{12}}{(a_{11} a_{22} - a_{12} a_{21})} \quad y = \frac{b_2 a_{11} - b_1 a_{21}}{(a_{11} a_{22} - a_{12} a_{21})}$$

Man sieht, dass eine eindeutige Lösung des Gleichungssystems voraussetzt, dass

$$|A| = a_{11} a_{22} - a_{12} a_{21} \neq 0$$

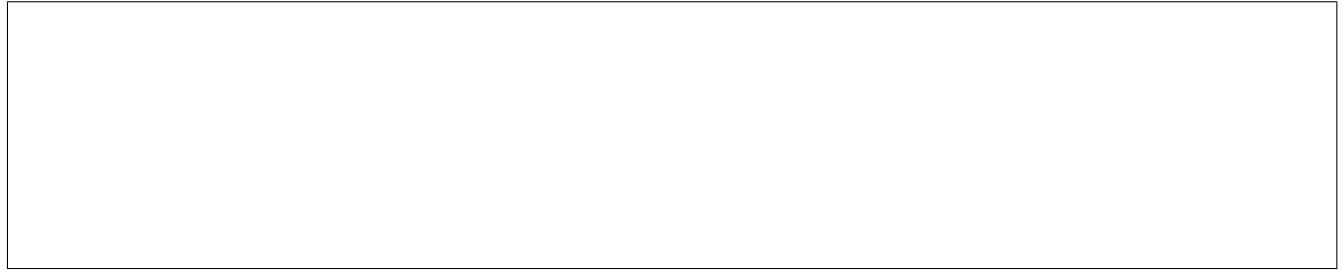
$|A|$ wird die Determinante von A genannt.

Beispiel 5.6. Finden sie die Determinanten der folgenden Matrizen

1. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}$



Wir fahren fort mit der 3×3 Matrix

$$B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Die Determinante dieser Matrix kann wie folgt berechnet werden

$$\begin{aligned} |B| &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \end{aligned}$$

Wir sehen, dass sich die Determinante der 3×3 Matrix B aus drei Produkten besteht, die jeweils die Determinante einer 2×2 Untermatrix enthalten

Definition 5.10 (Kofaktor). *Der Kofaktor des (i, j) ten Elements einer Matrix A ist die Determinante der Untermatrix, die sich ergibt wenn man die i te Zeile und die j te Spalte der Matrix A löscht multipliziert mit $(-1)^{(i+j)}$*

Beispiel 5.7. *Gegeben sei die Matrix*

$$B = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 3 & 5 \\ 4 & 5 & 2 \end{bmatrix}$$

Berechnen sie den Kofaktor des $(2, 2)$ ten Elements der Matrix. Wenn man die 2te Zeile und Spalte löscht ergibt sich die Untermatrix

$$\begin{bmatrix} 1 & 4 \\ 4 & 2 \end{bmatrix}$$

und der Kofaktor ist

$$K_{2,2} = (-1)^4 \begin{vmatrix} 1 & 4 \\ 4 & 2 \end{vmatrix} = -14$$

Wir sehen, dass die Determinante einer 3×3 Matrix gleich der Summe der mit ihren bestimmten Elementen multiplizierten Kofaktoren ist. Dabei bestimmen die Elemente der ersten Zeile die Kofaktoren. Allgemein gilt

Definition 5.11 (Berechnung der Determinante mit Kofaktoren). *Wir können die Determinante einer quadratischen Matrix B mittels der Elemente einer Spalte (Zeile) und deren Kofaktoren wie folgt berechnen:*

Gegeben sei eine $n \times n$ Matrix B mit Elementen b_{ij}

- Mit den Elementen der i -ten Zeile

$$|B| = b_{i1}K_{i1} + b_{i2}K_{i2} + b_{i3}K_{i3} + \dots + b_{in}K_{in}$$

- Mit den Elementen der j -ten Spalte

$$|B| = b_{1j}K_{1j} + b_{2j}K_{2j} + b_{3j}K_{3j} + \dots + b_{nj}K_{nj}$$

wobei K_{ij} der Kofaktor des (i, j) -ten Elements ist.

Beispiel 5.8. Gegeben sei die Matrix

$$B = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 3 & 5 \\ 4 & 5 & 2 \end{bmatrix}$$

1. Berechnen sie die Determinante der Matrix B mittels der Elemente und Kofaktoren der ersten Zeile von B
2. Berechnen sie die Determinante der Matrix B mittels der Elemente und Kofaktoren der zweiten Zeile von B

Und hier auch noch mal die Berechnung mittels der Elemente und Kofaktoren der ersten Spalte

$$\begin{aligned} |B| &= (-1)^2 \cdot 1 \cdot \begin{vmatrix} 3 & 5 \\ 5 & 2 \end{vmatrix} + (-1)^3 \cdot 4 \cdot \begin{vmatrix} 2 & 4 \\ 5 & 2 \end{vmatrix} + (-1)^4 \cdot 4 \cdot \begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} \\ &= -19 + 64 - 8 \\ &= 37 \quad \text{WOW!} \end{aligned}$$

Notiz 5.6. Die oben gelernte Art Determinanten zu berechnen ist unglaublich ineffizient. Wir zeigen dies anhand einer 25×25 Matrix. Die oben gelernte Berechnung mittels Kofaktoren benötigt ca. 25! Berechnungen. Wenn ein Computer eine Milliarde Berechnungen pro Sekunde durchführen kann, dauert die Bestimmung der Determinante ca. 500 Millionen Jahre! Glücklicherweise gibt es effizientere Verfahren.

Definition 5.12. *Führende Hauptminoren einer $n \times n$ Matrix A .*

Führende Hauptminoren k -ter Ordnung entstehen wenn sukzessive die letzten $(n-k)$ Spalten und $(n-k)$ Zeilen gelöscht werden.

Beispiel 5.9. *Gegeben sei die Matrix*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

3 führende Hauptminoren

$$A_1 = |a_{11}| \quad A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Beispiel 5.10. *Gegeben sei die Matrix*

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 3 & 4 \\ 3 & 4 & 3 \end{bmatrix}$$

Berechnen sie alle führenden Hauptminoren.

Hauptminoren und führende Hauptminoren werden verwendet um die Definitheit einer Matrix zu überprüfen

Definition 5.13 (Definitheit einer symmetrischen $(n \times n)$ Matrix). *Eine symmetrische $n \times n$ -Matrix A ist*

- 1. positiv definit, wenn und nur wenn alle führenden Hauptminoren positiv sind, $A_i > 0$ für $i = 1 \dots n$*
- 2. negativ definit, wenn und nur wenn alle führenden Hauptminoren ungerader Ordnung negativ und alle führenden Hauptminoren gerader Ordnung positiv sind, $(-1)^i A_i > 0$ für $i = 1 \dots n$*
- 3. In allen anderen Fällen ist die Matrix positiv semidefinit, negativ semidefinit oder indefinit*

Notiz 5.7. *Es muss hier noch einmal betont werden, dass diese Resultate nur für symmetrische Matrizen Gültigkeit besitzen.*

Beispiel 5.11. *Gegeben sei die Matrizen*

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad B = \begin{bmatrix} -3 & 2 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 3 \end{bmatrix}$$

Bestimmen sie, ob die die Matrizen positiv definit, negativ definit oder weder noch sind.

Die folgenden Theoreme liefern *notwendige* und *hinreichende* Bedingungen für lokale Extrema.

Theorem 5.3 (Notwendige Bedingungen zweiter Ordnung). Gegeben sei eine 2-fach differenzierbare Funktion $f(x, y)$ und ein innerer Punkt (x_0, y_0)

1. Wenn die Funktion an der Stelle (x_0, y_0) ein Maximum hat, dann ist $H(x_0, y_0)$ negative semidefinit
2. Wenn die Funktion an der Stelle (x_0, y_0) ein Minimum hat, dann ist $H(x_0, y_0)$ positiv semidefinit

Notiz 5.8. Ähnlich wie bei den Bedingungen erster Ordnung muss an dieser Stelle darauf hingewiesen werden, dass auch Theorem 5.3 nur notwendige Bedingungen liefert. Es besagt zum Beispiel nicht, dass ein Minimum vorliegt, wenn die Hesse-Matrix positiv semidefinit ist.

Theorem 5.4 (Hinreichende Bedingungen zweiter Ordnung). Gegeben sei eine 2-fach differenzierbare Funktion $f(x, y)$ und ein innerer Punkt (x_0, y_0) . Wenn (x_0, y_0) ein stationärer Punkt ist und die

1. Hesse-Matrix $H(x_0, y_0)$ positiv definit ist, dann ist (x_0, y_0) ein striktes lokales Minimum.
2. Hesse-Matrix $H(x_0, y_0)$ negativ definit ist, dann ist (x_0, y_0) ein striktes lokales Maximum.

Notiz 5.9. Dieses Theorem liefert eine hinreichende Bedingung. Wenn wir einen stationären Punkt haben, an dem die Hesse-Matrix positiv oder negativ definit ist, liegt ein striktes Extremum vor. Der umgekehrte Fall muss nicht immer gelten, d.h. es kann ein Extremum vorliegen obwohl potenziell beide Bedingungen (stationärer Punkt/Definitheit) nicht zutreffen.

Beispiel 5.12. In Beispiel 5.4 hatten wir die stationären Stellen der folgenden Funktionen berechnet

1. $f(x, y) = x^2 + y^2$ mit stationärem Punkt $(0, 0)$
2. $f(x, y) = -2x^2 - 2xy - 2y^2 + 36x + 42y - 158$ mit den stationärem Punkte $(5, 8)$

Bestimmen sie nun mittels der Hesse-Matrizen, ob sie mit Sicherheit sagen können, dass die stationären Stellen lokale Maxima oder Minima sind.

Wir sehen, dass die obigen Bedingungen zweiter Ordnung uns ein wenig weiterhelfen lokale Extrema zu identifizieren, jedoch keine Hilfe sind Extrema zu identifizieren, an denen die Hesse-Matrix semidefinit ist. Da zusätzlich auch die hinreichenden Bedingungen nicht in der Lage sind *globale* Extrema zu identifizieren, sind die Bedingungen zweiter Ordnung nur von limitiertem Nutzen.

Wir geben nun eine Abfolge von Schritten, mit denen man die globalen Extrema einer stetigen Funktion einer Variablen findet, die einen abgeschlossenem, beschränktem Definitionsbereich besitzt.

Definition 5.14. Gegeben sei eine Funktion $f(x)$, die auf dem Intervall $[a, b]$ stetig ist. Zum Auffinden der globalen Extrema werden folgende Schritte benutzt:

1. Finden sie alle kritischen Punkte der Funktion $f(x)$ im Intervall $[a, b]$
2. Evaluieren sie die Funktion an allen kritischen Punkten und an den Randpunkten
3. Identifizieren sie das globale Maximum und das globale Minimum

Beispiel 5.13. Finden sie die Extrema der folgenden Funktionen

1. $f(x) = 2x^3 + 3x^2 - 12x + 4$ auf $[-4, 2]$
2. $x^{\frac{2}{3}} - 2x$ auf $[-1, 1]$

Wir wenden uns nun Spezialfällen zu, in denen es möglich ist globale Extrema einfacher zu identifizieren. Dazu müssen wir der Funktion $f(x, y)$ aber noch mehr Struktur geben.

5.2.4 Konkavität/Konvexität und Optimierung

Wir beginnen diesen Abschnitt mit der Definition einer konvexen Menge.

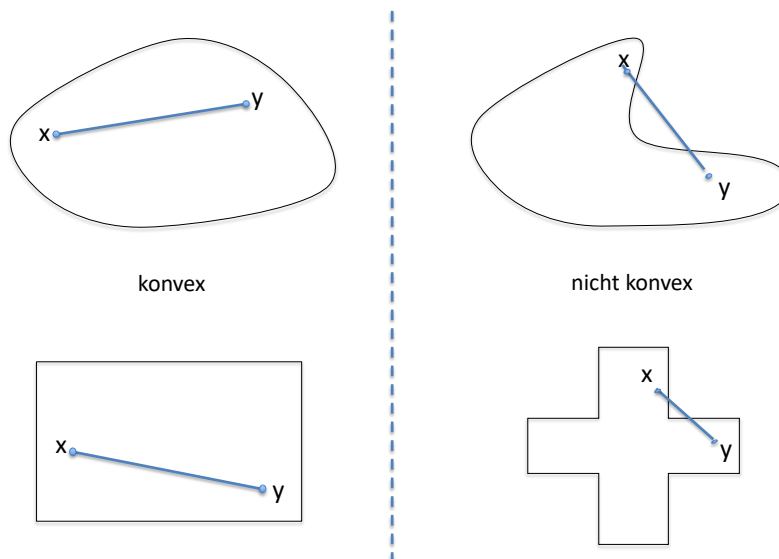
Konvexe Mengen

Definition 5.15. Eine Menge M heißt konvex, wenn für alle $\mathbf{x}, \mathbf{y} \in M$ und für alle $\lambda \in \mathbb{R}$ mit $0 \leq \lambda \leq 1$ stets gilt

$$\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in M$$

In Worten ausgedrückt ist eine Menge konvex, wenn eine Linie zwischen zwei Punkten dieser Menge komplett in der Menge liegt. Die Abbildungen verdeutlichen dies.

Abbildung 26: Konvexe und nicht-konvexe Mengen



Konkave und Konvexe Funktionen

Definition 5.16 (Konkave und Konvexe Funktionen). Eine Funktion $f(\mathbf{x})$, die auf einer konvexen Menge M definiert ist, heißt

1. konvex, wenn für alle $\mathbf{x}, \mathbf{y} \in M$ und $t \in \mathbb{R}$ mit $0 \leq t \leq 1$ gilt

$$f[t\mathbf{x} + (1 - t)\mathbf{y}] \leq tf(\mathbf{x}) + (1 - t)f(\mathbf{y})$$

2. konkav, wenn für alle $\mathbf{x}, \mathbf{y} \in M$ und $t \in \mathbb{R}$ mit $0 \leq t \leq 1$ gilt

$$f[t\mathbf{x} + (1 - t)\mathbf{y}] \geq tf(\mathbf{x}) + (1 - t)f(\mathbf{y})$$

3. *streng konvex*, wenn für alle $\mathbf{x}, \mathbf{y} \in M$ ($\mathbf{x} \neq \mathbf{y}$) und $t \in \mathbb{R}$ mit $0 < t < 1$ gilt

$$f[t\mathbf{x} + (1-t)\mathbf{y}] < tf(\mathbf{x}) + (1-t)f(\mathbf{y})$$

4. *streng konkav*, wenn für alle $\mathbf{x}, \mathbf{y} \in M$ ($\mathbf{x} \neq \mathbf{y}$) und $t \in \mathbb{R}$ mit $0 < t < 1$ gilt

$$f[t\mathbf{x} + (1-t)\mathbf{y}] > tf(\mathbf{x}) + (1-t)f(\mathbf{y})$$

Wir illustrieren diese Definition mit einem Beispiel einer reellen Funktion einer Variablen. Hier besagt die Definition, dass eine Funktion

1. konvex ist, wenn die Verbindungslinie zwischen zwei Funktionswerten komplett auf oder oberhalb des Graphen der Funktion liegt
2. konkav ist, wenn die Verbindungslinie zwischen zwei Funktionswerten komplett auf oder unterhalb des Graphen der Funktion liegt

Die Konkavität oder Konvexität einer Funktion kann somit durch deren Epi- und Hypograph definiert werden.

Definition 5.17. *Es sei $f(\mathbf{x})$ eine reelle Funktion, die auf einer konvexen Menge M definiert ist.*

1. *Der Epigraph der Funktion ist die Menge aller Punkte, die auf oder über dem Graph liegen. Die Funktion ist konvex, wenn der Epigraph eine konvexe Menge ist.*
2. *Der Hypograph der Funktion ist die Menge aller Punkte, die auf oder unter dem Graph liegen. Die Funktion ist konkav, wenn der Hypograph eine konvexe Menge ist.*

Die folgenden Abbildungen zeigen Beispiele im 2- und 3-dimensionalen Raum.

