

# Zur Minimierung einer reellwertigen Funktion

Jens-Peter M. Zemke  
zemke@tu-harburg.de

Institut für Numerische Simulation  
Technische Universität Hamburg-Harburg

22.10.2007



Zielsetzung

Einordnung

Abgrenzung

## Zielsetzung

- Einordnung
- Abgrenzung

## Mengen, Minima & Funktionen

- Mengen
- Minima
- Funktionen

## Zielsetzung

- Einordnung
- Abgrenzung

## Mengen, Minima & Funktionen

- Mengen
- Minima
- Funktionen

## Das Newton-Verfahren

- Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Übersicht

## Zielsetzung

Einordnung

Abgrenzung

## Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

## Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion.

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst **wenig** Material verwenden,

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst **sichere** Dinge erbauen,

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst sichere Dinge erbauen,
- ▶ möglichst **kurze** Rechenzeiten haben,

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst sichere Dinge erbauen,
- ▶ möglichst kurze Rechenzeiten haben,
- ▶ möglichst **lange** schlafen,

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst sichere Dinge erbauen,
- ▶ möglichst kurze Rechenzeiten haben,
- ▶ möglichst lange schlafen,
- ▶ möglichst **gute** Noten in

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst sichere Dinge erbauen,
- ▶ möglichst kurze Rechenzeiten haben,
- ▶ möglichst lange schlafen,
- ▶ möglichst gute Noten in
- ▶ möglichst **kurzer** Zeit erwirtschaften

# Einordnung

Viele Aufgaben der Ingenieurwissenschaften führen auf die Minimierung (oder Maximierung) einer reellwertigen Funktion. Oft möchte man z.B.

- ▶ möglichst wenig Material verwenden,
- ▶ möglichst sichere Dinge erbauen,
- ▶ möglichst kurze Rechenzeiten haben,
- ▶ möglichst lange schlafen,
- ▶ möglichst gute Noten in
- ▶ möglichst kurzer Zeit erwirtschaften . . .

Solche Aufgaben nennt man allgemein **Optimierungsaufgaben**. Wir werden hier nur die **Minimierung einer reellwertigen Funktion** betrachten (die Aufgabe der Maximierung ist dazu äquivalent, man muss ja nur die negative Funktion minimieren).

# Übersicht

## Zielsetzung

Einordnung

**Abgrenzung**

## Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

## Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Abgrenzung

Minimierungsaufgaben treten auch in ganzzahligen Kontexten auf, z.B. in der Minimierung der Kosten (Cent-Beträge sind im Allgemeinen ganzzahlig) oder Dezimierung der Mitarbeiter.

# Abgrenzung

Minimierungsaufgaben treten auch in ganzzahligen Kontexten auf, z.B. in der Minimierung der Kosten (Cent-Beträge sind im Allgemeinen ganzzahlig) oder Dezimierung der Mitarbeiter. Solche Optimierungsaufgaben betrachten wir hier nicht. Wir beschränken uns der Einfachheit halber auf Funktionen, die reelle Zahlen als Ergebnis liefern.

# Abgrenzung

Minimierungsaufgaben treten auch in ganzzahligen Kontexten auf, z.B. in der Minimierung der Kosten (Cent-Beträge sind im Allgemeinen ganzzahlig) oder Dezimierung der Mitarbeiter. Solche Optimierungsaufgaben betrachten wir hier nicht. Wir beschränken uns der Einfachheit halber auf Funktionen, die reelle Zahlen als Ergebnis liefern.

Wir setzen weiter voraus, dass die auftretenden Vorgänge/Funktionen “glatt” genug sind, so dass in einer geeigneten kleinen Umgebung eines optimalen Punktes die Funktion sich wie eine Parabel verhält. Das führt auf die zweifache stetige Differenzierbarkeit (dazu gleich mehr).

# Abgrenzung

Minimierungsaufgaben treten auch in ganzzahligen Kontexten auf, z.B. in der Minimierung der Kosten (Cent-Beträge sind im Allgemeinen ganzzahlig) oder Dezimierung der Mitarbeiter. Solche Optimierungsaufgaben betrachten wir hier nicht. Wir beschränken uns der Einfachheit halber auf Funktionen, die reelle Zahlen als Ergebnis liefern.

Wir setzen weiter voraus, dass die auftretenden Vorgänge/Funktionen “glatt” genug sind, so dass in einer geeigneten kleinen Umgebung eines optimalen Punktes die Funktion sich wie eine Parabel verhält. Das führt auf die zweifache stetige Differenzierbarkeit (dazu gleich mehr).

Es gibt sehr viele Verfahren, die eine näherungsweise Lösung von Optimierungsaufgaben anstreben. Unter diesen werden wir nur eines der bedeutenderen und am häufigsten verwendeten herauspicken, das **Newton-Verfahren**.

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

**Mengen**

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

Diese “Definition” ist problematisch, siehe dazu die [Russellsche Antinomie](#).

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

Diese “Definition” ist problematisch, siehe dazu die [Russellsche Antinomie](#).

Besser wäre die Verwendung einer axiomatische Mengenlehre, z.B. die der Zermelo-Fraenkel-Mengenlehre mit oder ohne Auswahlaxiom (ZF/ZFC). Das würde aber den Rahmen dieser einführenden Vorlesung sprengen.