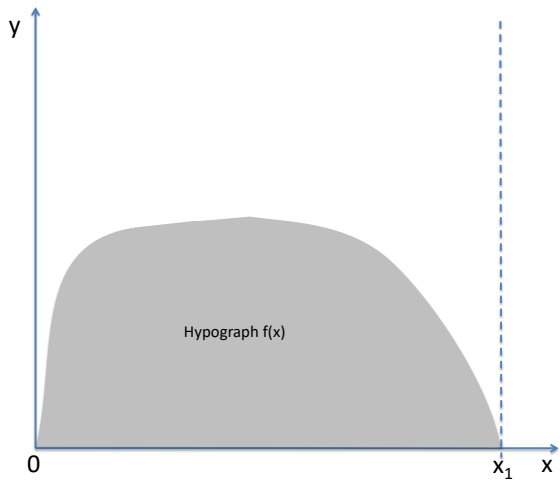


Abbildung 27: Konkave Funktion

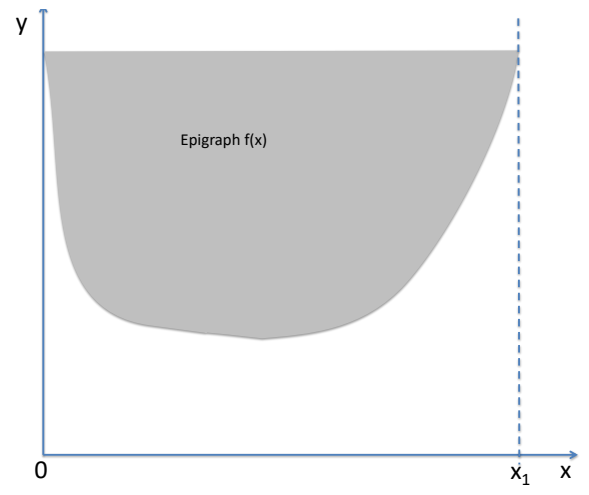


Abbildung 28: Konvexe Funktion

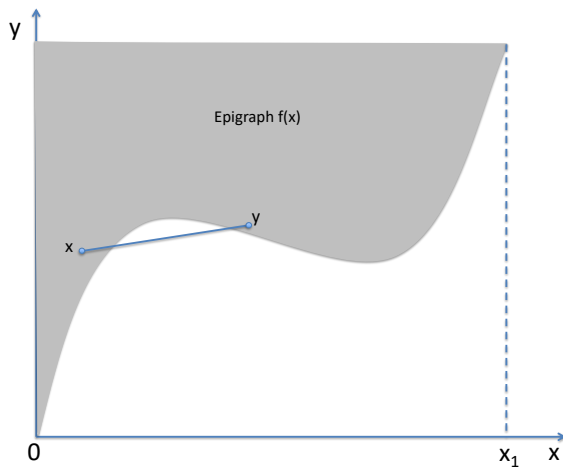


Abbildung 29: Nicht konkav/konvex

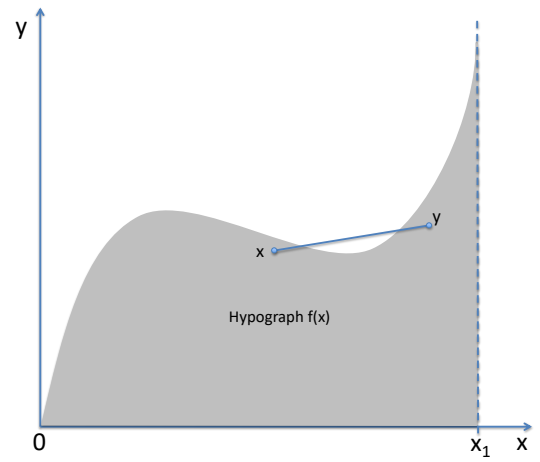
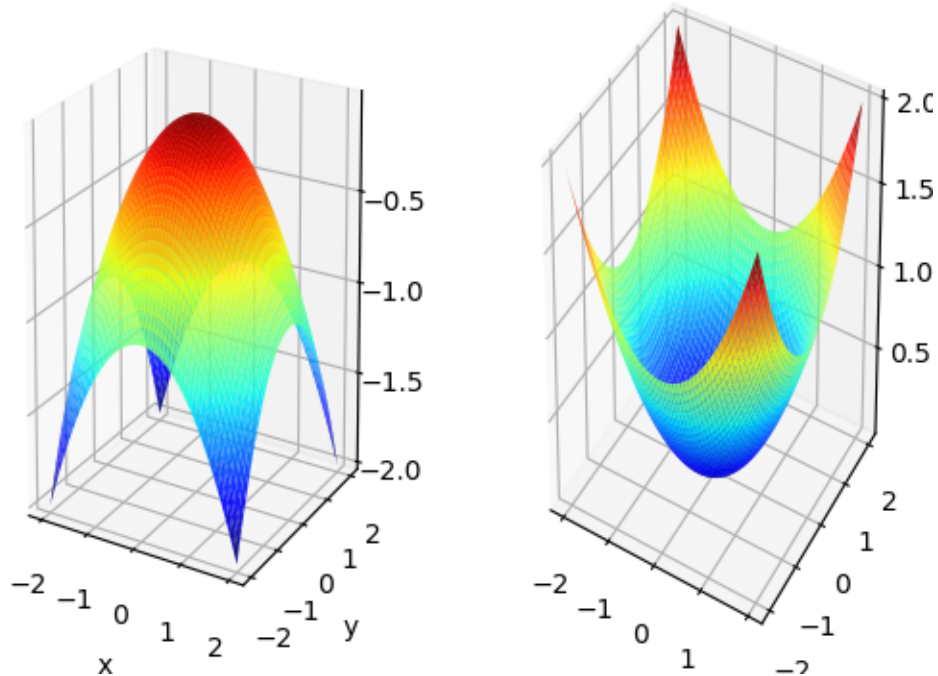


Abbildung 30: Nicht konkav/konvex

Abbildung 31: Links: Konkave Funktion in 3D Rechts: Konvexe Funktion in 3D



Wenn eine reelle Funktion einer Variablen differenzierbar ist, können wir Konkavität/Konvexität wie folgt bestimmen.

Definition 5.18. Gegeben sei $f(x)$, eine reelle, 2-fach differenzierbare Funktion einer Variablen definiert auf einer konvexen Menge M .

1. $f(x)$ ist konkav auf M , wenn und nur wenn $f''(x) \leq 0$ für alle inneren Punkte von M .
2. Wenn $f''(x) < 0$ für alle inneren Punkte von M , dann ist $f(x)$ streng konkav auf M .
3. $f(x)$ ist konvex auf M , wenn und nur wenn $f''(x) \geq 0$ für alle inneren Punkte von M .
4. Wenn $f''(x) > 0$ für alle inneren Punkte von M , dann ist $f(x)$ streng konvex auf M .

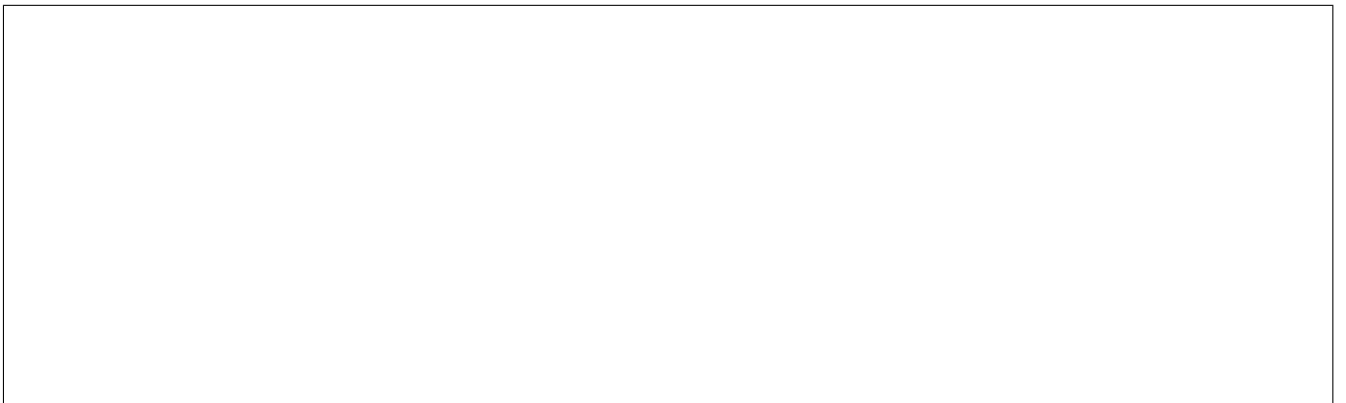
Beispiel 5.14. Bestimmen sie, ob die folgenden Funktionen konvex, konkav, beides (oder weder noch) sind.

1. $f(x) = a + bx$

3. $f(x) = x^3$

2. $f(x) = x^2$

4. $f(x) = \sqrt{x}$



Wenn eine reelle Funktion mehrerer Variablen zweifach differenzierbar ist, können wir Konkavität/Konvexität wie folgt bestimmen.

Definition 5.19 (Konkave/Konvexe 2-fach differenzierbare Funktionen). Gegeben sei eine 2-fach differenzierbare Funktion $f(\mathbf{x})$ mit stetigen partiellen Ableitungen zweiter Ordnung definiert auf einer konvexen Menge M . Die Hesse-Matrix der Funktion am Punkt x wird geschrieben als $H(x)$. Dann gilt:

1. $f(\mathbf{x})$ ist konkav, wenn und nur wenn $H(x)$ negativ semidefinit ist für alle $\mathbf{x} \in M$.
2. Wenn $H(x)$ für alle $x \in M$ negativ definit ist, ist $f(\mathbf{x})$ streng konkav.
3. $f(\mathbf{x})$ ist konvex, wenn und nur wenn $H(x)$ positiv semidefinit ist für alle $\mathbf{x} \in M$.
4. Wenn $H(x)$ für alle $x \in M$ positiv definit ist, ist $f(\mathbf{x})$ streng konvex.

Notiz 5.10. Es ist nicht zwingend wahr das $H(x)$ negativ definit ist, wenn $f(\mathbf{x})$ streng konkav ist.

Zum Abschluss unserer Diskussion über Konkavität und Konvexität präsentieren wir noch zwei Resultate:

Theorem 5.5. :

1. Wenn f konkav (konvex) ist, ist $-f$ konvex (konkav)
2. Wenn f und g konkav (konvex) sind, ist auch die Summe $af + bg$ mit $a, b > 0$ konkav (konvex)

Beispiel 5.15. Bestimmen sie, ob die folgenden Funktionen konvex, konkav (oder weder noch) sind.

1. $f(x) = -x^2$

2. $f(x) = 10x^2 - 5\sqrt{x}$



Wir kommen nun zum Hauptteil dieses Abschnitts. Hier sehen wir, dass die Identifikation von globalen Extrema sich sehr einfach gestaltet, wenn wir annehmen, dass die zu optimierende Funktion konkav (oder konvex) ist.

Theorem 5.6. Es sei $f(\mathbf{x})$ eine reelle Funktion, die auf einer konvexen Menge M definiert ist.

1. Wenn $f(\mathbf{x})$ konkav ist und mindestens ein Maximum besitzt, ist jedes lokale Maximum ein globales Maximum.
2. Wenn $f(\mathbf{x})$ streng konkav ist existiert entweder ein oder kein Maximum.

Natürlich gelten die obigen Theoreme auch wenn die Funktion konvex ist. Dann geht es um Minima.

Beispiel 5.16. Bestimmen sie die Extrema (wenn existent) der folgenden Funktionen.

1. $f(x) = \sqrt{x}$

2. $f(x) = \sqrt{x} - 2x^2$



Theorem 5.7. *Es sei $f(\mathbf{x})$ eine reelle Funktion, die auf einer konvexen Menge M definiert ist und \mathbf{x}^* ein innerer Punkt der Menge M . Wenn $f(\mathbf{x})$ konkav (konvex) und auf M differenzierbar ist, dann und nur dann ist \mathbf{x}^* ein globales Maximum (Minimum) wenn*

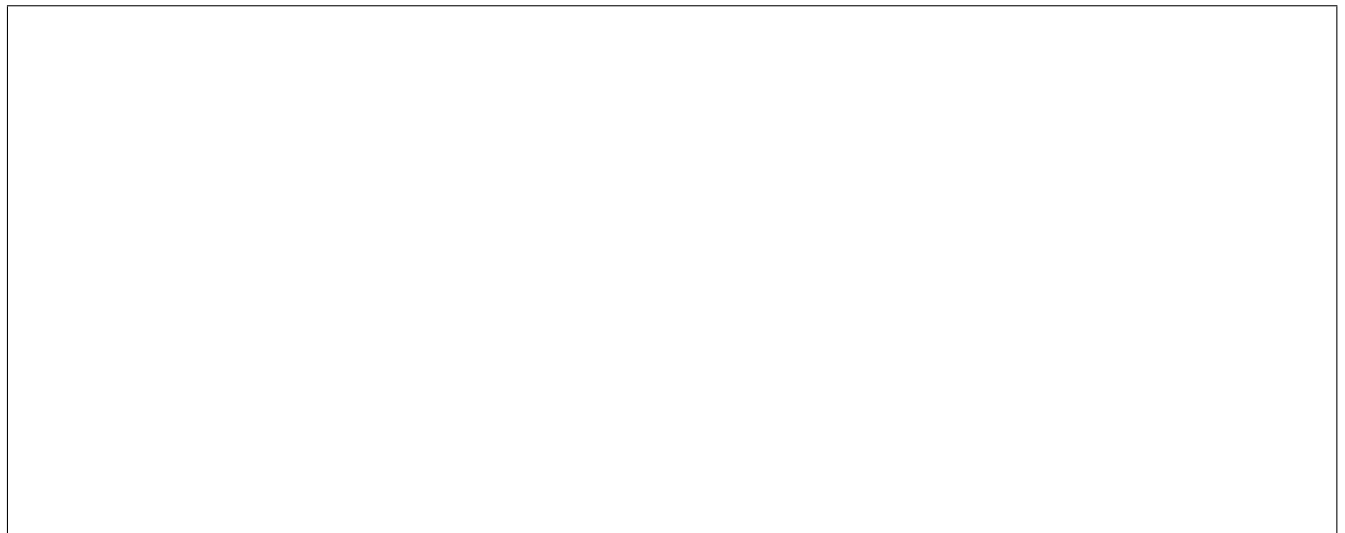
$$\nabla f(\mathbf{x}^*) = 0$$

‘dann und nur dann’ bedeutet hier, dass Theorem 5.7 eine notwendige UND hinreichende Bedingung darstellt, die uns das Auffinden von Extrema extrem erleichtert. Wir brauchen uns in diesem Fall keine Gedanken um Wende- oder Sattelpunkte zu machen. Wir müssen gegebenenfalls natürlich auch noch Randpunkte bewerten (dann gibt es keine stationären Stellen)

Beispiel 5.17. *Finden sie die Extrema der folgenden Funktionen:*

1. $f(x) = 6x - 3x^2$ auf $[-1, 2]$

2. $f(x) = 16\sqrt{x} - \frac{1}{2}x^2$ auf $[0, 6]$



5.3 Parameter, Wertfunktionen und das Envelope Theorem für Optimierungsprobleme ohne Nebenbedingungen

In den Wirtschaftswissenschaften kommt es oft vor, dass Optimierungsprobleme nicht nur von Variablen abhängen, die frei gewählt werden können (z.B. x und y), sondern auch noch von solchen, die vorgegeben sind (je nach Problem Preise, Steuersätze, Einkommen, Löhne). Die Letzteren werden Parameter des Problems genannt.

Beispiele:

- Ein Konsument maximiert seinen Nutzen aus dem Konsum von zwei Gütern. Hier sind die Preise der Güter und das Einkommen des Konsumenten die Parameter des Problems
- Eine Firma minimiert die Kosten einer bestimmten Produktionsmenge. Hier sind Löhne und der Preis des Kapitals die Parameter des Problems

Wir sind nun oft daran interessiert wie sich Veränderungen dieser Parameter auf den Optimalwert der Zielfunktion auswirken.

Gegeben sei ein Maximierungsproblem ohne Nebenbedingungen mit einem Parameter α

$$\max f(x, y; \alpha)$$

Um die Lösung zu finden werden die ersten partiellen Ableitungen nach x und y gebildet und gleich Null gesetzt

$$f_x(x, y, \alpha) = 0 \quad f_y(x, y, \alpha) = 0$$

Diese Gleichungen liefern die Werte für x und y , die die Zielfunktion maximieren (Wir nehmen an, dass die Bedingungen zweiter Ordnung erfüllt sind)

$$x^*(\alpha) \quad y^*(\alpha)$$

Wenn man nun diese Werte in die Zielfunktion einsetzt erhält man die **Wertfunktion**, auch Optimalwertfunktion genannt.

$$V(\alpha) = f[x^*(\alpha), y^*(\alpha), \alpha]$$

Die Wertfunktion liefert den maximierten Wert der Zielfunktion für einen gegebenen Wert des Parameters α .

Beispiel 5.18. Die Profitfunktion einer Unternehmung, die mit einem Produktionsfaktor (Arbeit) produziert ist gegeben durch

$$\pi(L) = pf(L) - wL$$

wobei p der Preis des Produktes ist und w der an den Faktor Arbeit (L) zu zahlende Lohn.

Die Bedingung erster Ordnung ist

$$\begin{aligned} \pi'(L; p, w) = pf'(L) - w = 0 &\implies pf'(L) = w \\ &\implies L^*(p, w) \end{aligned}$$

Die Wertfunktion ist gleich

$$V(p, w) = \pi[L^*(p, w), p, w]$$

5.3.1 Envelope Theorem bei Maximierung ohne Nebenbedingungen

Wir interessieren uns nun dafür wie sich dieser Maximalwert verändert, wenn sich der Parameter α verändert. Die Veränderung der Wertfunktion mit α wird durch die erste Ableitung der Wertfunktion nach α dargestellt. Diese ist gleich

Theorem 5.8 (Envelope Theorem ohne Nebenbedingungen).

$$\frac{dV(\alpha)}{d\alpha} = f_x \frac{dx^*}{d\alpha} + f_y \frac{dy^*}{d\alpha} + f_\alpha = f_\alpha$$

Der letzte Schritt resultiert aus der Realisierung, dass am stationären Punkt $f_x = f_y = 0$

Das Ergebnis, dass die Veränderung der Wertfunktion in Bezug auf α gleich der partiellen Ableitung der Zielfunktion am optimalen Punkt nach α entspricht wird als Envelope Theorem bezeichnet.

Beispiel 5.19. Gegeben ist die Funktion $y = -x^2 + ax$. Wir wollen wissen, wie sich der optimale Wert y^* mit a verändert. 2 Möglichkeiten

1. Wir optimieren:

(a) Wir finden die erste Ableitung nach x , setzen gleich Null und lösen für das optimale x^*

$$-2x + a = 0 \quad \implies \quad x^* = \frac{1}{2}a$$

(b) Wir setzen x^* in die Zielfunktion ein und erhalten

$$V(a) = -\left(\frac{1}{2}a\right)^2 + a\frac{1}{2}a = -\frac{1}{4}a^2 + \frac{1}{2}a^2 = \frac{1}{4}a^2$$

(c) Wir bilden $\frac{\partial V(a)}{\partial a} = \frac{1}{2}a = x^*$

2. Wir benutzen das Envelope Theorem

(a) Bilden der Wertfunktion

$$V(a) = f[x^*(a), a] = -(x^*)^2 + ax^*$$

(b) Laut Envelope Theorem

$$\frac{\partial V(a)}{\partial a} = \frac{\partial f(x^*(a), a)}{\partial a} = x^*$$

Beispiel 5.20. Gegeben sei die Funktion

$$f(x) = -x^2 + 2ax + 4a^2$$

Bestimmen sie $\frac{dV(a)}{da}$ durch Ableiten der Wertfunktion und mittels des Envelope-Theorems.

Zum Abschluss dieses Abschnitts noch eine VWL-Beispiel

Beispiel 5.21. (Hotelling's Lemma: Die Profitfunktion) Eine Firma maximiert ihren Profit, der gegeben ist durch

$$\pi(K, L) = P \cdot f(K, L) - wL - rK$$

. Wobei L und K die Mengen von Arbeit und Kapital darstellen. Die Parameter des Problems sind der Lohn (w), der Preis des Kapitals (r) und der Produktpreis (P). Die Maximierung des Profits mit den gegebenen

Parametern führt zu der Wertfunktion (=Profitfunktion)

$$V(w, r, P) = \pi[K^*(w, r, P), L^*(w, r, P)]$$

Benutzen sie das Envelope Theorem um die folgenden partiellen Ableitungen zu finden

$$\frac{dV(w, r, P)}{dw} \quad \frac{dV(w, r, P)}{dr} \quad \frac{dV(w, r, P)}{dP}$$

Dieses Ergebnis ist als Hotelling's Lemma bekannt und zeigt

- dass das Angebots der Firma gleich der partiellen Ableitung der Profitfunktion nach dem Produktpreis ist.
- dass die Nachfrage nach Arbeit gleich der partiellen Ableitung der Profitfunktion nach dem Lohn ist multipliziert mit -1 ist.
- dass die Nachfrage nach Kapital gleich der partiellen Ableitung der Profitfunktion nach dem Preis des Kapitals ist multipliziert mit -1 ist.

5.4 Multivariate Optimierung unter Nebenbedingungen

In den Wirtschaftswissenschaften müssen bei Optimierungsproblemen oft auch noch Nebenbedingungen erfüllt sein. Allgemein schreiben wir

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{unter den Nebenbedingungen} \quad g_i(\mathbf{x}) = (\leq) \mathbf{c} \quad \text{für } i = 1..k$$

Ein Beispiel ist die Nutzenmaximierung. Da ein Individuum (auch nicht Bill Gates) nicht unbegrenzt konsumieren kann, ist hier die Nebenbedingung, dass die gesamten Konsumausgaben entweder kleiner oder gleich dem Einkommen des Individuums sind. Formal

$$\max_{\mathbf{x}} u(\mathbf{x}) \quad \text{unter der Nebenbedingung} \quad \mathbf{p}\mathbf{x} = p_1x_1 + p_2x_2 + \dots + p_nx_n (\leq \text{ oder } =) I$$

Hier sind p_i die Preise der Güter, x_i die Gütermengen, I das Einkommen und $u(\mathbf{x})$ eine Funktion, die den Gesamtnutzen der konsumierten Güter als Funktionswert hat (Nutzenfunktion).

Die Nebenbedingungen können entweder Gleichungsnebenbedingungen ($g_i(\mathbf{x}) = \mathbf{c}$) oder Ungleichungsnebenbedingungen ($g_i(\mathbf{x}) \leq \mathbf{c}$) sein. Wir behandeln in dieser Vorlesung nur den ersten Fall. Im Falle von Gleichheitsnebenbedingungen werden zwei Methoden behandelt:

1. Substitutionsmethode
2. Methode der Lagrange-Multiplikatoren

5.4.1 Substitutionsmethode

Bei der Substitutionsmethode ist es möglich die Gleichungsnebenbedingungen direkt in die zu optimierende Zielfunktion einzusetzen und somit ein Optimierungsproblem *ohne* Nebenbedingungen zu erzeugen. Dies wird am besten durch ein Beispiel dargestellt.

Beispiel 5.22. Ein Individuum konsumiert 2 Güter, deren Mengen mit x und y dargestellt werden. Gegeben sei die Nutzenfunktion

$$u(x, y) = xy$$

. Das Individuum hat ein Einkommen in Höhe von I und die Preise der Güter sind $p_x = 2$ und $p_y = 4$. Wir nehmen an, dass mehr Konsum immer den Gesamtnutzen erhöht. Somit gilt die Nebenbedingung

$$p_x x + p_y y = I$$

Die Nebenbedingung kann nach y aufgelöst werden

$$y = \frac{I}{p_y} - \frac{p_x}{p_y} x$$

Wir können nun in der Zielfunktion für y substituieren

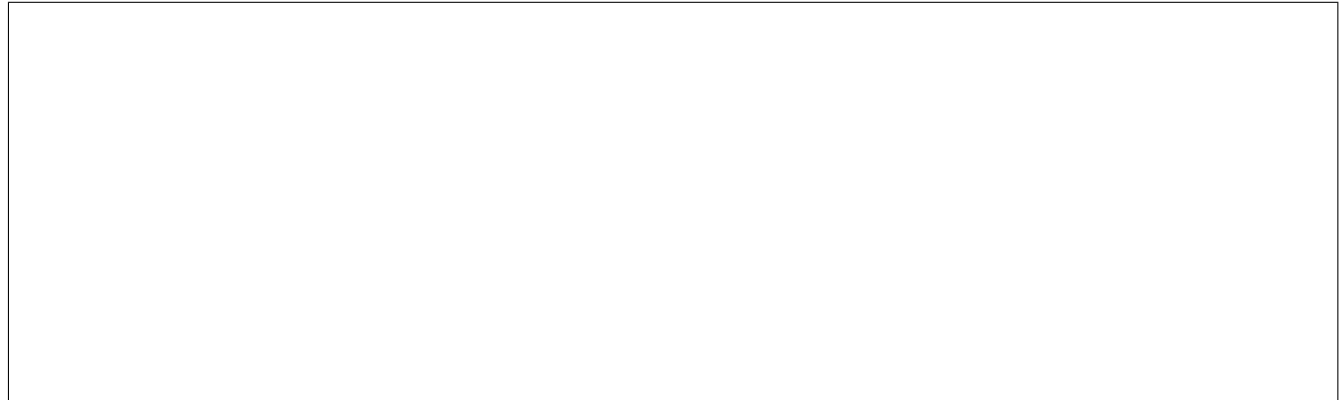
$$u(x) = x \left(\frac{I}{p_y} - \frac{p_x}{p_y} x \right) = \frac{I}{p_y} x - \frac{p_x}{p_y} x^2$$

Wir maximieren die (konkave) Funktion in dem wir die erste Ableitung gleich Null setzen und nach x und y (über die Gleichung der Nebenbedingung) auflösen³

³Wir müssen hier keine Bedingungen zweiter Ordnung prüfen. Warum?

$$u'(x) = \frac{I}{p_y} - 2\frac{p_x}{p_y}x = 0 \quad \implies \quad x^* = \frac{I}{2p_x} \text{ und } y^* = \frac{I}{2p_y}$$

Beispiel 5.23. *Beispiel: Maximieren sie $h(x, y) = 3y - 8x^2$ unter der Nebenbedingung $y - 6x = 4$.*



Mit der Substitutionsmethode lässt sich gut arbeiten, wenn die Zielfunktion und die Nebenbedingungen relativ einfache Funktionen sind. Wenn dies nicht der Fall ist und/oder mehrere Nebenbedingungen zu beachten sind, ist die Methode der Lagrange Multiplikatoren vorzuziehen, die jetzt behandelt wird. Bevor wir uns dieser Methode widmen, machen wir noch einen kurzen Ausflug ins Land der nicht-linearen Gleichungssysteme.

5.5 Exkurs: Nicht-Lineare Gleichungssysteme

In der später behandelten, in den Wirtschaftswissenschaften ungemein wichtigen Methode der LaGrange-Multiplikatoren taucht das folgende Problem auf

$$\begin{aligned} f_1(x, y, \lambda) &= 0 \\ f_2(x, y, \lambda) &= 0 \\ f_3(x, y) &= 0 \end{aligned}$$

Hierbei repräsentieren f_1, f_2, f_3 beliebige Gleichungen. Es geht also darum, ein Gleichungssystem mit 3 Gleichungen und 3 Variablen zu lösen. Wir behandeln hier nur den Spezialfall, in dem die dritte Gleichung nur die ersten zwei Variablen enthält, da wir bei der Lagrange-Methode mit solchen Systemen konfrontiert werden.

Ein Gleichungssystem lösen? Was heißt das überhaupt und wie kann man hier vorgehen? Eine Lösung besteht aus einem 3-Tupel (x^*, y^*, z^*) , die ALLE drei Gleichungen simultan erfüllt!

Was sind LINEARE Gleichungen? In linearen Gleichungen kommen die Variablen, deren Werte gesucht werden, nur in der ersten Potenz vor und sind nicht mit den anderen Variablen multiplikativ verknüpft. Beispiele:

$$10 - 5x = 9 \quad 5x = 10 \quad 5x + 3y = 0 \quad 4x + 6y - 7z = 8$$

Wir werden uns später mit dem Lösen von Gleichungssystemen beschäftigen, die nur lineare Gleichungen beinhalten.

Was sind NICHT-LINEARE Gleichungen? Antwort: Alle Gleichungen außer linearen Gleichungen! Hier geht alles! Beispiele:

$$ax + b\sqrt{x} = 4 \quad ax + \sqrt{y} = 2 \quad x^2 - e^x + 5 = 0 \quad \frac{ax^2 - 5e^y}{\sqrt{\ln z - y^3}} = 0$$

Im einführenden Beispiel haben wir bewusst die abstrakte Bezeichnung $f(x, y, z) = 0$ für eine Gleichung des Gleichungssystems benutzt, um darauf hinzuweisen, dass es sich hier um nicht-lineare Gleichungen handeln kann.

Definition 5.20. Ein Nicht-Lineares Gleichungssystem ist ein Gleichungssystem, das mindestens eine nicht-lineare Gleichung enthält.

Die schlechte Nachricht: Nicht jedes Gleichungssystem hat eine Lösung und bei nicht-linearen Gleichungssystemen gibt es nicht einmal eine allgemeine Methode, mit der man garantiert herausfinden kann, ob eine Lösung existiert.

Die gute Nachricht: Die von uns behandelten nicht-linearen Gleichungssysteme haben eine Lösung. Unsere Lösungsmöglichkeiten

- Die Variable λ aus den ersten zwei Gleichungen eliminieren, nach x (oder y) auflösen und in die dritte Gleichung einsetzen.
- Rumprobieren: Einen Wert einer Variablen finden, der eine der Gleichungen erfüllt und dann überprüfen, ob dieser Wert mit den anderen Gleichungen kompatibel ist.

Notiz:

- Keine der Möglichkeiten funktioniert garantiert.
- Besonders bei der ersten Methode ist besonders darauf zu achten, dass nicht durch NULL geteilt wird!

Es gilt hier besonders zu verstehen, dass eine Lösung ALLE drei Gleichungen erfüllen muss! Wenn ein Wert für zum Beispiel x eine Gleichung erfüllt, dann muss immer noch kontrolliert werden, ob dieser Wert auch mit den restlichen Gleichungen kompatibel ist.

Die folgenden drei Beispiele sollen uns helfen, ein Gefühl für das Lösen zu entwickeln. Der erste Weg ist relativ standardisiert und in vielen wirtschaftswissenschaftlichen Problemen einsetzbar. Es gilt aber immer die Devise: Nicht durch NULL teilen!

Beispiel 5.24. Lösen Sie das nicht-lineare Gleichungssystem. Kann es sein, dass $x = 0$ oder $y = 0$ oder $\lambda = 0$

$$5 - 2\lambda x = 0 \tag{1}$$

$$-3 - 2\lambda y = 0 \tag{2}$$

$$x^2 + y^2 = 136 \tag{3}$$

- Wenn $x = 0$ in (1), dann steht dort $5 = 0$. Geht nicht.
- Wenn $y = 0$ in (2) dann muss gelten $-3 = 0$. Geht nicht.
- Aus den gleichen Gründen scheidet $\lambda = 0$ als Lösung aus.

- Wir schlussfolgern, dass weder $x = 0$ noch $y = 0$ noch $\lambda = 0$ Teil einer Lösung sein können.
- Wir schreiben die ersten zwei Gleichungen um und teilen (1) durch (2)

$$5 = 2\lambda x \quad -3 = 2\lambda y \implies \frac{5}{-3} = \frac{2\lambda x}{2\lambda y} = \frac{x}{y}$$

- Wir lösen auf zu $x = -\frac{5}{3}y$, setzen in (3) ein und erhalten

$$\left(-\frac{5}{3}y\right)^2 + y^2 = 136$$

- Dies liefert die Lösungen $y = \pm 6$. Daraus folgt $x = \pm 6$ und $\lambda = \pm \frac{1}{4}$
- Die Lösungen dieses Systems sind also

$$(x^*, y^*, \lambda^*) = \left(10, -6, \frac{1}{4}\right) \quad \text{und} \quad \left(-10, 6, \lambda = -\frac{1}{4}\right)$$

Beispiel 5.25. Lösen Sie das nicht-lineare Gleichungssystem. Kann es sein, dass $x = 0$ oder $y = 0$ oder $\lambda = 0$

$$y - 2\lambda = 0 \tag{1}$$

$$x - 5\lambda = 0 \tag{2}$$

$$2x + 5y = 100 \tag{3}$$

- Wenn $x = 0$ in (2), dann muss gelten $\lambda = 0$. Dann gilt auch $y = 0$ in (1) und die dritte Gleichung kann nicht erfüllt werden.
- Wenn $y = 0$, dann muss gelten $\lambda = 0$. Dann gilt auch $x = 0$ und die dritte Gleichung kann nicht erfüllt werden.
- Wenn $\lambda = 0$, dann muss gelten $x = y = 0$ und die dritte Gleichung kann nicht erfüllt werden.
- Wir schlussfolgern, dass weder $x = 0$ noch $y = 0$ noch $\lambda = 0$ Teil einer Lösung sein können.
- Wir schreiben die ersten zwei Gleichungen um und teilen (1) durch (2)

$$y = 2\lambda \quad x = 5\lambda \implies \frac{y}{x} = \frac{2\lambda}{5\lambda} = \frac{2}{5}$$

- Wir lösen auf zu $y = \frac{2}{5}x$, setzen in (3) ein und erhalten

$$2x + 5\left(\frac{2}{5}x\right) = 4x = 100$$

- Dies liefert die Lösungen $x = 25$. Daraus folgt $y = 10$ und $\lambda = 5$
- Die Lösung dieses Systems ist also

$$(x^*, y^*, \lambda^*) = (25, 10, 5)$$

Unser drittes Beispiel zeigt, dass man manchmal einfach ausprobieren muss, um zur Lösung zu gelangen.

Beispiel 5.26. Lösen Sie das nicht-lineare Gleichungssystem. Kann es sein, dass $x = 0$ oder $y = 0$ oder $\lambda = 0$?

$$2x - 2x\lambda = 0 \quad (1)$$

$$8y - 2y\lambda = 0 \quad (2)$$

$$1 - x^2 - y^2 = 0 \quad (3)$$

- Nun löst $x = 0$ die erste Gleichung und wir müssen schauen, ob dies auch mit den anderen Gleichungen kompatibel ist.
- In (3) führt $x = 0$ zu $y = \pm 1$ und somit in (2) zu $\lambda = 4$
- $y = 0$ löst die zweite Gleichung und wir müssen schauen, ob dies auch mit den anderen Gleichungen kompatibel ist.
- In (3) führt $y = 0$ zu $x = \pm 1$ und somit in (1) zu $\lambda = 1$
- Was ist mit $x \neq 0$? Dann gilt laut (1) $\lambda = 1$. Laut (2) ist dann $y = 0$ und die dritte Gleichung wird nur von $x = \pm 1$ erfüllt. Diese Lösung haben wir schon gefunden.
- Was ist mit $y \neq 0$? Dann gilt laut (2) $\lambda = 4$. Laut (1) ist dann $x = 0$ und die dritte Gleichung wird nur von $y = \pm 1$ erfüllt. Diese Lösung haben wir schon gefunden.
- Die Lösungen des Systems sind

$$(x^*, y^*, \lambda^*) = (0, 1, 4) \quad (0, -1, 4) \quad (1, 0, 1) \quad (-1, 0, 1)$$

Weiter geht es mit unserem eigentlichen Thema.

5.5.1 Optimierung unter Gleichheitsnebenbedingungen: Methode der Lagrange Multiplikatoren

Der Mathematiker Lagrange hat die folgende Methode zur Lösung von Optimierungsproblemen mit Gleichheitsnebenbedingungen entwickelt. Wir beginnen einer Beschreibung der Methode und lassen dann ein Beispiel folgen.

Gegeben sei das (2-dimensionale) Optimierungsproblem

$$\max_{x,y} f(x,y) \text{ u.d.N. } g(x,y) = c \quad (4)$$

(u.d.N. = unter den Nebenbedingungen) Die Methode der Lagrange Multiplikatoren durchläuft die folgenden Schritte (egal ob Minimierung oder Maximierung)

1. Bildung der Lagrange Funktion

$$\mathbb{L}(x,y,\lambda) = f(x,y) + \lambda(c - g(x,y)) \quad (5)$$

Diese Gleichung ist der zentrale Punkt des Ansatzes.

Sie führt die Lagrange-Multiplikatoren (λ) ein, deren Existenz an Extrema der Zielfunktion im Theorem von Lagrange bewiesen werden. Das Theorem liefert die notwendigen Bedingungen für ein Extremum, die im nächsten Punkt beschrieben werden.

2. Bildung der ersten partiellen Ableitungen der Lagrange-Funktion nach x , y und λ . Nullsetzen liefert uns die drei Bedingungen erster Ordnung und resultiert in drei Gleichungen in drei Unbekannten (x, y, λ)

$$\frac{\partial \mathbb{L}}{\partial x} = f_x(x, y) - \lambda g_x(x, y) = 0 \quad (6)$$

$$\frac{\partial \mathbb{L}}{\partial y} = f_y(x, y) - \lambda g_y(x, y) = 0 \quad (7)$$

$$\frac{\partial \mathbb{L}}{\partial \lambda} = c - g(x, y) = 0 \quad (8)$$

3. Auflösen der Gleichungen nach den Unbekannten x^* , y^* und λ^* , die die Zielfunktion maximieren. Normalerweise werden hier die ersten zwei Gleichungen kombiniert (oft durch Division) und nach x oder y aufgelöst. Dann wird die Nebenbedingung benutzt um die Werte für x und y zu finden. Wir werden dieses Vorgehen weiter unten noch intuitiv/graphisch beleuchten. Zunächst ein Beispiel

Beispiel 5.27. Gegeben sei das Maximierungsproblem mit Nebenbedingung

$$u(x, y) = xy$$

$$\text{u.d.N. } p_x x + p_y y = I$$

die Lagrange-Funktion des Problems ist

$$\mathbb{L} = xy + \lambda(I - p_x x - p_y y)$$

Wir bilden die partiellen Ableitungen nach x, y, λ

$$\frac{\partial \mathbb{L}}{\partial x} = y - \lambda p_x = 0 \quad (9)$$

$$\frac{\partial \mathbb{L}}{\partial y} = x - \lambda p_y = 0 \quad (10)$$

$$\frac{\partial \mathbb{L}}{\partial \lambda} = I - p_x x - p_y y = 0 \quad (11)$$

Wir bringen die Terme mit λ auf die rechte Seite, teilen Gleichung 9 durch Gleichung 10 und lösen auf nach y

$$\begin{aligned} \frac{y}{x} &= \frac{\lambda p_x}{\lambda p_y} \\ \frac{y}{x} &= \frac{p_x}{p_y} \\ y &= \frac{p_x}{p_y} x \end{aligned} \quad (12)$$

Schließlich setzen wir das Ergebnis in die Nebenbedingung ein und lösen für die Werte von x und y , die die Zielfunktion unter der Nebenbedingung maximieren.

$$I - p_x x - p_y \frac{p_x}{p_y} x = 0$$

$$\implies x^* = \frac{I}{2p_x} \quad y^* = \frac{I}{2p_y} \quad \lambda^* = \frac{I}{2p_x p_y}$$

Dies ist natürlich die selbe Lösung wie bei der Benutzung der Substitutionsmethode, aber wir haben eine zusätzliche Größe λ berechnet. Im Folgenden werden wir besprechen

1. Die Charakterisierung des Optimalen Punktes (x^*, y^*, λ^*)
2. Die Interpretation des Lagrange Multiplikators λ

Doch zunächst ein weiteres Beispiel.