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

Diese “Definition” ist problematisch, siehe dazu die [Russellsche Antinomie](#).

Besser wäre die Verwendung einer axiomatische Mengenlehre, z.B. die der Zermelo-Fraenkel-Mengenlehre mit oder ohne Auswahlaxiom (ZF/ZFC). Das würde aber den Rahmen dieser einführenden Vorlesung sprengen.

Mengen werden im Folgenden mit Buchstaben wie  $\mathbb{M}$ ,  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $\mathbb{Q}$ ,  $\mathbb{R}$ ,  $\mathbb{C}$ ,  $\mathbb{H}$  und  $\mathbb{O}$  bezeichnet, dabei steht  $\mathbb{M}$  für eine beliebige Menge.

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

Diese “Definition” ist problematisch, siehe dazu die [Russellsche Antinomie](#).

Besser wäre die Verwendung einer axiomatische Mengenlehre, z.B. die der Zermelo-Fraenkel-Mengenlehre mit oder ohne Auswahlaxiom (ZF/ZFC). Das würde aber den Rahmen dieser einführenden Vorlesung sprengen.

Mengen werden im Folgenden mit Buchstaben wie  $M, N, Z, Q, R, C, H$  und  $\mathbb{O}$  bezeichnet, dabei steht  $M$  für eine beliebige Menge. Die Mengen  $N, Z, Q$  und  $R$  sollten aus der Schule bereits bekannt sein als die Mengen der natürlichen, ganzen, rationalen und reellen Zahlen.

# Mengen

Eine Menge ist “naiv” nach Cantor definiert als:

*„[...] eine Zusammenfassung von bestimmten wohl unterschiedenen Objekten der Anschauung oder des Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.“*

Diese “Definition” ist problematisch, siehe dazu die [Russellsche Antinomie](#).

Besser wäre die Verwendung einer axiomatische Mengenlehre, z.B. die der Zermelo-Fraenkel-Mengenlehre mit oder ohne Auswahlaxiom (ZF/ZFC). Das würde aber den Rahmen dieser einführenden Vorlesung sprengen.

Mengen werden im Folgenden mit Buchstaben wie  $M, N, Z, Q, R, C, H$  und  $\mathbb{O}$  bezeichnet, dabei steht  $M$  für eine beliebige Menge. Die Mengen  $N, Z, Q$  und  $R$  sollten aus der Schule bereits bekannt sein als die Mengen der natürlichen, ganzen, rationalen und reellen Zahlen.

Die Buchstaben  $C, H$  und  $\mathbb{O}$  bezeichnen die Mengen der komplexen Zahlen, Quaternionen (Hamilton-Zahlen) und Oktonionen (Cayley-Zahlen).

# Verknüpfungen

Die Schul“definition” der reellen Zahlen  $\mathbb{R}$  ist leider meist zweifelhaft. Wir geben hier eine “saubere” Definition an.

# Verknüpfungen

Die Schul“definition“ der reellen Zahlen  $\mathbb{R}$  ist leider meist zweifelhaft. Wir geben hier eine “saubere“ Definition an.

Dazu müssen wir einige weitere Definitionen angeben, die nachher nur dazu dienen, die “Eigenschaften“ der reellen Zahlen zu erfassen. Die ersten dieser Eigenschaften betreffen das “Rechnen“ mit den zu definierenden reellen Zahlen.

# Verknüpfungen

Die Schul“definition“ der reellen Zahlen  $\mathbb{R}$  ist leider meist zweifelhaft. Wir geben hier eine “saubere“ Definition an.

Dazu müssen wir einige weitere Definitionen angeben, die nachher nur dazu dienen, die “Eigenschaften“ der reellen Zahlen zu erfassen. Die ersten dieser Eigenschaften betreffen das “Rechnen“ mit den zu definierenden reellen Zahlen:

## Definition (Verknüpfung)

Eine Verknüpfung  $\circ$  auf einer Menge  $\mathbb{M}$  ist eine Abbildung

$$\circ : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{M}, \quad \circ : (m, n) \rightarrow m \circ n. \quad (1)$$

# Verknüpfungen

Die Schul“definition“ der reellen Zahlen  $\mathbb{R}$  ist leider meist zweifelhaft. Wir geben hier eine “saubere“ Definition an.

Dazu müssen wir einige weitere Definitionen angeben, die nachher nur dazu dienen, die “Eigenschaften“ der reellen Zahlen zu erfassen. Die ersten dieser Eigenschaften betreffen das “Rechnen“ mit den zu definierenden reellen Zahlen:

## Definition (Verknüpfung)

Eine Verknüpfung  $\circ$  auf einer Menge  $M$  ist eine Abbildung

$$\circ : M \times M \rightarrow M, \quad \circ : (m, n) \rightarrow m \circ n. \quad (1)$$

Beispiele für Verknüpfungen sind die Addition  $+$  und die Multiplikation  $\cdot$  auf  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $\mathbb{Q}$  und  $\mathbb{R}$ , aber z.B. nicht die Division auf  $\mathbb{N}$ , da  $1/2 \notin \mathbb{N}$ .