Beispiel 5.28. Maximieren sie $f(x, y) = 3y - 8x^2$ unter der Nebenbedingung $y - 6x = 4$ mit Hilfe der Lagrange Methode.



Charakterisierung des Optimums Gegeben sei das allgemeine Optimierungsproblem in Gleichung 4 mit der dazugehörigen Lagrange-Funktion (Gleichung 5). Der Quotient der ersten beiden Bedingungen erster Ordnung (Gleichungen 6 und 7) ist gleich

$$\frac{f_x(x, y)}{f_y(x, y)} = \frac{g_x(x, y)}{g_y(x, y)} \quad (13)$$

Laut der Ergebnisse im Abschnitt über das implizite Differenzieren gilt

$$-\frac{f_x(x, y)}{f_y(x, y)} = \text{Steigung einer Höhenlinie von } f(x, y) \text{ im Punkt } (x, y)$$

$$-\frac{g_x(x, y)}{g_y(x, y)} = \text{Steigung der Höhenlinie } g(x, y) = c \text{ im Punkt } (x, y)$$

Ein Optimum liegt also an einem Punkt, an dem die Steigung *einer* Höhenlinie der Zielfunktion und *die* Steigung der Nebenbedingung gleich sind. Diese Bedingung kann man graphisch darstellen. Abbildung 32 zeigt zunächst zwei Punkte, an denen die Bedingung von Gleichung 13 *nicht* erfüllt sind. Wie wir sehen gibt es Punkte wie (x_2, y_2) , die

1. die Nebenbedingung erfüllen (auf $g(x, y) = c$ liegen) **und**
2. auf höheren Höhenlinien liegen

Die Punkte (x_0, y_0) und (x_1, y_1) können somit nicht optimal sein.

Abbildung 32: Kein Maximum an (x_0, y_0) und (x_1, y_1)

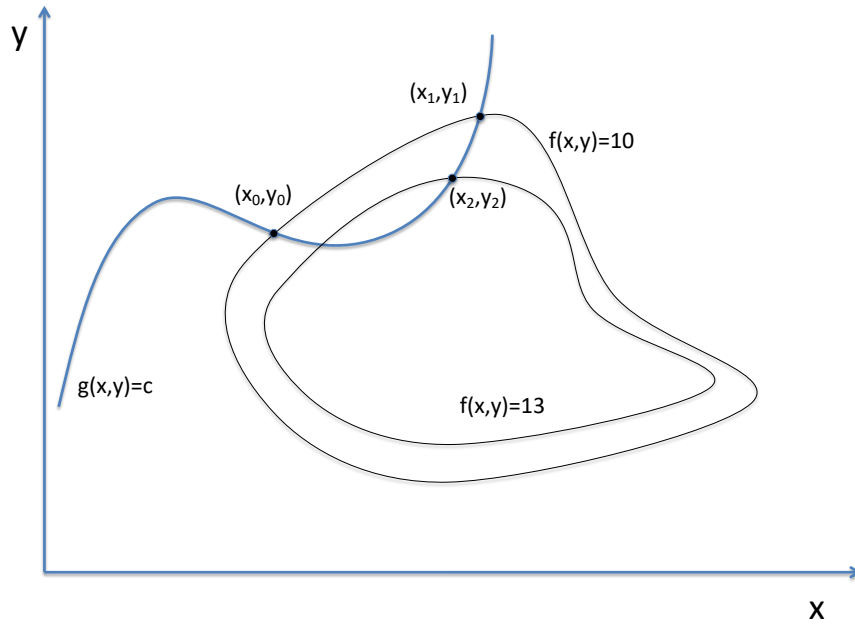
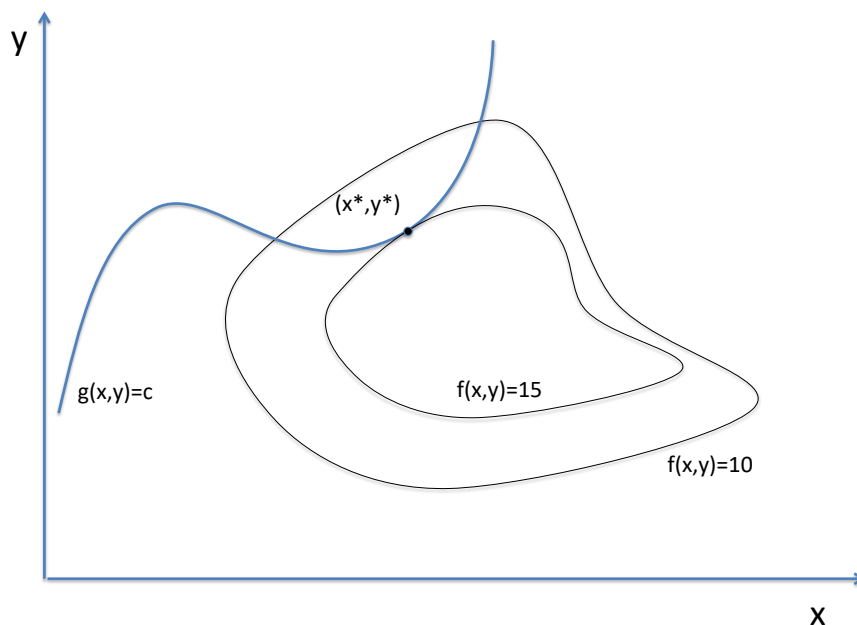


Abbildung 33 zeigt den optimalen Punkt, an dem die Steigung einer Höhenlinie der Zielfunktion gleich der Steigung der Nebenbedingung ist. Am Maximum (x^*, y^*) ist ersichtlich, dass

1. es nicht möglich ist eine Höhenlinie mit einem größeren Funktionswert zu erreichen.
2. jede Veränderung, die auf der Höhenlinie $g(x, y) = c$ bleibt, notwendigerweise zu einem niedrigerem Funktionswert der Zielfunktion führt

Abbildung 33: Maximum an (x^*, y^*)



Beispiel 5.29. Lösen sie das folgende Optimierungsproblem mittels der Lagrange-Methode

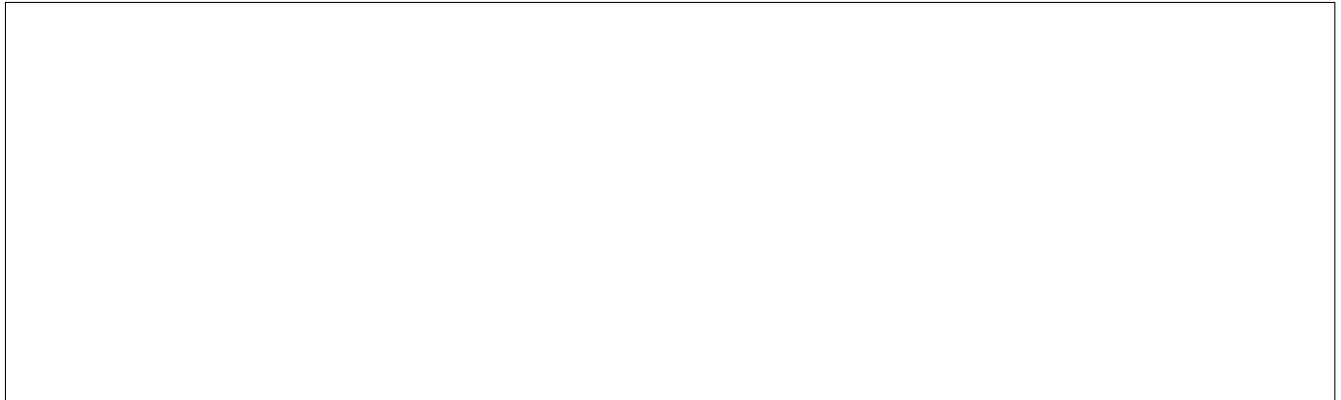
$$f(x, y) = x^2 + 4y^2$$

$$\text{u.d.N. } x^2 + y^2 = 1$$



Wir werden nun noch eine Optimierung durchführen mit einer Funktion, die oft in den Wirtschaftswissenschaften benutzt wird und bei der die graphische Darstellung relativ einfach ist.

Beispiel 5.30. Maximieren sie die Nutzenfunktion $u(x, y) = x^{\frac{1}{2}}y^{\frac{1}{2}}$ unter der Nebenbedingung $I = p_x x + p_y y$.



Wir werden nun die im Beispiel 5.30 gefundene Lösung graphisch darstellen. Dabei beleuchten wir wie die drei Bedingungen erster Ordnung zusammenspielen um die Lösung des Problems zu finden. Die Optimalitätsbedingung, dass am optimalen Punkt die Steigung einer Höhenlinie gleich der Steigung der Nebenbedingung sein muss ist in diesem Fall

$$\frac{y}{x} = \frac{p_x}{p_y}$$

Wir können diese Gleichung umformen zu

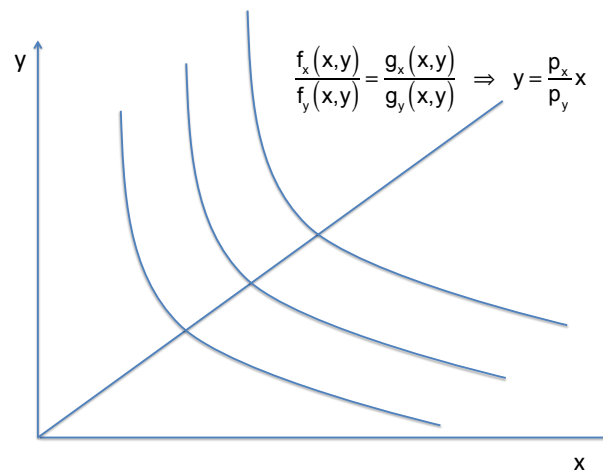
$$y = \frac{p_x}{p_y} x$$

Dies ist eine lineare Funktion ⁴ und kann wie folgt interpretiert werden. Die Funktion ist die Men-

⁴Die Linearität dieser Beziehung ist ein Spezialfall und nicht allgemein gültig.

ge aller Punkte, die die ersten zwei Bedingungen erster Ordnung erfüllen. An jedem dieser Punkte gleicht die Steigung einer Höhenlinie der Steigung der Nebenbedingung. Abbildung 34 verdeutlicht dies.

Abbildung 34

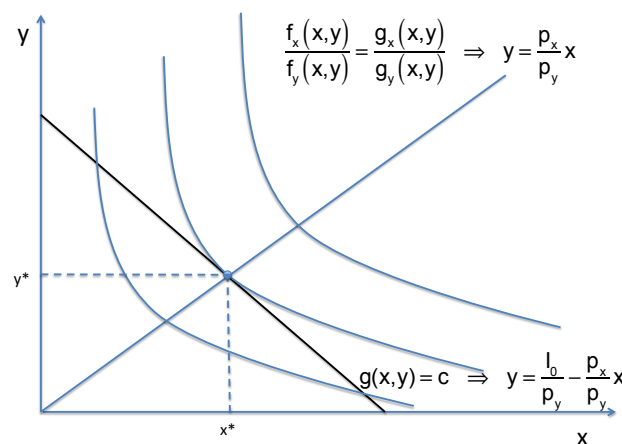


Doch welcher Punkt ist nun optimal? Um den optimalen Punkt zu finden müssen wir die dritte Bedingung erster Ordnung, die Nebenbedingung, zu Rate ziehen, die in unserem Fall auch eine lineare Funktion ist.

$$y = \frac{I}{p_y} - \frac{p_x}{p_y} x$$

Abbildung 35 zeigt die Nebenbedingung für einen speziellen Wert von I . Wir sehen, dass nun alle drei Bedingungen erster Ordnung zusammen den optimalen Punkt bestimmen. Zusätzlich zeigen die beiden Abbildungen auch die Höhenlinien der Funktion.

Abbildung 35



Abschließend soll noch darauf hingewiesen werden, dass die Methode nach Lagrange nur dann mit Sicherheit funktioniert (d.h. die Extrema sind unter den stationären Stellen des Problems), wenn die gesuchten Extrema existieren und der Gradientvektor der Nebenbedingung am Punkt

(x^*, y^*) ungleich dem 0-Vektor ist

$$\nabla g(x^*, y^*) \begin{bmatrix} g_x(x^*, y^*) \\ g_y(x^*, y^*) \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Wir werden diesen Fall ($\nabla g(x^*, y^*) = 0$) aber nicht weiter verfolgen.

5.5.2 Interpretation des Lagrange Multiplikators

Betrachten wir das Problem

$$\max_{x,y} f(x,y) \quad \text{u.d.N.} \quad g(x,y) = c$$

Wir sind nun daran interessiert wie sich der maximierte Wert der Zielfunktion ändert, wenn sich der Wert der Nebenbedingung c ändert. Dazu finden wir die Werte von $x^*(c)$ und $y^*(c)$, die die Zielfunktion maximieren. Diese sind natürlich abhängig vom Parameter c . Wenn wir die Maximalen in die Zielfunktion einsetzen erhalten wir (die Wertfunktion/Optimalwertfunktion)

$$F(c) = f[x^*(c), y^*(c)]$$

Die Wertfunktion

$$F(c)$$

ist nur noch eine Funktion von c . Wir können nun mittels der Kettenregel nach c differenzieren

$$\frac{dF(c)}{dc} = f_x \frac{dx}{dc} + f_y \frac{dy}{dc}$$

Von den Bedingungen erster Ordnung (Gleichungen 6 und 7) des Lagrange Optimierungsproblems wissen wir, dass am Optimum folgendes gilt (für bessere Lesbarkeit schreiben wir $x^* = x^*(c)$ und $y^* = y^*(c)$)

$$f_x = \lambda^*(c)g_x(x^*, y^*) \quad f_y = \lambda g_y(x^*, y^*)$$

Einsetzen in Gleichung 5.5.2 liefert

$$\begin{aligned} \frac{dF(c)}{dc} &= f_x \frac{dx^*}{dc} + f_y \frac{dy^*}{dc} \\ &= \lambda^*(c)g_x(x^*, y^*) \frac{dx^*}{dc} + \lambda^*(c)g_y(x^*, y^*) \frac{dy^*}{dc} \\ &= \lambda^*(c) \left[g_x(x^*, y^*) \frac{dx^*}{dc} + g_y(x^*, y^*) \frac{dy^*}{dc} \right] \end{aligned}$$

Der Term in Klammern ist gleich der totalen Ableitung der Nebenbedingung $g(x^*, y^*) = c$ nach c . Also gilt

$$\frac{dg(x^*, y^*)}{dc} = g_x(x^*, y^*) \frac{dx^*}{dc} + g_y(x^*, y^*) \frac{dy^*}{dc} = 1$$

Somit ist der gesuchte Einfluss des Parameters c auf den Wert der Zielfunktion am Optimum gleich

$$\frac{df(x^*(c), y^*(c))}{dc} = \lambda^*(c)$$

$\lambda^*(c)$ gibt also Auskunft darüber wie sich der Wert der Zielfunktion im Optimum verändert, wenn man den Wert von c um eine kleine Einheit verändert. Besonders in den Wirtschaftswissenschaften hat dieser Parameter wichtige Anwendungen.

Beispiele:

- Nutzenmaximierung: Hier ist λ gleich dem Grenznutzen des Einkommens, da es anzeigt wie sich der maximale Nutzen verändert, wenn sich das Einkommen um eine kleine Einheit verändert.

$$\max_{x_1, x_2} u(x_1, x_2) \quad \text{u.d.N.} \quad p_1 x_1 + p_2 x_2 = \text{Inc}$$

- Kostenminimierung: Hier minimiert eine Firma ihre Kosten unter der Nebenbedingung das eine bestimmte Menge an Output produziert wird. λ repräsentiert hier die Grenzkosten.

$$\min_{x, y} C = wL + rK \quad \text{u.d.N.} \quad Y(K, L) = \bar{y}$$

Beispiel 5.31. Gegeben sei das Maximierungsproblem mit Nebenbedingung

$$u(x, y) = xy \quad \text{u.d.N.} \quad p_x x + p_y y = I$$

mit $p_x = 2$ und $p_y = 4$. Bestimmen sie die Maximalnutzen und $\lambda(I_1)$ für die Einkommen $I_1 = 100$ und $I_2 = 101$. Die Lagrange-Funktion des Problems ist

$$\mathbb{L} = xy + \lambda(I - p_x x - p_y y)$$

Laut Gleichung 12 muss im Optimum gelten

$$y = \frac{p_x}{p_y} x = \frac{2}{4} x$$

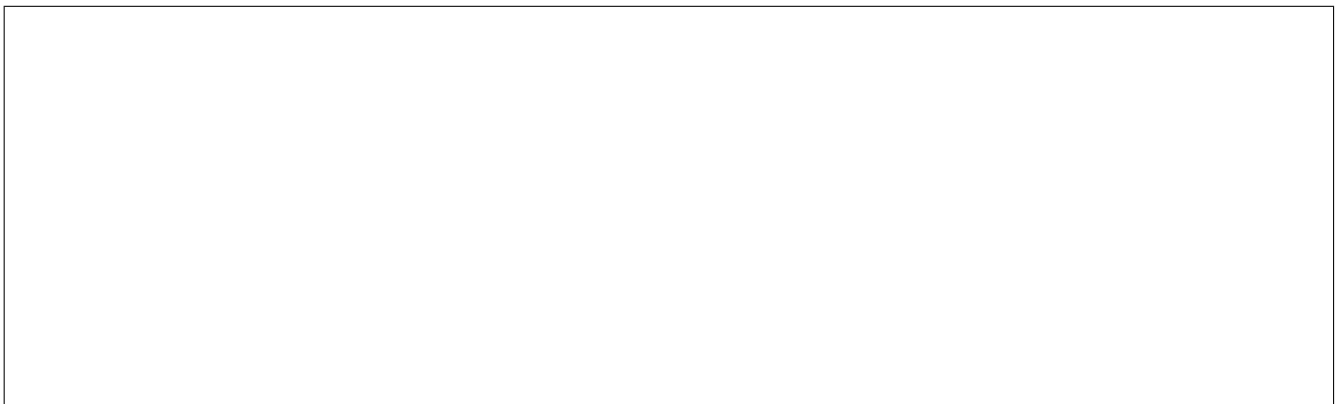
Zusätzlich wird $\lambda^*(c)$ allgemein bestimmt durch (siehe Gleichungen 6 und 7)

$$\lambda^*(c) = \frac{f_x}{g_x} = \frac{f_y}{g_y}$$

- Lösung für $I = 100$

Wir setzen ein in die Nebenbedingung $2x + 4y = 100$ und lösen für

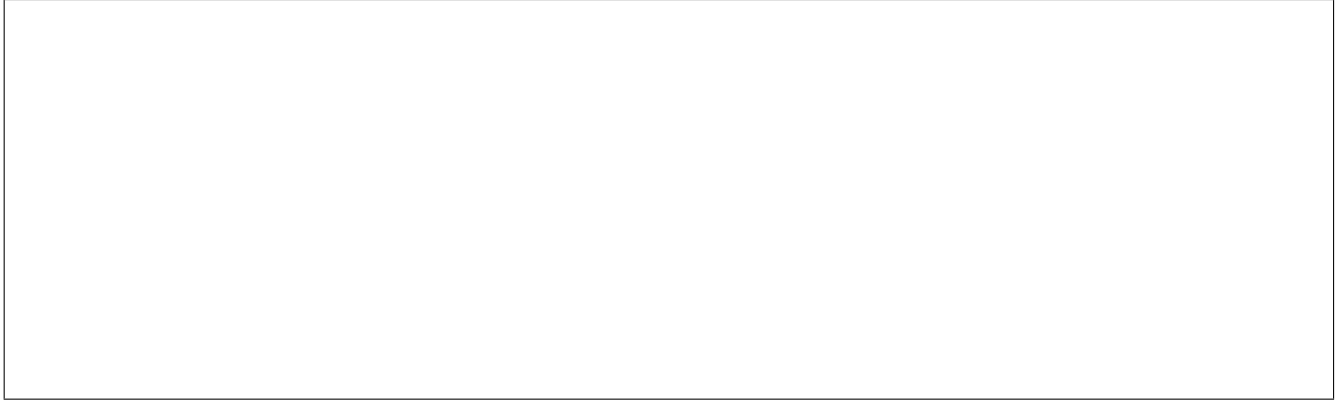
$$x^*(I = 100) = 25 \quad y^*(I = 100) = 12,5 \quad \lambda^*(I = 100) = 6,25 \quad u(x_1^*, x_2^*) = 312,5$$



- Lösung für $I = 101$

Wir setzen ein in die Nebenbedingung $2x + 4y = 101$ und lösen für

$$x^*(I = 101) = 25,5 \quad y^*(I = 101) = 12,625 \quad u(x_1^*, x_2^*) = 318,78$$



Die Differenz zwischen den beiden Maximalnutzen ist $318,78 - 312,5 = 6,28 \approx 6,25 = \lambda^*(I = 100)$
Diese wäre natürlich noch präziser, wenn wir die Veränderung der Nebenbedingung kleiner halten.

Abschließend wollen wir (wie im Fall der Optimierung ohne Nebenbedingungen) noch den Fall beleuchten, bei dem die Zielfunktion konkav/oder konvex ist.

Theorem 5.9. *Konkavität/Konvexität der Lagrange Funktion: Gegeben sei ein Optimierungsproblem mit Gleichungsnebenbedingungen. Gegeben sei ein stationärer Punkt der Lagrange Funktion (x^*). Wenn die Lagrange Funktion*

1. *konkav ist, löst x^* das Maximierungsproblem*
2. *konvex ist, löst x^* das Minimierungsproblem*

5.5.3 Envelope Theorem bei Maximierung mit Nebenbedingungen

Gegeben sei ein Optimierungsproblem mit Nebenbedingung, das einen Parameter α beinhaltet

$$\max_{x,y} f(x,y;\alpha) \quad \text{u.d.N} \quad g(x,y;\alpha) = 0$$

Die Lagrange-Funktion ist gegeben durch

$$\mathbb{L}(x,y,\lambda) = f(x,y;\alpha) + \lambda g(x,y;\alpha)$$

Nach der LaGrange Methode erhalten wir die Lösungen

$$x^*(\alpha) \quad y^*(\alpha) \quad \lambda^*(\alpha)$$

und die Wertfunktion

$$V(\alpha) = f[x^*(\alpha), y^*(\alpha), \alpha]$$

Auch hier sind wir an der Veränderung des Maximalwertes in Bezug auf α interessiert

$$\frac{dV(\alpha)}{d\alpha} = f_x \frac{dx^*}{d\alpha} + f_y \frac{dy^*}{d\alpha} + f_\alpha \quad (14)$$

Hier sind f_x und f_y ungleich Null, aber wir können die Nebenbedingung $g[x^*(\alpha), y^*(\alpha), \alpha] = 0$ evaluiert am Optimum nach α ableiten und bekommen

$$\frac{dg[x^*(\alpha), y^*(\alpha), \alpha]}{d\alpha} = g_x \frac{dx^*}{d\alpha} + g_y \frac{dy^*}{d\alpha} + g_\alpha = 0$$

Wir können diese Gleichung mit λ^* malnehmen und zu (14) addieren. Das Resultat ist

$$\begin{aligned} \frac{dV(\alpha)}{d\alpha} &= f_x \frac{dx^*}{d\alpha} + f_y \frac{dy^*}{d\alpha} + f_\alpha + \lambda^* g_x \frac{dx^*}{d\alpha} + \lambda^* g_y \frac{dy^*}{d\alpha} + \lambda^* g_\alpha \\ &= \underbrace{(f_x + \lambda^* g_x)}_{=0} \frac{dx^*}{d\alpha} + \underbrace{(f_y + \lambda^* g_y)}_{=0} \frac{dy^*}{d\alpha} + f_\alpha + \lambda^* g_\alpha \\ &= f_\alpha + \lambda^* g_\alpha \end{aligned}$$

In diesem Fall besagt das Envelope Theorem, dass die Veränderung der Wertfunktion in Bezug auf α gleich der partiellen Ableitung der LaGrange Funktion nach α ist.

Notiz: Wenn der Parameter (hier α) nicht in der Nebenbedingung vorkommt gilt wie im Falle ohne Nebenbedingungen

$$\frac{dV(\alpha)}{d\alpha} = f_\alpha$$

Beispiel 5.32. Gegeben sei die Funktion $f(x,y) = x^a y$ mit $a > 1$ und Nebenbedingung $2x + y = 12$. Bestimmen sie das Maximum unter der gegebenen Nebenbedingung. Die LaGrange Funktion ist

$$\mathbb{L}(x,y,\lambda;a) = x^a y + \lambda(12 - 2x - y)$$

Die Bedingungen erster Ordnung sind

$$\begin{aligned}\frac{\partial \mathbb{L}(x, y, \lambda; a)}{\partial x} &= ax^{a-1}y - \lambda 2 = 0 \\ \frac{\partial \mathbb{L}(x, y, \lambda; a)}{\partial y} &= x^a - \lambda = 0 \\ \frac{\partial \mathbb{L}(x, y, \lambda; a)}{\partial \lambda} &= 12 - 2x - y = 0\end{aligned}$$

Lösen ergibt

$$x^*(a) = \frac{6a}{1+a} \quad y^*(a) = \frac{12}{1+a} \quad \lambda^*(a) = \left(\frac{6a}{1+a}\right)^a$$

und $(0, 12, 0)$

Der Funktionswert an der ersten Stelle ist größer, also ist dort das Maximum. Die Wertfunktion ist gleich

$$V(a) = \left(\frac{6a}{1+a}\right)^a \left(\frac{12}{1+a}\right)$$

Die Ableitung der Wertfunktion ist direkt nicht leicht zu finden (zu Hause machen). Wir können aber das Envelope Theorem benutzen welches besagt, dass

$$\frac{dV(a)}{da} = \frac{\partial \mathbb{L}(x^*(a), y^*(a), a)}{\partial a}$$

Also, bilden wir die partielle Ableitung

$$\frac{\partial (x^*)^a y^*}{\partial a} = (x^*)^a \ln x^* y^* = \left(\frac{6a}{1+a}\right)^a \ln \left(\frac{6a}{1+a}\right) \left(\frac{12}{1+a}\right)$$

Beispiel 5.33. (Shepards Lemma) Hier minimiert ein Konsument seine Ausgaben (Zielfunktion) unter einem gegebenen Nutzenniveau (Nebenbedingung). Hier sind die Preise der Güter (p_1, p_2) und das zu erreichende Nutzenniveau (\bar{U}) die Parameter des Minimierungsproblems. Die LaGrange Funktion ist

$$\mathbb{L}(x_1, x_2, \lambda) = p_1 x_1 + p_2 x_2 + \lambda(\bar{U} - U(x_1, x_2))$$

Die Benutzung des Envelope Theorems liefert folgende Ergebnisse, die als Shepard's lemma bekannt sind.

$$\frac{dV(p_1, p_2, \bar{U})}{dp_1} = x_1^* \quad \frac{dV(p_1, p_2, \bar{U})}{dp_2} = x_2^* \quad \frac{dV(p_1, p_2, \bar{U})}{d\bar{U}} = \lambda^*$$

Das Lemma zeigt

- dass der optimale (Hicks'sche) Konsum von Gut 1 gleich der partiellen Ableitung der Ausgabenfunktion nach p_1 ist.
- dass der optimale (Hicks'sche) Konsum von Gut 2 gleich der partiellen Ableitung der Ausgabenfunktion nach p_2 ist.
- dass λ^* die Grenzkosten einer Nutzenveränderung sind.

Ein weiteres Beispiel:

Roy's Identity: Hier wird ein Nutzenmaximierungsproblem mit gegebenen Preisen und Einkom-

men betrachtet. Die LaGrange Funktion ist

$$\mathbb{L}(x_1, x_2, \lambda) = U(x_1, x_2) + \lambda(I - p_1x_1 - p_2x_2)$$

Roy's Identity besagt, dass

$$x_1^* = \frac{\frac{\partial V(p_1, p_2, I)}{\partial p_1}}{\frac{\partial V(p_1, p_2, I)}{\partial I}} \quad x_2^* = \frac{\frac{\partial V(p_1, p_2, I)}{\partial p_2}}{\frac{\partial V(p_1, p_2, I)}{\partial I}}$$

Hier handelt es sich um die optimalen (Marshall'schen) Nachfragen nach den Gütern

6 Lineare Algebra

Wir erinnern uns an das Gleichungssystem

$$\begin{aligned}a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2\end{aligned}$$

Dieses lineare System mit 2 Gleichungen und 2 Unbekannten war leicht zu lösen, aber wie sieht es mit dem folgenden System von m Gleichungen und n Unbekannten (x_1, x_2, \dots, x_m) aus? Zur Erinnerung sind a_{ij} und b_i Koeffizienten und die x_j die Unbekannten.

$$\begin{aligned}a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1m}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2m}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3m}x_n &= b_3 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nm}x_n &= b_m\end{aligned}$$

Wir werden sehen, dass für dieses Gleichungssystem drei Möglichkeiten bestehen

1. Es gibt genau eine Lösung
2. Es gibt keine Lösung
3. Es gibt unendlich viele Lösungen

Bevor wir lernen diese Fälle zu unterscheiden, transformieren wird das Gleichungssystem in Matrix-Form

Es sei A die $m \times n$ Matrix (m Zeilen und n Spalten)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix}$$

und \mathbf{x} und \mathbf{b} die folgenden Vektoren

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}$$

Dann können wir das obige lineare Gleichungssystem schreiben als

$$A\mathbf{x} = \mathbf{b}$$

Definition 6.1 (Konsistentes lineares Gleichungssystem). Gegeben sei das lineare Gleichungssystem

$$A\mathbf{x} = \mathbf{b}$$

mit $n \times m$ Koeffizientenmatrix A und Koeffizientenvektor \mathbf{b} . Eine Lösung dieses Systems ist ein Vektor $m \times 1$ Vektor \mathbf{s} , der in \mathbf{x} eingesetzt alle Gleichungen des Systems simultan löst. Das System heißt

- konsistent, wenn es eine oder unendlich viele Lösungen hat.
- inkonsistent, wenn es keine Lösung hat

6.1 Gauß'sche Elimination

Das Lösungsverfahren der Gaußschen Elimination überführt ein Gleichungssystem in eine äquivalente Form, die sich einfach lösen lässt, die sogenannte Zeilenstufenform

Beispiel 6.1. Gegeben sei das Gleichungssystem

$$\begin{array}{rcl} & 2x_2 & -x_3 & = & -7 \\ x_1 & +x_2 & +3x_3 & = & 2 \\ -3x_1 & +2x_2 & +2x_3 & = & -10 \end{array}$$

Die Zeilenstufenform dieses Systems ist

$$\begin{array}{rcl} x_1 & +x_2 & +3x_3 & = & 2 \\ & x_2 & -\frac{1}{2}x_3 & = & -\frac{7}{2} \\ & & x_3 & = & 1 \end{array}$$

Wie sehen nun was mit Zeilenstufenform gemeint ist

1. x_1 kommt nur in der ersten Gleichung vor und hat den Koeffizienten 1
2. In der zweiten Gleichung kommen nur x_2 und x_3 vor und x_2 hat den Koeffizienten 1
3. x_3 ist die einzige Unbekannte in der dritten Gleichung und hat den Koeffizienten 1

In Matrixform können wir die zwei Gleichungssysteme schreiben als

$$\begin{bmatrix} 0 & 2 & -1 \\ 1 & 1 & 3 \\ -3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -7 \\ 2 \\ 10 \end{bmatrix} \iff \begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -\frac{7}{2} \\ 1 \end{bmatrix}$$

Es sollte klar sein, dass sich das System nun lösen lässt. Die Lösungen sind

$$x_1^* = 2 \quad x_2^* = -3 \quad x_3^* = 1$$

Dieser Lösungsansatz wird Rückwärtseinsetzen genannt, weil man mit dem gefundenen Wert von x_3 anfängt und dann rückwärts durch Einsetzen die Lösungen von x_2 und x_1 findet.

Die Zeilenstufenform der Koeffizientenmatrix ist also eine obere Dreiecksmatrix (= alle Elemente unterhalb der Hauptdiagonalen sind 0). Hierbei ist zu beachten, dass Elemente auf der Hauptdiagonalen auch Null sein können.

Definition 6.2 (Elementare Zeilenoperationen). Wir können folgende Operationen an einem Gleichungssystem durchführen ohne die Lösung des Systems zu verändern. Die Idee ist, ein lineares Gleichungssystem durch ein anderes zu ersetzen, dass einfacher zu lösen ist. Elementare Zeilenoperationen sind:

1. Der Tausch von zwei Zeilen
2. Die Multiplikation einer Zeile mit einem Skalar $\lambda \neq 0$
3. Die Addition des vielfachen einer Zeile zu einer anderen Zeile

Es ist das Ziel der Gaußschen Eliminierung ein Gleichungssystem durch elementare Zeilenoperationen in die Zeilenstufenform zu überführen. Natürlich führt dieses Lösungsverfahren nicht zwangsläufig zu einer eindeutigen Lösung. Wir bearbeiten jetzt drei Beispiele, in denen ein simples System nach dem Eliminierungsverfahren eine, keine oder unendlich viele Lösungen hat.

Beispiel 6.2. :

1. Keine Lösung.

Gegeben sei das Gleichungssystem

$$\begin{array}{rcl} x & -2y & = 1 \\ 3x & -6y & = 11 \end{array}$$

Eliminierung (Addition des dreifachen der ersten Zeile zur zweiten) führt zu

$$\begin{array}{rcl} x & -2y & = 1 \\ 0 & \mathbf{0}y & = 8 \end{array}$$

Es ist ersichtlich, dass das System keine Lösung hat. Intuitiv lässt sich das erklären, wenn man sich das Problem graphisch betrachtet. Beide Gleichungen sind die Geraden

$$\begin{array}{l} y = -1 + \frac{1}{2}x \\ y = -\frac{11}{6} + \frac{1}{2}x \end{array}$$

Die zwei Geraden sind parallel zueinander und können sich damit nicht schneiden. Es gibt keine Lösung

2. Unendlich viele Lösungen.

Gegeben sei das Gleichungssystem

$$\begin{array}{rcl} x & -2y & = 1 \\ 3x & -6y & = 3 \end{array}$$

Eliminierung führt zu

$$\begin{array}{rcl} x & -2y & = 1 \\ 0 & \mathbf{0}y & = 0 \end{array}$$

Es ist ersichtlich, dass y nun jeden Wert annehmen kann. Intuitiv lässt sich das erklären, wenn man sich das Problem graphisch betrachtet. Beide Gleichungen sind Geraden

$$\begin{array}{l} y = -\frac{1}{2} + \frac{1}{2}x \\ y = -\frac{1}{2} + \frac{1}{2}x \end{array}$$

Die zwei Geraden sind identisch und haben somit unendlich viele Schnittpunkte, die alle das System lösen.

3. Eine Lösung.

Gegeben sei das Gleichungssystem

$$\begin{array}{rcl} x & -2y & = 1 \\ 3x & -2y & = 5 \end{array}$$

Eliminierung führt zu

$$\begin{array}{rcl} x & -2y & = 1 \\ 0 & y & = \frac{1}{2} \end{array}$$

Die eindeutige Lösung des Systems ist nun $x = 2, y = \frac{1}{2}$. Graphisch betrachtet liefert uns das System zwei Gerade mit den Gleichungen

$$\begin{array}{l} y = -1 + \frac{1}{2}x \\ y = -\frac{5}{2} + \frac{3}{2}x \end{array}$$

Die zwei Geraden schneiden sich in genau einem Punkt. Die Lösung ist eindeutig.

Wir werden nun Gleichungssysteme lösen. Dazu führen wir zunächst die erweiterte Matrix des linearen Gleichungssystems ein

Definition 6.3 (Erweiterte Matrix eines linearen Gleichungssystems). Gegeben sein ein lineares Gleichungssystem in Matrixform

$$A\mathbf{x} = \mathbf{b}$$

Die erweiterte Matrix dieses System entsteht, wenn der Vektor \mathbf{b} rechts an die Matrix A angefügt wird.

$$(A, \mathbf{b}) = A_b = \left[\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & \dots & a_{1m} & b_1 \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} & b_2 \\ a_{31} & a_{32} & a_{33} & \dots & a_{3m} & b_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} & b_n \end{array} \right]$$

Wir können nun die elementare Zeilenoperationen benutzen, um die erweiterte Matrix in Zeilenstufenform zu bringen.