# Gruppen

Nun sind die genannten Verknüpfungen oftmals “umkehrbar”, was in der nächsten Definition erfasst wird:

# Gruppen

Nun sind die genannten Verknüpfungen oftmals “umkehrbar”, was in der nächsten Definition erfasst wird:

## Definition (Abelsche Gruppe)

Eine abelsche Gruppe  $(\mathbb{G}, \circ)$  ist eine Menge  $\mathbb{G}$  mit einer Verknüpfung  $\circ$ , so dass

$$\forall a, b, c \in \mathbb{G} : (a \circ b) \circ c = a \circ (b \circ c), \quad (\text{Assoziativität von } \circ)$$

$$\exists e \in \mathbb{G} : \forall a \in \mathbb{G} \quad e \circ a = a \circ e = a, \quad (\text{Existenz neutraler Elemente } e)$$

$$\forall a \in \mathbb{G} \quad \exists a' \in \mathbb{G} : a \circ a' = a' \circ a = e, \quad (\text{Existenz des inversen Elementes})$$

$$\forall a, b \in \mathbb{G} : a \circ b = b \circ a. \quad (\text{Kommutativität von } \circ)$$

# Gruppen

Nun sind die genannten Verknüpfungen oftmals “umkehrbar”, was in der nächsten Definition erfasst wird:

## Definition (Abelsche Gruppe)

Eine abelsche Gruppe  $(\mathbb{G}, \circ)$  ist eine Menge  $\mathbb{G}$  mit einer Verknüpfung  $\circ$ , so dass

$$\forall a, b, c \in \mathbb{G} : (a \circ b) \circ c = a \circ (b \circ c), \quad (\text{Assoziativität von } \circ)$$

$$\exists e \in \mathbb{G} : \forall a \in \mathbb{G} \quad e \circ a = a \circ e = a, \quad (\text{Existenz neutraler Elemente } e)$$

$$\forall a \in \mathbb{G} \quad \exists a' \in \mathbb{G} : a \circ a' = a' \circ a = e, \quad (\text{Existenz des inversen Elementes})$$

$$\forall a, b \in \mathbb{G} : a \circ b = b \circ a. \quad (\text{Kommutativität von } \circ)$$

Oft werden die Zeichen  $+$ ,  $\cdot$  statt  $\circ$  verwendet. Die Verknüpfung  $+$  nennt man eine **Addition** und bezeichnet das neutrale Element mit 0. Die Verknüpfung  $\cdot$  nennt man eine **Multiplikation** und bezeichnet das neutrale Element mit 1.

# Gruppen

Ein Beispiel einer abelschen Gruppe ist  $(\mathbb{Z}, +)$ :

- ▶ das neutrale Element ist  $0 \in \mathbb{Z}$ , und
- ▶ das inverse Element von  $z \in \mathbb{Z}$  ist  $-z \in \mathbb{Z}$ .

# Gruppen

Ein Beispiel einer abelschen Gruppe ist  $(\mathbb{Z}, +)$ :

- ▶ das neutrale Element ist  $0 \in \mathbb{Z}$ , und
- ▶ das inverse Element von  $z \in \mathbb{Z}$  ist  $-z \in \mathbb{Z}$ .

Ein anderes Beispiel einer abelschen Gruppe ist  $(\mathbb{Q} \setminus \{0\}, \cdot)$ :

- ▶ das neutrale Element ist  $1 \in \mathbb{Q}$ , und
- ▶ das inverse Element von  $x = p/q \in \mathbb{Q}$ , wobei  $p \in \mathbb{Z}$ ,  $q \in \mathbb{N}$ , ist  $q/p \in \mathbb{Q}$ .

# Gruppen

Ein Beispiel einer abelschen Gruppe ist  $(\mathbb{Z}, +)$ :

- ▶ das neutrale Element ist  $0 \in \mathbb{Z}$ , und
- ▶ das inverse Element von  $z \in \mathbb{Z}$  ist  $-z \in \mathbb{Z}$ .

Ein anderes Beispiel einer abelschen Gruppe ist  $(\mathbb{Q} \setminus \{0\}, \cdot)$ :

- ▶ das neutrale Element ist  $1 \in \mathbb{Q}$ , und
- ▶ das inverse Element von  $x = p/q \in \mathbb{Q}$ , wobei  $p \in \mathbb{Z}$ ,  $q \in \mathbb{N}$ , ist  $q/p \in \mathbb{Q}$ .

$(\mathbb{N}, +)$  ist keine (abelsche) Gruppe, da (z.B.) das Element  $2 \in \mathbb{N}$  kein inverses Element **in**  $\mathbb{N}$  hat.

# Gruppen

Ein Beispiel einer abelschen Gruppe ist  $(\mathbb{Z}, +)$ :

- ▶ das neutrale Element ist  $0 \in \mathbb{Z}$ , und
- ▶ das inverse Element von  $z \in \mathbb{Z}$  ist  $-z \in \mathbb{Z}$ .

Ein anderes Beispiel einer abelschen Gruppe ist  $(\mathbb{Q} \setminus \{0\}, \cdot)$ :

- ▶ das neutrale Element ist  $1 \in \mathbb{Q}$ , und
- ▶ das inverse Element von  $x = p/q \in \mathbb{Q}$ , wobei  $p \in \mathbb{Z}$ ,  $q \in \mathbb{N}$ , ist  $q/p \in \mathbb{Q}$ .

$(\mathbb{N}, +)$  ist keine (abelsche) Gruppe, da (z.B.) das Element  $2 \in \mathbb{N}$  kein inverses Element **in**  $\mathbb{N}$  hat.

Dass es ein inverses Element  $-2$  von  $2 \in \mathbb{Z}$  gibt, ändert daran auch nichts ...

# Gruppen

Ein Beispiel einer abelschen Gruppe ist  $(\mathbb{Z}, +)$ :

- ▶ das neutrale Element ist  $0 \in \mathbb{Z}$ , und
- ▶ das inverse Element von  $z \in \mathbb{Z}$  ist  $-z \in \mathbb{Z}$ .

Ein anderes Beispiel einer abelschen Gruppe ist  $(\mathbb{Q} \setminus \{0\}, \cdot)$ :

- ▶ das neutrale Element ist  $1 \in \mathbb{Q}$ , und
- ▶ das inverse Element von  $x = p/q \in \mathbb{Q}$ , wobei  $p \in \mathbb{Z}$ ,  $q \in \mathbb{N}$ , ist  $q/p \in \mathbb{Q}$ .

$(\mathbb{N}, +)$  ist keine (abelsche) Gruppe, da (z.B.) das Element  $2 \in \mathbb{N}$  kein inverses Element **in**  $\mathbb{N}$  hat.

Dass es ein inverses Element  $-2$  von  $2 \in \mathbb{Z}$  gibt, ändert daran auch nichts ...

Auch  $(\mathbb{Q}, \cdot)$  ist keine abelsche Gruppe, da  $0$  kein inverses Element besitzt.

# Körper

Nun haben wir bei den reellen Zahlen **zwei** Verknüpfungen und müssen auch die Verträglichkeit der beiden Verknüpfungen untereinander erfassen.

# Körper

Nun haben wir bei den reellen Zahlen **zwei** Verknüpfungen und müssen auch die Verträglichkeit der beiden Verknüpfungen untereinander erfassen:

## Definition (Körper)

Ein Körper  $(\mathbb{K}, +, \cdot)$  ist eine Menge  $\mathbb{K}$  mit zwei Verknüpfungen  $+$ ,  $\cdot$ , so dass

$(\mathbb{K}, +)$  ist eine abelsche Gruppe,

$(\mathbb{K}$  ist additive Gruppe)

$(\mathbb{K} \setminus \{0\}, \cdot)$  ist eine abelsche Gruppe,

$(\mathbb{K} \setminus \{0\}$  ist multiplikative Gruppe)

$\forall a, b, c \in \mathbb{K} : a \cdot (b + c) = a \cdot b + a \cdot c.$

(Distributivität)

# Körper

Nun haben wir bei den reellen Zahlen **zwei** Verknüpfungen und müssen auch die Verträglichkeit der beiden Verknüpfungen untereinander erfassen:

## Definition (Körper)

Ein Körper  $(\mathbb{K}, +, \cdot)$  ist eine Menge  $\mathbb{K}$  mit zwei Verknüpfungen  $+$ ,  $\cdot$ , so dass

$(\mathbb{K}, +)$  ist eine abelsche Gruppe,  $(\mathbb{K}$  ist additive Gruppe)

$(\mathbb{K} \setminus \{0\}, \cdot)$  ist eine abelsche Gruppe,  $(\mathbb{K} \setminus \{0\}$  ist multiplikative Gruppe)

$\forall a, b, c \in \mathbb{K} : a \cdot (b + c) = a \cdot b + a \cdot c$  (Distributivität)

Beispiele von Körpern sind  $(\mathbb{Q}, +, \cdot)$ ,  $(\mathbb{R}, +, \cdot)$  und  $(\mathbb{C}, +, \cdot)$ .

# Körper

Nun haben wir bei den reellen Zahlen **zwei** Verknüpfungen und müssen auch die Verträglichkeit der beiden Verknüpfungen untereinander erfassen:

## Definition (Körper)

Ein Körper  $(\mathbb{K}, +, \cdot)$  ist eine Menge  $\mathbb{K}$  mit zwei Verknüpfungen  $+$ ,  $\cdot$ , so dass

$(\mathbb{K}, +)$  ist eine abelsche Gruppe,  $(\mathbb{K}$  ist additive Gruppe)

$(\mathbb{K} \setminus \{0\}, \cdot)$  ist eine abelsche Gruppe,  $(\mathbb{K} \setminus \{0\}$  ist multiplikative Gruppe)

$\forall a, b, c \in \mathbb{K} : a \cdot (b + c) = a \cdot b + a \cdot c$  (Distributivität)

Beispiele von Körpern sind  $(\mathbb{Q}, +, \cdot)$ ,  $(\mathbb{R}, +, \cdot)$  und  $(\mathbb{C}, +, \cdot)$ .