Beispiel 6.3. Gegeben sei das lineare Gleichungssystem

$$A\mathbf{x} = \mathbf{b}$$

mit

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \\ 3 & 8 & 12 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad b = \begin{bmatrix} 4 \\ 5 \\ 20 \end{bmatrix}$$

Konstruieren sie die erweiterte Matrix des Systems, bringen sie diese in Zeilenstufenform und bestimmen sie die Lösung (wenn existent).

Die erweiterte Matrix des Systems ist

$$A_b = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 6 & 5 & 4 & 5 \\ 3 & 8 & 12 & 20 \end{array} \right]$$

Elementare Zeilenoperationen

1. Das (-6)-fache der ersten Zeile zur zweiten Zeile addieren

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 6 & 5 & 4 & 5 \\ 3 & 8 & 12 & 20 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -7 & -14 & -19 \\ 3 & 8 & 12 & 20 \end{array} \right]$$

2. Das (-3)-fache der ersten Zeile zur dritten Zeile addieren

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -7 & -14 & -19 \\ 3 & 8 & 12 & 20 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -7 & -14 & -19 \\ 0 & 2 & 3 & 8 \end{array} \right]$$

3. Die zweite Zeile mit $-\frac{1}{7}$ multiplizieren

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -7 & -14 & -19 \\ 0 & 2 & 3 & 8 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{19}{7} \\ 0 & 2 & 3 & 8 \end{array} \right]$$

4. Das (-2)-fache der zweiten Zeile zur dritten Zeile addieren

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{19}{7} \\ 0 & 2 & 3 & 8 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{19}{7} \\ 0 & 0 & -1 & \frac{18}{7} \end{array} \right]$$

5. Die dritte Zeile mit -1 multiplizieren

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{19}{7} \\ 0 & 0 & -1 & \frac{18}{7} \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & \frac{19}{7} \\ 0 & 0 & 1 & -\frac{18}{7} \end{array} \right]$$

6. Die Lösungen sind

$$x_3 = -\frac{18}{7} \quad x_2 = \frac{19}{7} - 2x_3 = \frac{55}{7} \quad x_1 = 4 - 2x_2 - 3x_3 = -4$$

Beispiel 6.4. Lösen sie die folgenden zwei Gleichungssysteme mittels Gauß'scher Elimination und Rückwärtseinsetzen.

1.

$$\begin{array}{rcl} x_1 & + & x_2 = 3 \\ 3x_1 & + & 5x_2 = 5 \end{array}$$

2.

$$\begin{array}{rcl} x_1 & + & 2x_2 + x_3 = 4 \\ x_1 & - & x_2 + x_3 = 5 \\ 2x_1 & + & 3x_2 - x_3 = 1 \end{array}$$

Bisher haben wir nur Systeme behandelt, die auch eine eindeutige Lösung besitzen. Wir wenden uns nun Gleichungssystemen zu, die entweder keine oder unendlich viele Lösungen haben. Gegeben sei das Gleichungssystem

$$A = \begin{array}{rclcrcl} & x_1 & +3x_2 & -x_3 & = & 4 \\ & 2x_1 & +x_2 & +x_3 & = & 7 \\ & 2x_1 & -4x_2 & +4x_3 & = & 6 \\ & 3x_1 & +4x_2 & & = & 11 \end{array}$$

Die erweiterte Matrix A_b ist gleich

$$A_b = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 2 & 1 & 1 & 7 \\ 2 & -4 & 4 & 6 \\ 3 & 4 & 0 & 11 \end{array} \right]$$

- Addition des -2 , -2 und -3 -fachen der ersten Zeile zur 2,3 und 4 Zeile ergibt

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & -5 & 3 & -1 \\ 2 & -4 & 4 & 6 \\ 3 & 4 & 0 & 11 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & -5 & 3 & -1 \\ 0 & -10 & 6 & -2 \\ 3 & 4 & 0 & 11 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & -5 & 3 & -1 \\ 0 & -10 & 6 & -2 \\ 0 & -5 & 3 & -1 \end{array} \right]$$

- Addition des -2 und -2 -fachen der zweiten Zeile zur 3. und 4. Zeile ergibt

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & -5 & 3 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & -5 & 3 & -1 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & -5 & 3 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

- Multiplikation der zweiten Zeile mit $-\frac{1}{5}$ ergibt

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & 1 & -\frac{3}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

- Die Lösung des Gleichungssystems ist

$$x_2 = \frac{1}{5} + \frac{3}{5}x_3 \quad x_1 = 4 - 3x_2 + x_3 = 4 - 3\left(\frac{1}{5} + \frac{3}{5}x_3\right) + x_3 = \frac{17}{5} - \frac{4}{5}x_3 \quad x_3 = \text{beliebig}$$

Man sieht, dass x_3 innerhalb der Lösung dieses Gleichungssystems einen beliebigen Wert annehmen kann. Das System hat unendlich viele Lösungen, die wir folgendermaßen darstellen.

$$(x_1, x_2, x_3) = \left(\frac{17}{5} - \frac{4}{5}t, \frac{1}{5} + \frac{3}{5}t, t\right)$$

wobei t eine beliebige reelle Zahl ist.

Elementare Zeilenoperationen erzeugen immer ein äquivalentes Gleichungssystem. In unserem Beispiel waren somit die folgenden zwei Gleichungssysteme äquivalent:

$$\begin{array}{rclcl} x_1 & +3x_2 & -x_3 & = & 4 \\ 2x_1 & +x_2 & +x_3 & = & 7 \\ 2x_1 & -4x_2 & +4x_3 & = & 6 \\ 3x_1 & +4x_2 & & = & 11 \end{array} \sim \begin{array}{rcl} x_1 & +\frac{4}{5}x_3 & = & \frac{17}{5} \\ x_2 & -\frac{3}{5}x_3 & = & \frac{1}{5} \end{array}$$

Wir konnten das LGS auf zwei Gleichungen reduzieren.

Wir führen nun die Begriffe der Basisvariablen und der freien Variablen ein. Im obigen LGS in Zeilenstufenform konnte man unterscheiden in

1. Zeilen, die als führenden Eintrag (= erster Eintrag, der nicht Null ist) eine 1 besitzen. Dieses Element wird Pivotelement genannt
2. Zeilen, ohne eine 1 als führenden Eintrag.

Definition 6.4 (Basisvariablen und Freie Variablen). *Gegeben sei eine Matrix in Zeilenstufenform. Als Basisvariablen werden die Variablen bezeichnet, die einer Spalte mit Pivotelement zugeordnet sind. Alle anderen Variablen sind freie Variablen. Im obigen Beispiel hatten wir die Zeilenstufenform*

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 4 \\ 0 & 1 & -\frac{3}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

Hier sind x_1 und x_2 die Basisvariablen und x_3 ist eine freie Variable

Beispiel 6.5. Bestimmen sie die Basisvariablen und die freien Variablen der folgenden Matrizen.

1.

$$\left[\begin{array}{ccccc|c} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 1 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{array} \right]$$

2.

$$\left[\begin{array}{ccccc|c} 1 & 6 & 0 & 0 & 4 & -2 \\ 0 & 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 2 & 10 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Natürlich kann es auch vorkommen, dass ein LGS keine Lösung hat. Wir behandeln diesen Fall jetzt und betrachten das LGS

$$\begin{aligned} x_1 - 2x_2 + x_3 + 2x_4 &= a \\ x_1 + x_2 - x_3 + x_4 &= b \\ x_1 + 7x_2 - 5x_3 - x_4 &= c \end{aligned}$$

Für welche Werte von a, b, c existiert für das LGS eine Lösung?

Das LGS kann durch elementare Zeilenoperationen umgeformt werden zu

$$\left[\begin{array}{cccc|c} 1 & -2 & 1 & 2 & a \\ 1 & 1 & -1 & 1 & b \\ 1 & 7 & -5 & -1 & c \end{array} \right] \sim \left[\begin{array}{cccc|c} 1 & -2 & 1 & 2 & a \\ 1 & 1 & -1 & 1 & b-a \\ 0 & 0 & 0 & 0 & 2a-3b+c \end{array} \right]$$

Dieses LGS hat unendlich viele Lösungen, wenn $2a - 3b + c = 0$ und keine Lösung, wenn $2a - 3b + c \neq 0$. Wir werden nun verschiedene LGS auf deren Lösbarkeit untersuchen und LGS anhand ihrer Anzahl von Zeilen und Spalten unterscheiden.

Wir führen nun ein wichtiges Konzept ein, dass uns hilft zu bestimmen, ob und wie viele Lösungen ein LGS hat.

Definition 6.5 (Rang einer Matrix). *Der Rang einer Matrix ist gleich der Anzahl der Zeilen der Matrix in Zeilenstufenform die nicht komplett gleich Null sind. Es sei A eine $m \times n$ Matrix und A_b die erweiterte Matrix von A . Dann gilt*

1. $\text{Rang } A \leq \text{Rang } A_b$
2. $\text{Rang } A \leq \text{Anzahl der Zeilen von } A$
3. $\text{Rang } A \leq \text{Anzahl der Spalten von } A$

Beispiel 6.6. Bestimmen sie den Rang der folgenden Matrizen.

1.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

2.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

Wir können nun den Rang einer Matrix mit dem Rang der erweiterten Matrix vergleichen um festzustellen, ob das LGS Lösungen hat.

Theorem 6.1. Gegeben sei das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$. Das LGS hat eine Lösung (nicht unbedingt eindeutig), wenn und nur wenn

$$\text{Rang}A = \text{Rang}A_{\mathbf{b}}$$

Beispiel 6.7. Bestimmen sie, ob die folgenden LGS Lösungen haben.

1.

$$\left[\begin{array}{cccc|c} 1 & -2 & 1 & 2 & 3 \\ 1 & 1 & -1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 9 \end{array} \right]$$

2.

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 3 \\ 1 & 1 & -1 & 4 \\ 3 & 3 & -3 & 12 \\ 6 & 4 & 3 & -5 \\ 9 & -1 & 4 & 10 \\ 5 & 0 & 5 & 18 \end{array} \right]$$

unter der Annahme, dass $\text{Rang} A_{\mathbf{b}} = 4$

Definition 6.6. Ein lineares LGS $A\mathbf{x} = \mathbf{0}$ wird homogenes LGS genannt.

Zusammenfassend können wir feststellen, dass

Definition 6.7. Gegeben sei das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$, wobei A eine $n \times m$ Matrix ist.

1. Wenn $n < m$ (unterbestimmt), dann

- hat $A\mathbf{x} = \mathbf{0}$ unendlich viele Lösungen.
- hat $A\mathbf{x} = \mathbf{b}$ keine oder unendlich viele Lösungen

2. Wenn $n > m$ (überbestimmt)

- hat $A\mathbf{x} = \mathbf{0}$ eine oder unendlich viele Lösungen.
- hat $A\mathbf{x} = \mathbf{b}$ keine, eine oder unendlich viele Lösungen

3. Wenn $n = m$

- hat $A\mathbf{x} = \mathbf{0}$ eine oder unendlich viele Lösungen.

- hat $A\mathbf{x} = \mathbf{b}$ keine, eine oder unendlich viele Lösungen

Beispiel 6.8. $n < m$.

Das LGS hat mehr Unbekannte als Gleichungen. Wir wissen das $\text{Rang}(A) \leq \text{Rang}(A_b)$. Wir unterscheiden zwischen vollem $\text{Rang}(A) = n$ und $\text{Rang}(A) < m$

1. $\text{Rang}(A) = n$. In diesem Fall gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es muss also mindestens eine Lösung existieren.

- $A\mathbf{x} = \mathbf{0}$. Eine eindeutige Lösung scheidet aus, da immer freie Variablen existierten.
- $A\mathbf{x} = \mathbf{b}$. Eine eindeutige Lösung scheidet aus, da immer freie Variablen existierten.

2. $\text{Rang}(A) < n$.

- $A\mathbf{x} = \mathbf{0}$. In diesem Fall gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es muss also mindestens eine Lösung existieren. Eine eindeutige Lösung scheidet aus, da immer freie Variablen existierten.
- $A\mathbf{x} = \mathbf{b}$. Wenn $\text{Rang}(A) < \text{Rang}(A_b)$ existiert keine Lösung. Wenn $\text{Rang}(A) = \text{Rang}(A_b)$ existieren unendlich viele Lösungen, weil es immer freie Variablen gibt.

Also können wir bei einem linearen Gleichungssystem mit weniger Gleichungen als Unbekannten eine eindeutige Lösung ausschließen!

Beispiel 6.9. $n > m$.

Das LGS hat mehr Gleichungen als Unbekannte. Wir wissen das $\text{Rang}(A) \leq \text{Rang}(A_b)$. Wir unterscheiden zwischen vollem $\text{Rang}(A) = m$ und $\text{Rang}(A) < m$

1. $\text{Rang}(A) = m$. Es gibt keine freien Variablen. Es kann also eine oder keine Lösung existieren.

- $A\mathbf{x} = \mathbf{0}$. Es gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es existiert eine Lösung. $\mathbf{x} = \mathbf{0}$ ist eine Lösung und somit die einzige.
- $A\mathbf{x} = \mathbf{b}$. Wenn $\text{Rang}(A) = \text{Rang}(A_b)$ existiert genau eine Lösung, wenn $\text{Rang}(A) < \text{Rang}(A_b)$ existiert keine.

2. $\text{Rang}(A) < m$.

- $A\mathbf{x} = \mathbf{0}$. In diesem Fall gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es muss also mindestens eine Lösung existieren. Eine eindeutige Lösung scheidet aus, da immer freie Variablen existierten. Es existieren unendlich viele Lösungen.
- $A\mathbf{x} = \mathbf{b}$. Wenn $\text{Rang}(A) < \text{Rang}(A_b)$ existiert keine Lösung. Wenn $\text{Rang}(A) = \text{Rang}(A_b)$ existieren unendlich viele Lösungen, weil es immer freie Variablen gibt.

Beispiel 6.10. $n = m$.

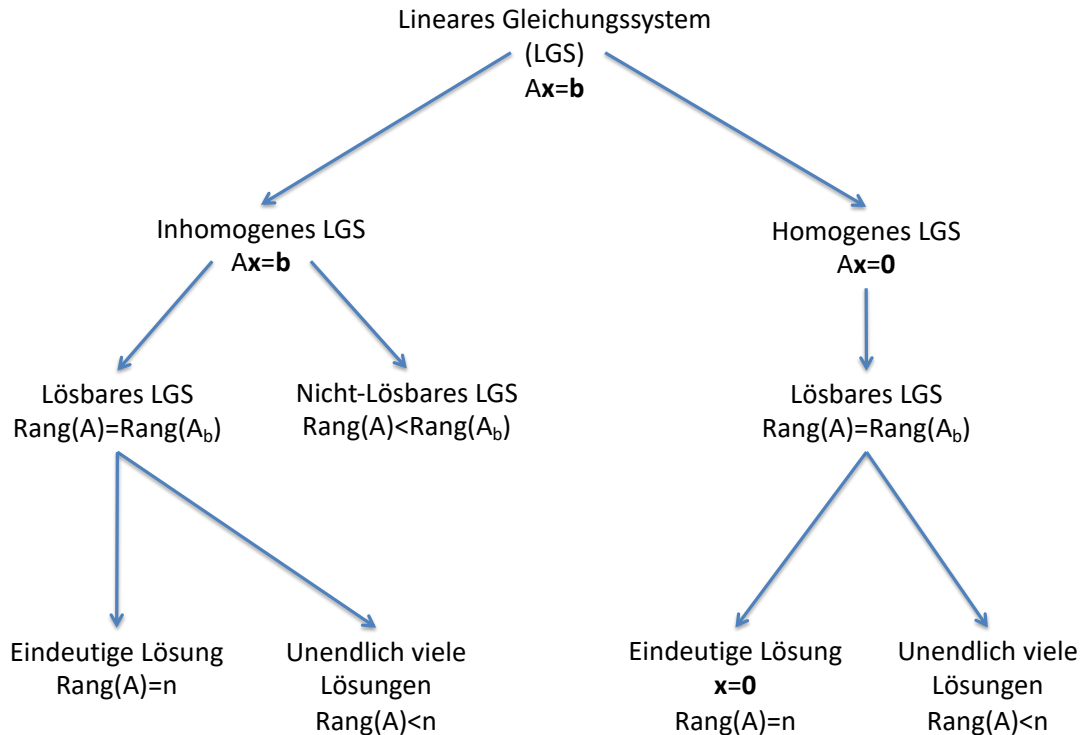
Das LGS hat die gleiche Anzahl von Gleichungen und Unbekannten. Wir wissen das $\text{Rang}(A) \leq \text{Rang}(A_b)$. Wir unterscheiden zwischen vollem $\text{Rang}(A) = m = n$ und $\text{Rang}(A) < m, n$

1. $\text{Rang}(A) = n = m$. Es gibt keine freien Variablen. Es kann also eine oder keine Lösung existieren.

- $A\mathbf{x} = \mathbf{0}$. Es gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es existiert eine Lösung. $\mathbf{x} = \mathbf{0}$ ist eine Lösung und somit die einzige.
- $A\mathbf{x} = \mathbf{b}$. Da gelten muss $\text{Rang}(A) = \text{Rang}(A_b)$ existiert eine eindeutige Lösung, da es in diesem Fall keine freien Variablen gibt.

2. $\text{Rang}(A) < m$.

- $A\mathbf{x} = \mathbf{0}$. In diesem Fall gilt $\text{Rang}(A) = \text{Rang}(A_b)$. Es muss also mindestens eine Lösung existieren. Eine eindeutige Lösung scheidet aus, da immer freie Variablen existieren. Es existieren unendlich viele Lösungen.
- $A\mathbf{x} = \mathbf{b}$. Wenn $\text{Rang}(A) < \text{Rang}(A_b)$ existiert keine Lösung. Wenn $\text{Rang}(A) = \text{Rang}(A_b)$ existieren unendlich viele Lösungen, weil es immer freie Variablen gibt.