Auch  $(\{0, 1\}, +, \cdot)$  ist ein Körper, wobei (sogenannte “Rechnung modulo 2”)

$$\begin{array}{c|cc}
 + & 0 & 1 \\
 \hline
 0 & 0 & 1 \\
 1 & 1 & 0
 \end{array}, \quad
 \begin{array}{c|cc}
 \cdot & 0 & 1 \\
 \hline
 0 & 0 & 0 \\
 1 & 0 & 1
 \end{array}. \quad (2)$$

# Notationen

Wir verwenden folgende abkürzende Notationen:

# Notationen

Wir verwenden folgende abkürzende Notationen:

## Definition

- ▶ Das neutrale Element bzgl. der Addition wird mit  $0$  bezeichnet.

# Notationen

Wir verwenden folgende abkürzende Notationen:

## Definition

- ▶ Das neutrale Element bzgl. der Addition wird mit 0 bezeichnet.
- ▶ Das neutrale Element bzgl. der Multiplikation wird mit 1 bezeichnet.

# Notationen

Wir verwenden folgende abkürzende Notationen:

## Definition

- ▶ Das neutrale Element bzgl. der Addition wird mit 0 bezeichnet.
- ▶ Das neutrale Element bzgl. der Multiplikation wird mit 1 bezeichnet.
- ▶ Das inverse Element von  $a \in \mathbb{K}$  bzgl. der Addition wird mit  $-a$  bezeichnet.

# Notationen

Wir verwenden folgende abkürzende Notationen:

## Definition

- ▶ Das neutrale Element bzgl. der Addition wird mit  $0$  bezeichnet.
- ▶ Das neutrale Element bzgl. der Multiplikation wird mit  $1$  bezeichnet.
- ▶ Das inverse Element von  $a \in \mathbb{K}$  bzgl. der Addition wird mit  $-a$  bezeichnet.
- ▶ Das inverse Element von  $a \in \mathbb{K}$  bzgl. der Multiplikation wird mit  $a^{-1}$  bezeichnet.

# Notationen

Wir verwenden folgende abkürzende Notationen:

## Definition

- ▶ Das neutrale Element bzgl. der Addition wird mit 0 bezeichnet.
- ▶ Das neutrale Element bzgl. der Multiplikation wird mit 1 bezeichnet.
- ▶ Das inverse Element von  $a \in \mathbb{K}$  bzgl. der Addition wird mit  $-a$  bezeichnet.
- ▶ Das inverse Element von  $a \in \mathbb{K}$  bzgl. der Multiplikation wird mit  $a^{-1}$  bezeichnet.
- ▶ Wir verwenden die folgenden Kurzschreibweisen:

$$a - b := a + (-b) \quad (3)$$

$$\frac{a}{b} := a/b := a \cdot b^{-1}. \quad (4)$$

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

**Minima**

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Ordnungen

Reelle Zahlen sind untereinander “vergleichbar”, also “geordnet”.

# Ordnungen

Reelle Zahlen sind untereinander “vergleichbar”, also “geordnet”:

## Definition (Ordnung)

Eine Relation  $\leq$  über einer nichtleeren Menge  $\mathbb{M}$  heißt eine totale Ordnung, wenn für alle  $a, b, c \in \mathbb{M}$  gilt

$$a \leq a, \quad \text{(Reflexivität)}$$

$$a \leq b \wedge b \leq a \Rightarrow a = b, \quad \text{(Antisymmetrie)}$$

$$a \leq b \wedge b \leq c \Rightarrow a \leq c, \quad \text{(Transitivität)}$$

$$a \leq b \vee b \leq a. \quad \text{(Totalität)}$$

Die Menge  $\mathbb{M}$  heißt dann eine **total geordnete Menge**.

# Ordnungen

Reelle Zahlen sind untereinander “vergleichbar”, also “geordnet”:

## Definition (Ordnung)

Eine Relation  $\leq$  über einer nichtleeren Menge  $\mathbb{M}$  heißt eine totale Ordnung, wenn für alle  $a, b, c \in \mathbb{M}$  gilt

$$a \leq a, \quad \text{(Reflexivität)}$$

$$a \leq b \wedge b \leq a \Rightarrow a = b, \quad \text{(Antisymmetrie)}$$

$$a \leq b \wedge b \leq c \Rightarrow a \leq c, \quad \text{(Transitivität)}$$

$$a \leq b \vee b \leq a. \quad \text{(Totalität)}$$

Die Menge  $\mathbb{M}$  heißt dann eine **total geordnete Menge**.

Sind nur die ersten drei Axiome erfüllt, so spricht man von einer **partiellen Ordnung** und nennt  $\mathbb{M}$  dann eine **partiell geordnete Menge**.

# Ordnungen

Als abgekürzte Notation definieren wir noch:

# Ordnungen

Als abgekürzte Notation definieren wir noch:

## Definition (Schreibweisen)

Als abkürzende Schreibweisen definieren wir

$$a \geq b := b \leq a, \quad (5)$$

$$a < b := (a \leq b \wedge a \neq b), \quad (6)$$

$$a > b := b < a. \quad (7)$$

# Ordnungen

Als abgekürzte Notation definieren wir noch:

## Definition (Schreibweisen)

Als abkürzende Schreibweisen definieren wir

$$a \geq b := b \leq a, \quad (5)$$

$$a < b := (a \leq b \wedge a \neq b), \quad (6)$$

$$a > b := b < a. \quad (7)$$

Haben wir nun eine (partielle) Ordnung auf einer Menge definiert/gefunden, so können wir Minima und Maxima suchen.

# Extrema

Die Definition einer Ordnung ermöglicht erst die Frage nach dem “größten” oder “kleinsten” Element einer (Teil-)Menge.

# Extrema

Die Definition einer Ordnung ermöglicht erst die Frage nach dem “größten” oder “kleinsten” Element einer (Teil-)Menge:

## Definition

Sei  $(M, \leq)$  eine (partielle) Ordnung und  $T \subset M$ . Ein Element  $s \in T$  heißt

größtes Element oder Maximum:  $t \leq s \quad \forall t \in T$ ,                      (Maximum)

kleinstes Element oder Minimum:  $s \leq t \quad \forall t \in T$ .                      (Minimum)

Ein Element ist ein Extremum, wenn es ein Minimum oder ein Maximum ist.

# Extrema

Die Definition einer Ordnung ermöglicht erst die Frage nach dem “größten” oder “kleinsten” Element einer (Teil-)Menge:

## Definition

Sei  $(M, \leq)$  eine (partielle) Ordnung und  $T \subset M$ . Ein Element  $s \in T$  heißt

größtes Element oder Maximum:  $t \leq s \quad \forall t \in T$ ,                    (Maximum)

kleinstes Element oder Minimum:  $s \leq t \quad \forall t \in T$ .                    (Minimum)

Ein Element ist ein Extremum, wenn es ein Minimum oder ein Maximum ist.

Nun muss es nicht so sein, dass eine Teilmenge ein kleinstes Element enthält, z.B. hat jedes nach unten offene Intervall  $(a, b] \subset \mathbb{R}$ ,  $a < b$ , kein Minimum.

# Extrema

Die Definition einer Ordnung ermöglicht erst die Frage nach dem “größten” oder “kleinsten” Element einer (Teil-)Menge:

## Definition

Sei  $(M, \leq)$  eine (partielle) Ordnung und  $T \subset M$ . Ein Element  $s \in T$  heißt

größtes Element oder Maximum:  $t \leq s \quad \forall t \in T$ ,                      (Maximum)

kleinstes Element oder Minimum:  $s \leq t \quad \forall t \in T$ .                      (Minimum)

Ein Element ist ein Extremum, wenn es ein Minimum oder ein Maximum ist.

Nun muss es nicht so sein, dass eine Teilmenge ein kleinstes Element enthält, z.B. hat jedes nach unten offene Intervall  $(a, b] \subset \mathbb{R}$ ,  $a < b$ , kein Minimum.

Um auch solche Fälle behandeln zu können, werden noch zwei weitere Begriffe definiert.

# Ordnungen

## Definition

Sei  $(\mathbb{M}, \leq)$  eine (partielle) Ordnung und  $\mathbb{T} \subset \mathbb{M}$ . Ein Element  $s \in \mathbb{M}$  heißt kleinste obere Schranke oder **Supremum**:

$$t \leq s \leq s' \quad \forall t \in \mathbb{T}, \quad \forall s' \in \mathbb{M}. \quad (\text{Supremum})$$

Ein Element  $s \in \mathbb{M}$  heißt größte untere Schranke oder **Infimum**:

$$s' \leq s \leq t \quad \forall t \in \mathbb{T}, \quad \forall s' \in \mathbb{M}. \quad (\text{Infimum})$$

# Ordnungen

## Definition

Sei  $(\mathbb{M}, \leq)$  eine (partielle) Ordnung und  $\mathbb{T} \subset \mathbb{M}$ . Ein Element  $s \in \mathbb{M}$  heißt kleinste obere Schranke oder **Supremum**:

$$t \leq s \leq s' \quad \forall t \in \mathbb{T}, \quad \forall s' \in \mathbb{M}. \quad (\text{Supremum})$$

Ein Element  $s \in \mathbb{M}$  heißt größte untere Schranke oder **Infimum**:

$$s' \leq s \leq t \quad \forall t \in \mathbb{T}, \quad \forall s' \in \mathbb{M}. \quad (\text{Infimum})$$

Das vorher genannte Intervall  $(a, b]$  hat jetzt das Infimum  $a$ . Bei beliebigen abgeschlossenen Intervallen der Form  $[a, b]$  ist  $a$  sowohl Infimum als auch Minimum (Minimum  $\Rightarrow$  Infimum).

# Ordnungen

Es kann nun passieren, dass eine Teilmenge einer Menge kein Supremum (oder Infimum oder beides) hat. Das erschwert die Suche nach dem kleinsten Element.

# Ordnungen

Es kann nun passieren, dass eine Teilmenge einer Menge kein Supremum (oder Infimum oder beides) hat. Das erschwert die Suche nach dem kleinsten Element.

Ein Beispiel dafür ist die Menge

$$\mathbb{S} := \{x \in \mathbb{Q} : 1 \leq x < \sqrt{2}\}. \quad (8)$$

# Ordnungen

Es kann nun passieren, dass eine Teilmenge einer Menge kein Supremum (oder Infimum oder beides) hat. Das erschwert die Suche nach dem kleinsten Element.