6.2 Die Inverse

Definition 6.8. Eine quadratische $n \times n$ Matrix A heißt invertierbar, wenn eine quadratische $n \times n$ Matrix A^{-1} existiert für die gilt

$$A^{-1}A = AA^{-1} = I_n$$

wobei I_n eine $n \times n$ Identitätsmatrix ist.

Theorem 6.2. Eine quadratische $n \times n$ Matrix A ist invertierbar, wenn

1. man sie durch elementare Zeilenoperationen in eine Einheitsmatrix umwandeln kann.
2. sie den maximalen Rang $= n$ hat.
3. die Determinante von A ungleich Null ist.

Wenn eine quadratische Matrix eine Inverse hat kann man das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$ für jeden Vektor \mathbf{b} wie folgt lösen

$$\begin{aligned}
A\mathbf{x} &= \mathbf{b} \\
A^{-1}A\mathbf{x} &= A^{-1}\mathbf{b} \\
I_n\mathbf{x} &= A^{-1}\mathbf{b} \\
\mathbf{x} &= A^{-1}\mathbf{b}
\end{aligned}$$

6.3 Lineare Unabhängigkeit

Definition 6.9 (Linearkombination). Gegeben seien die Vektoren $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Dann ist der Vektor

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = \sum_i^n a_i\mathbf{v}_i$$

mit den Konstanten a_1, a_2, \dots, a_n eine Linearkombination von $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

Definition 6.10 (Lineare Unabhängigkeit). Gegeben seien die Vektoren $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Wenn

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = \mathbf{0}$$

impliziert, dass $a_1 = a_2 = \dots = a_n = 0$, dann sind die Vektoren **linear unabhängig**. Ansonsten sind die Vektoren linear abhängig.

Definition 6.11 (Lineares Gleichungssystem als Linearkombination von Vektoren). Gegeben sei das lineare Gleichungssystem $A\mathbf{x} = \mathbf{b}$. Es sei \mathbf{a}_i die i^{te} Spalte der Matrix A . Dann können wir dieses System auch schreiben als

$$A\mathbf{x} = \mathbf{b} \iff [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \mathbf{b}$$

Man versucht bei der Lösung des LGS also Werte für x_1, x_2, \dots, x_n zu finden, so dass \mathbf{b} als Linearkombination der Spalten von A geschrieben werden kann!

Definition 6.12 (Prüfung auf Lineare Unabhängigkeit). Gegeben seien die $m \times 1$ -Vektoren $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Die Vektoren sind linear unabhängig, wenn das homogene Gleichungssystem

$$V\mathbf{x} = \mathbf{0}$$

mit $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]$ nur die eindeutige Lösung $\mathbf{x} = \mathbf{0}$ hat.

Notiz 6.1. Die Vektoren $m \times 1$ -Vektoren $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ können also nur linear unabhängig sein, wenn das System mindestens so viele Gleichungen wie Unbekannte besitzt.

6.4 Spanne von Vektoren

In diesem Abschnitt beschäftigen wir uns damit welchen Raum Vektoren „aufspannen“. Hiermit ist gemeint welche Menge an Punkten die Linearkombination einer Gruppe von Vektoren reproduzieren können.

Beispiel 6.11. Gegeben seien die Vektoren $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ und $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Es ist ersichtlich, dass die Linearkombination

$$a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$$

alle Vektoren in \mathbb{R}^2 reproduzieren können. Z.B. ist

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\mathbf{e}_1 + 3\mathbf{e}_2$$

Die Vektoren spannen \mathbb{R}^2 .

Theorem 6.3 (Spanne einer Gruppe von Vektoren). Um \mathbb{R}^n zu spannen benötigt man mindestens n linear unabhängige Vektoren mit n Elementen.

6.5 Lineare Projektion

Wir wenden nun unser Wissen auf den Fall eines linearen Gleichungssystems $A\mathbf{x} = \mathbf{b}$ an, das mehr Gleichungen als Variablen besitzt. Solche linearen Gleichungssysteme werden sie in der Ökonometrie behandeln. Wir wissen, dass \mathbf{b} ein $m \times 1$ -Vektor ist, also im m -dimensionalen Raum lebt. Wir wissen auch, dass dieses GS in Regelfall keine Lösung hat. Wir können nun intuitiv betrachten warum dies der Fall ist.

Die Matrix A besitzt $n < m$ Vektoren und kann somit unmöglich den Raum \mathbb{R}^m spannen, in dem der Vektor \mathbf{b} existiert!

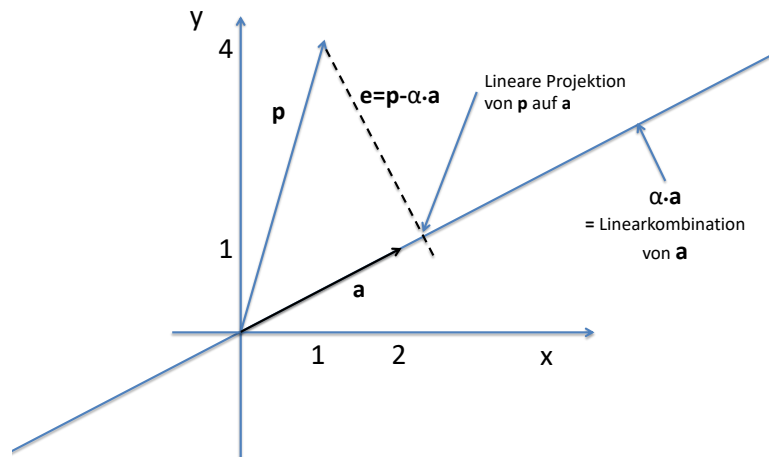
Was kann man in diesem Fall tun? Nehmen wir an, dass die Matrix A vollen Rang hat. Wir finden dann die Linearkombination der Spalten von A , die am nächsten am Vektor \mathbf{b} dran ist. Wir verdeutlichen diese Vorgehensweise mit einem Beispiel. Wir wollen den Vektor

$$\mathbf{p} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

als Linearkombination des Vektors

$$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

darstellen. Da zwei linear unabhängige Vektoren benötigt werden um \mathbb{R}^2 zu spannen, ist dies unmöglich.



Wie finden wir α ? Wir sehen, dass am Punkt, an dem \mathbf{p} dem Vektor $\alpha \mathbf{a}$ am nächsten ist, $\mathbf{e} = \mathbf{p} - \alpha \mathbf{a}$ und \mathbf{a} orthogonal sind. Wir können also rechnen

$$\begin{aligned} \mathbf{a}'(\mathbf{p} - \alpha \mathbf{a}) &= 0 \\ \mathbf{a}'\mathbf{p} - \alpha \mathbf{a}'\mathbf{a} &= 0 \\ \mathbf{a}'\mathbf{p} &= \alpha \mathbf{a}'\mathbf{a} \\ \alpha &= \frac{\mathbf{a}'\mathbf{p}}{\mathbf{a}'\mathbf{a}} \end{aligned}$$

$\alpha \mathbf{a}$ ist dann die lineare Projektion von \mathbf{p} auf \mathbf{a} . In unserem Beispiel ist also

$$\alpha = \frac{\mathbf{a}'\mathbf{p}}{\mathbf{a}'\mathbf{a}} = \frac{6}{5} = 1,2 \quad \text{und} \quad \alpha \mathbf{a} = \begin{bmatrix} 2,4 \\ 1,2 \end{bmatrix}$$

Höhere Dimensionen Abschließend behandeln wir ein Beispiel aus der VWL. Wir wollen die Löhne von Arbeitnehmern mittels der Variablen Bildung (b) und Berufserfahrung (e) erklären. Die lineare Gleichung ist

$$w_i = \beta_1 b_i + \beta_2 e_i$$

Wir nehmen an, dass wir Daten für 1000 Arbeitnehmer haben. Das LGS ist

$$X\boldsymbol{\beta} = \mathbf{w}$$

, wobei

$$X = [\mathbf{b} \quad \mathbf{e}] \quad \boldsymbol{\beta}' = [\beta_1 \quad \beta_2]$$

Unter der Annahme, dass \mathbf{b} und \mathbf{e} linear unabhängig sind ist $\text{Rang}(X) = 2$ und \mathbf{w} ist ein (1000×1) -Vektor. Eine eindeutige Lösung ist also mehr als unwahrscheinlich. Wir können aber \mathbf{w} auf $X\boldsymbol{\beta}$ linear projizieren. Wie im letzten Beispiel suchen wir den Vektor $\hat{\boldsymbol{\beta}}$, an dem $\mathbf{w} - X\hat{\boldsymbol{\beta}}$ und X orthogonal sind

$$\begin{aligned} X'(\mathbf{w} - X\hat{\boldsymbol{\beta}}) &= 0 \\ X'\mathbf{w} - X'X\hat{\boldsymbol{\beta}} &= 0 \\ X'\mathbf{w} &= X'X\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}} &= (X'X)^{-1}X'\mathbf{w} \end{aligned}$$

Somit ist $X\hat{\beta}$ die lineare Projektion des 1000-dimensionalen Vektors \mathbf{w} in die 2-dimensionale Spanne von X

6.6 Eigenwerte und Eigenvektoren

Matrizen verbringen einen Großteil der Zeit damit Vektoren zu multiplizieren. Betrachten wir zum Beispiel das Gleichungssystem

$$A\mathbf{x} = \mathbf{b}$$

Wenn man eine Matrix mit einem beliebigen Vektor multipliziert, verändert sich in der Regel die Richtung des Vektors (\mathbf{x} zeigt in eine andere Richtung als \mathbf{b}). Jetzt fragt man sich natürlich, ob es für jede quadratische $n \times n$ Matrix A Vektoren \mathbf{x} (mit $\mathbf{x} \neq 0$) und einen Skalar $\lambda \in \mathbb{R}$ gibt, so dass

$$A\mathbf{x} = \lambda\mathbf{x}$$

Man sieht in diesem Fall, dass die Matrix A den Vektor \mathbf{x} nur streckt, staucht, umkehrt oder gleich Null setzt, jedoch nicht seine Richtung verändert. Vektoren mit dieser Eigenschaft werden Eigenvektoren genannt. Welche der vier Möglichkeiten der Fall ist, liegt an den mit den Eigenvektoren assoziierten Eigenwerten - den λ s.

Die Eigenwerte einer Matrix können verwendet werden

1. Feststellung der Definitheit einer Matrix
2. Stabilität dynamischer Systeme
3. Auswahl von Variablen in statistischen Analysen
4. und andere

Wir stellen nun vor wie man die Eigenwerte und Eigenvektoren einer Matrix bestimmt. Als Basis haben wir die Gleichung

$$A\mathbf{x} = \lambda\mathbf{x} \implies A\mathbf{x} = \lambda I_n \mathbf{x} \implies (A - \lambda I)\mathbf{x} = 0$$

wobei I_n eine $n \times n$ Einheitsmatrix ist. Da $\mathbf{x} \neq \mathbf{0}$ hat das System ganz rechts nur Lösungen, wenn

$$\det(A - \lambda I) = 0$$

Die Determinante liefert das Charakteristische Polynom, dessen Lösung die Eigenwerte sind.

Definition 6.13. Als Charakteristisches Polynom einer $n \times n$ Matrix A , wird bezeichnet

$$P_A(\lambda) = \det(A - \lambda I)$$

wobei I eine $n \times n$ Einheitsmatrix ist. Wird $P_A(\lambda) = 0$ gelöst, werden die gefundenen Werte für die λ s Eigenwerte genannt.

$$P_A(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} c_1 \lambda^{n-1} + (-1)^{n-2} c_2 \lambda^{n-2} + \dots + c_n = 0$$

Hierbei sind die c_i sich aus der Determinantenbildung ergebende Koeffizienten. Hier ist erwähnenswert, dass c_1 immer die Spur von A (=Summe der Elemente auf der Diagonalen) und c_n immer die Determinante sind. Als Beispiel zeigen wir hier die Determinante einer 2×2 -Matrix.

Charakteristisches Polynom einer 2×2 Matrix

$$\lambda^2 - \text{Spur}(A)\lambda + \det(A) = 0$$

Theorem 6.4. Es sei A eine symmetrische Matrix reeller Zahlen, d.h. $a_{i,j} \in \mathbb{R}$. Dann sind auch die Eigenwerte reelle Zahlen.

Wir können die Eigenwerte einer Matrix bestimmen, indem wir die Wurzeln des charakteristischen Polynoms finden. Der Rechenaufwand bei 3×3 und 4×4 Matrizen ist groß, danach gibt es nur noch numerische Lösungen, so dass man die Eigenwerte (und Eigenvektoren) von größeren Matrizen mit dem Computer bestimmt.

Wenn man die Eigenwerte λ_i bestimmt hat, kann man sie benutzen, um die dazu gehörigen Eigenvektoren zu bestimmen. Jeder Eigenwert liefert einen Eigenvektor. Man setzt die Eigenwerte in das homogene Gleichungssystem

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

ein und löst für \mathbf{x} .

Beispiel 6.12. Bestimmen sie die Eigenwerte und Eigenvektoren der Matrix

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Wir können dann die Eigenwerte bestimmen, welche die Determinante von $(A - \lambda I)$ gleich Null setzen.

$$\det \begin{bmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{bmatrix} = (2 - \lambda)(2 - \lambda) - 1 = 0$$

Wir lösen $\lambda^2 - 4\lambda + 3 = 0$ für

$$\lambda_1 = 1 \quad \lambda_2 = 3$$

Wir sehen, dass

$$\det(A) = \lambda_1 \cdot \lambda_2 \quad \text{und} \quad \text{tr}(A) = \lambda_1 + \lambda_2$$

Also gilt hier

1. Die Summe der Eigenwerte ist gleich der Spur von A
2. Das Produkt der Eigenwerte ist gleich der Determinante von A

Für die Bestimmung der Eigenvektoren setzen wir die Eigenwerte in $(A - \lambda I)\mathbf{x} = \mathbf{0}$ ein

1. (für $\lambda = 3$)

$$\begin{bmatrix} 2 - 3 & -1 \\ -1 & 2 - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0} \iff \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}$$

Wir multiplizieren die erste Zeile mit (-1) und addieren Sie zur zweiten Zeile und erhalten

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{0}$$

Die erste Zeile des Systems ist jetzt gleich

$$x_1 + x_2 = 0$$

Dies impliziert den folgenden Eigenvektor

$$\mathbf{x}_1 = t \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

2. für $\lambda = 1$ erhalten wir den Eigenvektor

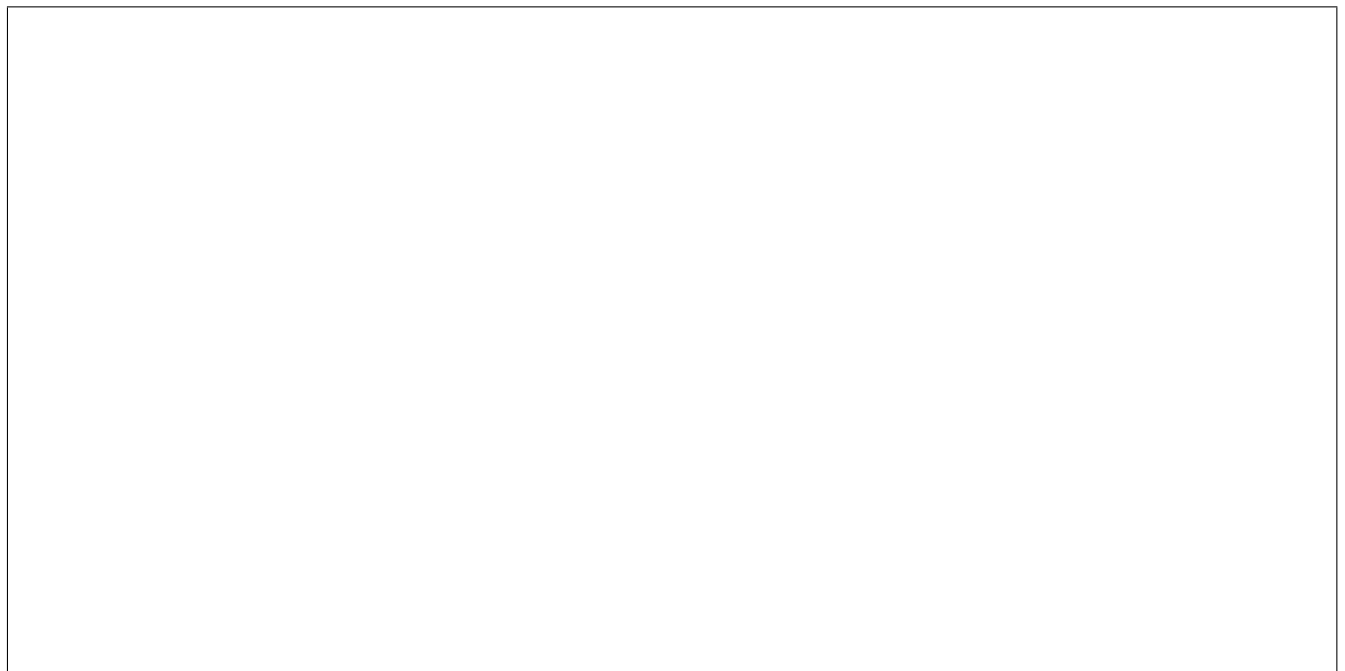
$$\mathbf{x}_2 = t \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Wir sehen, dass die Eigenvektoren einer Matrix nicht eindeutig bestimmt sind.

Beispiel 6.13. Bestimmen Sie die Eigenwerte und Eigenvektoren der folgenden Matrizen

1. $\begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$



Eigenwerte und die Definitheit einer Matrix Eine quadratische, symmetrische Matrix reeller Zahlen ist

1. positiv definit, wenn $\lambda_i > 0 \quad \forall i$
2. positiv semidefinit, wenn $\lambda_i \geq 0 \quad \forall i$
3. negativ definit, wenn $\lambda_i < 0 \quad \forall i$
4. negativ semidefinit, wenn $\lambda_i \leq 0 \quad \forall i$
5. indefinit, wenn positive und negative Eigenwerte existieren

Zum Schluss noch ein (nicht-klausurrelevantes) Beispiel, das die Wichtigkeit von Eigenwerten in dynamischen Systemen darstellt.