Ein Beispiel dafür ist die Menge

$$S := \{x \in \mathbb{Q} : 1 \leq x < \sqrt{2}\}. \quad (8)$$

Daher definiert man

## Definition (Ordnungsvollständigkeit)

Eine Menge heißt **ordnungsvollständig**, wenn jede nicht-leere nach oben beschränkte Teilmenge ein Supremum besitzt.

# Ordnungen

Es kann nun passieren, dass eine Teilmenge einer Menge kein Supremum (oder Infimum oder beides) hat. Das erschwert die Suche nach dem kleinsten Element.

Ein Beispiel dafür ist die Menge

$$S := \{x \in \mathbb{Q} : 1 \leq x < \sqrt{2}\}. \quad (8)$$

Daher definiert man

## Definition (Ordnungsvollständigkeit)

Eine Menge heißt **ordnungsvollständig**, wenn jede nicht-leere nach oben beschränkte Teilmenge ein Supremum besitzt.

Das Beispiel zeigt, dass  $\mathbb{Q}$  nicht ordnungsvollständig ist.

# Ordnungen

Die Ordnung der reellen Zahlen muss noch mit den bereits erfassten Rechenregeln harmonisieren.

# Ordnungen

Die Ordnung der reellen Zahlen muss noch mit den bereits erfassten Rechenregeln harmonisieren:

## Definition (Angeordneter Körper)

Ein Körper  $(\mathbb{K}, +, \cdot)$  heißt angeordneter Körper  $(\mathbb{K}, +, \cdot, \leq)$ , wenn für alle  $a, b, c \in \mathbb{K}$  gilt

$(\mathbb{K}, \leq)$  ist eine total geordnete Menge,

$$a \leq b \Rightarrow a + c \leq b + c,$$

$$(a \leq b \wedge 0 \leq c) \Rightarrow a \cdot c \leq b \cdot c.$$

( $\mathbb{K}$  ist total geordnet)

(Verträglichkeit mit Addition)

(Verträglichkeit mit Multiplikation)

# Ordnungen

Die Ordnung der reellen Zahlen muss noch mit den bereits erfassten Rechenregeln harmonisieren:

## Definition (Angeordneter Körper)

Ein Körper  $(\mathbb{K}, +, \cdot)$  heißt angeordneter Körper  $(\mathbb{K}, +, \cdot, \leq)$ , wenn für alle  $a, b, c \in \mathbb{K}$  gilt

$(\mathbb{K}, \leq)$  ist eine total geordnete Menge,

$$a \leq b \Rightarrow a + c \leq b + c,$$

$$(a \leq b \wedge 0 \leq c) \Rightarrow a \cdot c \leq b \cdot c.$$

( $\mathbb{K}$  ist total geordnet)

(Verträglichkeit mit Addition)

(Verträglichkeit mit Multiplikation)

## Definition (Vollständiger (angeordneter) Körper)

Ein angeordneter Körper  $(\mathbb{K}, +, \cdot, \leq)$  heißt **vollständig**, wenn  $(\mathbb{K}, \leq)$  ordnungsvollständig ist.

# Definition der reellen Zahlen

Wenn sich zwei Begrifflichkeiten nur durch eine “Umbenennung” unterscheiden, so handelt es sich um dieselbe Sache, wenn auch anders gesehen. Die Konkretisierung dieser Aussage führt auf den Begriff der Isomorphie.

# Definition der reellen Zahlen

Wenn sich zwei Begrifflichkeiten nur durch eine “Umbenennung” unterscheiden, so handelt es sich um dieselbe Sache, wenn auch anders gesehen. Die Konkretisierung dieser Aussage führt auf den Begriff der Isomorphie:

Zwei “algebraische Strukturen”  $\mathcal{S}_1$  und  $\mathcal{S}_2$  heißen **isomorph**, wenn es eine bijektive Abbildung  $\iota$  gibt, so dass die “Eigenschaften” (von z.B.  $\cdot_1, +_1, \leq_1$  und  $\cdot_2, +_2, \leq_2$ ) aufeinander abgebildet werden.

# Definition der reellen Zahlen

Wenn sich zwei Begrifflichkeiten nur durch eine “Umbenennung” unterscheiden, so handelt es sich um dieselbe Sache, wenn auch anders gesehen. Die Konkretisierung dieser Aussage führt auf den Begriff der Isomorphie:

Zwei “algebraische Strukturen”  $\mathcal{S}_1$  und  $\mathcal{S}_2$  heißen **isomorph**, wenn es eine bijektive Abbildung  $\iota$  gibt, so dass die “Eigenschaften” (von z.B.  $\cdot_1, +_1, \leq_1$  und  $\cdot_2, +_2, \leq_2$ ) aufeinander abgebildet werden.

Jetzt können wir die reellen Zahlen eindeutig charakterisieren, wobei wir stillschweigend die Aussage der Isomorphie glauben.