Beispiel 6.14 (Dynamisches System). Gegeben sei das dynamische System

$$\mathbf{x}_1 = A\mathbf{x}_0$$

Wir können dieses System fortsetzen als

$$\mathbf{x}_2 = A\mathbf{x}_1 = AA\mathbf{x}_0 = A^2\mathbf{x}_0$$

und allgemein schreiben als

$$\mathbf{x}_t = A^t\mathbf{x}_0$$

Unter gewissen Voraussetzungen lässt sich die Matrix A schreiben als

$$A = P\Lambda P^{-1} \implies A^t = P\Lambda^t P^{-1}$$

wobei die Spalten von P gleich den Eigenvektoren sind und die diagonale Matrix Λ die Eigenwerte auf der Diagonalen hat. Nun ist (für eine 3×3 Matrix)

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & 0 & 0 \\ 0 & \lambda_2^t & 0 \\ 0 & 0 & \lambda_3^t \end{bmatrix}$$

Wir sehen, dass das langfristige Verhalten des dynamischen Systems von den Beträgen der Eigenwerte abhängt. Wenn alle Eigenwerte im Betrag kleiner als 1 sind, ist das System langfristig stabil und konvergiert gegen den Nullvektor. Wenn die Eigenwerte im Betrag größer als 1 sind, "explodiert" der Vektor \mathbf{x}_t langfristig.

7 Integralrechnung

7.1 Unbestimmte Integrale

Gegeben sei eine Funktion $f(x)$, die die erste Ableitung einer Funktion $F(x)$ ist, also $F'(x) = f(x)$. In diesem Abschnitt der Vorlesung stellen wir die Frage

Ist es möglich von $f(x)$ auf $F(x)$ zu schließen?

Die Funktion $F(x)$ wird als *Stammfunktion* der Ableitung $f(x)$ bezeichnet. Wir beginnen mit einem Beispiel.

Beispiel 7.1. Bestimmen sie die Stammfunktion von $f(x) = x^2$.

Wir rechnen zurück: Wir wissen, dass die Ableitung von $F(x) = x^n$ gleich $F'(x) = nx^{n-1}$ ist. Für unser Beispiel gilt $n - 1 = 2 \implies n = 3$. Da die Ableitung aber auch noch einen Koeffizienten gleich 1 hat, müssen wir die Stammfunktion noch mit $\frac{1}{n}$ multiplizieren. Man sieht nun

$$F(x) = \frac{1}{3}x^3 \implies F'(x) = x^2$$

Wir sind aber noch nicht ganz fertig. Wir sehen, dass alle folgenden Funktionen die Ableitung x^2 haben.

$$y = \frac{1}{3}x^3 + 5 \quad y = \frac{1}{3}x^3 - 5 \quad y = \frac{1}{3}x^3 \quad y = \frac{1}{3}x^3 - 4$$

Wir müssen deshalb an unsere gefundene Stammfunktion noch eine (unbestimmte) Konstante anfügen, die beim Ableiten verloren ging. Am Ende ist das unbestimmte Integral von $f(x) = x^2$

$$F(x) = \frac{1}{3}x^3 + C$$

Dieses gerade demonstrierte *Gegenableiten* ist Gegenstand der Integralrechnung.

Definition 7.1 (Das Unbestimmte Integral). Gegeben seien $F(x)$ und $f(x)$ mit $F'(x) = f(x)$ Dann schreiben wir

$$\int f(x)dx = F(x) + C$$

wobei C eine beliebige Konstante ist und dx impliziert, dass x die Integrationsvariable ist. \int ist das Integralzeichen, $f(x)$ ist der Integrand und $F(x) + C$ das unbestimmte Integral von $f(x)$. Wir lesen

Das unbestimmte Integral von $f(x)$ bezüglich x ist $F(x)$ plus eine Konstante. Ferner sehen wir, dass

- $\frac{d}{dx} \int f(x)dx = f(x)$, die Ableitung eines unbestimmten Integrals ist gleich dem Integranden ist.
- $\int F'(x)dx = F(x) + C$, Integration und Ableitung heben sich gegenseitig auf.

Wir formulieren nun unsere erste Integrationsregel

Definition 7.2 (Integrationsregeln I). Für $n \neq -1$ gilt

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + C$$

Beispiel 7.2. *Bestimmen sie die folgenden unbestimmten Integrale*

1. $\int 5dx$

3. $\int x^3 dx$

2. $\int x dx$

4. $\int 5x^4 dx$



Die obige Integrationsregel gilt nicht für $n = -1$ und es gibt weitere Spezialfälle, die die folgenden Integrationsregeln nach sich ziehen

Definition 7.3. *Integrationsregeln II:*

1. $\int \frac{1}{x} dx = \ln|x| + C$

2. $\int e^{ax} dx = \frac{1}{a} e^{ax} + C$

3. $\int a^x dx = \frac{1}{\ln a} a^x + C$ für $a > 0$ und $a \neq 1$

Integrale können wie folgt umgeformt werden:

1. $\int af(x) dx = a \int f(x) dx$ wobei $a \neq 0$ eine Konstante ist.

2. $\int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx$

Beispiel 7.3. *Bestimmen sie die folgenden unbestimmten Integrale*

1. $\int \frac{3}{x} dx$

3. $\int e^x dx$

2. $\int (3x^4 + 5x^2 + 2) dx$

4. $\int \frac{x^3 - 3x + 4}{x} dx$

Nachdem wir nun viele unbestimmte Integrale bestimmt haben, kann man leicht in den Glauben verfallen, dass man für jede Funktion ein unbestimmtes Integral finden kann. Dem ist nicht so. Ein Paradebeispiel ist

$$\int e^{x^2} dx$$

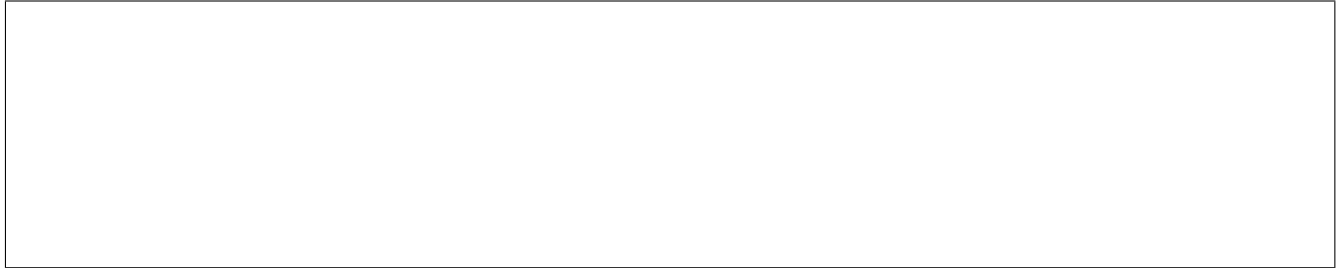
Dieses unbestimmte Integral existiert nicht und es ist bei vielen Funktion sehr schwierig das unbestimmte Integral zu bestimmen. Es ist jedoch leicht zu überprüfen, ob ein Ausdruck das unbestimmte Integral einer Funktion ist. Man braucht den Ausdruck nur abzuleiten. Beispielsweise kann überprüft werden, ob

$$\int \ln x dx = x \ln x - x + C$$

Man braucht das vorgeschlagene unbestimmte Integral nur zu differenzieren und bekommt die Bestätigung!

Beispiel 7.4. Überprüfen sie die Korrektheit des folgenden unbestimmten Integrals

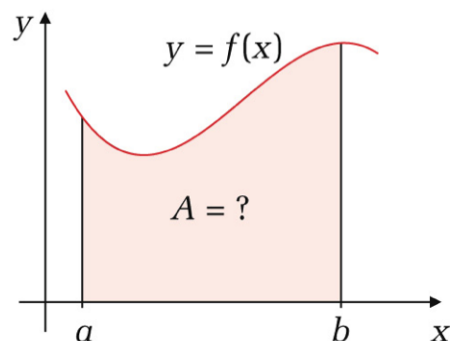
$$\int (2xe^x + x^2e^x) dx = x^2e^x + C$$



7.2 Bestimmte Integrale

Wir wenden uns nun dem Problem zu Flächen zu berechnen. Gegeben sei zum Beispiel die nicht-negative, stetige Funktion $f(x)$

Wie berechnen wir die im folgenden Diagramm mit A gekennzeichnete Fläche unter der Funktion im Intervall $[a, b]$?



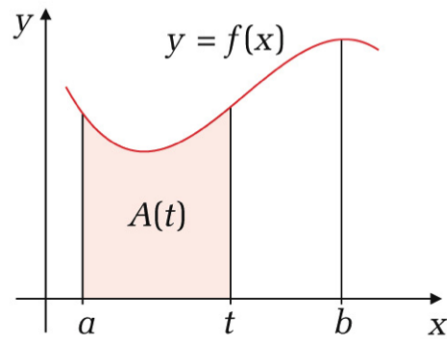
Wir beginnen mit einer Fläche, die etwas kleiner als die gesuchte ist. Wir bezeichnen mit $A(t)$ die Fläche unter der Funktion im Intervall $[a, t]$, die in der nächsten Graphik gezeigt wird.

Wie wir sehen nimmt die Fläche $A(t)$ mit t zu. Es sei $\Delta t > 0$ und wir können auch die Fläche $A(t + \Delta t)$ und die Differenz $\Delta A = A(t + \Delta t) - A(t)$ definieren

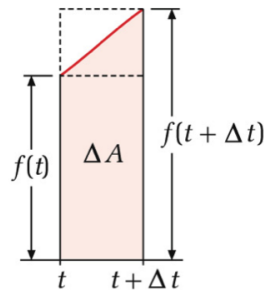
$$A(t) = \text{Fläche unter der Funktion } f(x) \text{ im Intervall } [a, t]$$

$$A(t + \Delta t) = \text{Fläche unter der Funktion } f(x) \text{ im Intervall } [a, t + \Delta t]$$

$$\Delta A = A(t + \Delta t) - A(t) = \text{Fläche unter der Funktion } f(x) \text{ im Intervall } [t, t + \Delta t]$$



Die folgende Graphik zeigt, dass die Fläche ΔA zwischen den beiden Flächen $\Delta t \cdot f(t)$ und $\Delta t \cdot f(t + \Delta t)$ liegen muss



$$\Delta t \cdot f(t) \leq A(t + \Delta t) - A(t) \leq \Delta t \cdot f(t + \Delta t)$$

Da $\Delta t > 0$ folgt

$$f(t) \leq \frac{A(t + \Delta t) - A(t)}{\Delta t} \leq f(t + \Delta t)$$

Wenn wir nun den Grenzwert dieses Ausdrucks betrachten wenn Δt gegen Null geht. Wir bemerken folgendes

- $\lim_{\Delta t \rightarrow 0} f(t + \Delta t) = f(t)$, weil $f(t)$ stetig ist. (d.h. Grenzwert existiert und ist gleich Funktionswert an der Stelle)
- $\lim_{\Delta t \rightarrow 0} \frac{A(t + \Delta t) - A(t)}{\Delta t} = f(t)$, die Funktion $A(t)$ ist differenzierbar
- $A'(t) = f(t)$
- Der gesuchte Flächeninhalt ist ein unbestimmtes Integral von $f(t)$

Um nun den Flächeninhalt im Intervall $[a, b]$ zu berechnen müssen wir ein unbestimmtes Integral von $f(t)$ verwenden

$$A(t) = F(t) + C$$

Zusammen mit $A(a) = 0$ liefert dies

$$\begin{aligned} A(a) &= F(a) + C = 0 \implies C = -F(a) \\ A(b) &= F(b) + C \end{aligned}$$

Zusammengefasst ergibt sich somit die gesuchte Fläche als

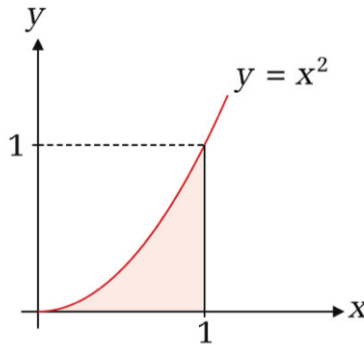
$$A = F(b) - F(a)$$

Beispiel 7.5. Bestimmen sie die Fläche unter der Funktion $f(x) = x^2$ im Intervall $[0, 1]$. Das unbestimmte Integral ist

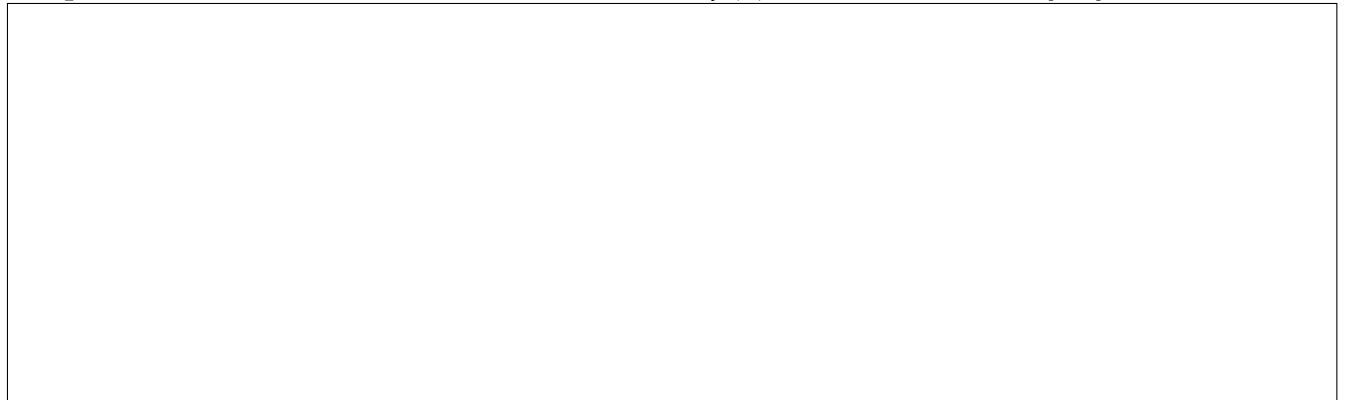
$$\int x^2 dx = \frac{1}{3}x^3 + C$$

und die gesuchte Fläche ist gleich

$$F(1) - F(0) = \frac{1}{3}(1)^3 - \frac{1}{3}(0)^3 = \frac{1}{3}$$



Beispiel 7.6. Bestimmen sie die Fläche unter der Funktion $f(x) = 3 + 5x$ im Intervall $[3, 4]$.



Definition 7.4 (Das bestimmte Integral). Wir nehmen an, dass die Funktion $F(x)$ auf dem Intervall $[a, b]$ stetig ist und die Ableitung $F'(x) = f(x)$ für jedes x in (a, b) hat. Dann heißt die Differenz

$$F(b) - F(a)$$

das bestimmte Integral von f über $[a, b]$. Wir schreiben

$$\int_a^b f(x) dx = \Big|_a^b F(x) = [F(x)]_a^b = F(b) - F(a)$$

Beispiel 7.7. Berechnen sie das bestimmte Integral $\int_2^5 e^{2x} dx$

Wir bilden das unbestimmte Integral

$$\int e^{2x} dx = \frac{1}{2}e^{2x} + C$$

und berechnen dann das bestimmte Integral

$$\int_2^5 F(x) = F(5) - F(2) = \frac{1}{2}e^{10} - \frac{1}{2}e^4 = \frac{1}{2}e^4(e^6 - 1)$$

Beispiel 7.8. Berechnen sie die folgenden bestimmten Integrale

1. $\int_0^1 x^3 dx$

2. $\int_1^2 (x^2 + 2x) dx$



Definition 7.5. Eigenschaften bestimmter Integrale:

1. $\int_a^b f(x) dx = - \int_b^a f(x) dx$

2. $\int_a^a f(x) dx = 0$

3. $\int_a^b \alpha f(x) dx = \alpha \int_a^b f(x) dx$ für $\alpha \in \mathbb{R}$

4. $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$

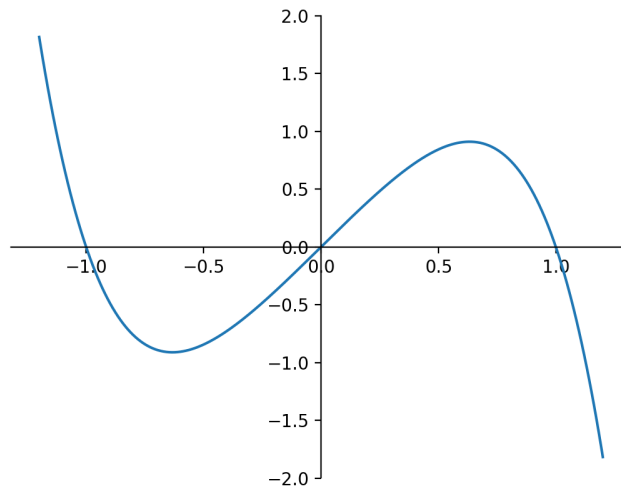
Sind wir nun auch in der Lage die Fläche zwischen einer Funktion und der x-Achse zu berechnen, die nicht nur positive Werte annimmt? Das folgende Beispiel zeigt, dass die bisher verwendete Methode nicht zum Ziel führt

Beispiel 7.9. Berechnen sie $\int_{-1}^1 (2x - x^3 - x^5)$.

$$\int_{-1}^1 (x^2 - \frac{1}{4}x^4 - \frac{1}{6}x^6) = \left(1 - \frac{1}{4} - \frac{1}{6}\right) - \left(1 - \frac{1}{4} - \frac{1}{6}\right) = 0$$

Der Graph der Funktion zeigt jedoch, dass die Fläche zwischen Funktion und x-Achse positiv ist.

Das bestimmte Integral ist gleich Null, weil Flächen unterhalb der x-Achse ein negatives Vorzeichen haben und weil der Graph der Funktion symmetrisch ist. Wir können Abhilfe schaffen in dem wir das Integral unterteilen und Flächen unterhalb der x-Achse mit (-1) multiplizieren. Dazu müssen wir die Nullstellen der Funktion finden, die in diesem Fall $-1, 1$ und 0 sind.



$$\begin{aligned}
 \int_{-1}^1 (2x - x^3 - x^5) &= - \int_{-1}^0 (2x - x^3 - x^5) + \int_0^1 (2x - x^3 - x^5) \\
 &= - \left|_{-1}^0 \left(x^2 - \frac{1}{4}x^4 - \frac{1}{6}x^6 \right) \right. + \left|_0^1 \left(x^2 - \frac{1}{4}x^4 - \frac{1}{6}x^6 \right) \right. \\
 &= 2 \cdot \left(1 - \frac{1}{4} - \frac{1}{6} \right) \\
 &= \frac{7}{6}
 \end{aligned}$$

Beispiel 7.10. Bestimmen sie die Fläche zwischen der Funktion $x^3 - x^2 - 2x$ und der x -Achse.