# Definition der reellen Zahlen

Wenn sich zwei Begrifflichkeiten nur durch eine “Umbenennung” unterscheiden, so handelt es sich um dieselbe Sache, wenn auch anders gesehen. Die Konkretisierung dieser Aussage führt auf den Begriff der Isomorphie:

Zwei “algebraische Strukturen”  $\mathcal{S}_1$  und  $\mathcal{S}_2$  heißen **isomorph**, wenn es eine bijektive Abbildung  $\iota$  gibt, so dass die “Eigenschaften” (von z.B.  $\cdot_1, +_1, \leq_1$  und  $\cdot_2, +_2, \leq_2$ ) aufeinander abgebildet werden.

Jetzt können wir die reellen Zahlen eindeutig charakterisieren, wobei wir stillschweigend die Aussage der Isomorphie glauben:

## Definition (Reelle Zahlen)

Die reellen Zahlen  $\mathbb{R}$  sind der bis auf Isomorphie eindeutig bestimmte vollständige angeordnete Körper  $(\mathbb{K}, +, \cdot, \leq)$ .

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

## Mengen, Minima & Funktionen

Mengen

Minima

**Funktionen**

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Funktionen

## Definition (Kartesisches Produkt)

Das kartesische Produkt zweier Mengen  $\mathbb{M}_1$  und  $\mathbb{M}_2$  ist definiert als die Menge  $\mathbb{M}_1 \times \mathbb{M}_2$  der geordneten Tupel,

$$\mathbb{M}_1 \times \mathbb{M}_2 = \{(m_1, m_2) : m_1 \in \mathbb{M}_1, m_2 \in \mathbb{M}_2\}. \quad (9)$$

# Funktionen

## Definition (Kartesisches Produkt)

Das kartesische Produkt zweier Mengen  $\mathbb{M}_1$  und  $\mathbb{M}_2$  ist definiert als die Menge  $\mathbb{M}_1 \times \mathbb{M}_2$  der geordneten Tupel,

$$\mathbb{M}_1 \times \mathbb{M}_2 = \{(m_1, m_2) : m_1 \in \mathbb{M}_1, m_2 \in \mathbb{M}_2\}. \quad (9)$$

## Definition (Funktion)

Eine Funktion  $f : \mathbb{M}_1 \rightarrow \mathbb{M}_2$  ist eine Teilmenge  $\mathbb{T} \subset \mathbb{M}_1 \times \mathbb{M}_2$ , so dass

$$\forall m_1 \in \mathbb{M}_1 \quad \exists! m_2 \in \mathbb{M}_2 : (m_1, m_2) \in \mathbb{T}. \quad (10)$$

# Funktionen

## Definition (Kartesisches Produkt)

Das kartesische Produkt zweier Mengen  $\mathbb{M}_1$  und  $\mathbb{M}_2$  ist definiert als die Menge  $\mathbb{M}_1 \times \mathbb{M}_2$  der geordneten Tupel,

$$\mathbb{M}_1 \times \mathbb{M}_2 = \{(m_1, m_2) : m_1 \in \mathbb{M}_1, m_2 \in \mathbb{M}_2\}. \quad (9)$$

## Definition (Funktion)

Eine Funktion  $f : \mathbb{M}_1 \rightarrow \mathbb{M}_2$  ist eine Teilmenge  $\mathbb{T} \subset \mathbb{M}_1 \times \mathbb{M}_2$ , so dass

$$\forall m_1 \in \mathbb{M}_1 \quad \exists! m_2 \in \mathbb{M}_2 : (m_1, m_2) \in \mathbb{T}. \quad (10)$$

$\mathbb{M}_1$  heißt Definitionsbereich von  $f$ ,  $\mathbb{M}_2$  Wertebereich von  $f$ , in Zeichen

$$f : \mathbb{M}_1 \rightarrow \mathbb{M}_2. \quad (11)$$

# Funktionen

## Definition (Kartesisches Produkt)

Das kartesische Produkt zweier Mengen  $\mathbb{M}_1$  und  $\mathbb{M}_2$  ist definiert als die Menge  $\mathbb{M}_1 \times \mathbb{M}_2$  der geordneten Tupel,

$$\mathbb{M}_1 \times \mathbb{M}_2 = \{(m_1, m_2) : m_1 \in \mathbb{M}_1, m_2 \in \mathbb{M}_2\}. \quad (9)$$

## Definition (Funktion)

Eine Funktion  $f : \mathbb{M}_1 \rightarrow \mathbb{M}_2$  ist eine Teilmenge  $\mathbb{T} \subset \mathbb{M}_1 \times \mathbb{M}_2$ , so dass

$$\forall m_1 \in \mathbb{M}_1 \quad \exists! m_2 \in \mathbb{M}_2 : (m_1, m_2) \in \mathbb{T}. \quad (10)$$

$\mathbb{M}_1$  heißt Definitionsbereich von  $f$ ,  $\mathbb{M}_2$  Wertebereich von  $f$ , in Zeichen

$$f : \mathbb{M}_1 \rightarrow \mathbb{M}_2. \quad (11)$$

Das eindeutig bestimmte  $m_2 \in \mathbb{M}_2$  zu  $m_1 \in \mathbb{M}_1$  wird mit  $f(m_1)$  bezeichnet.

# Reellwertige und reelle Funktionen

Wir betrachten nur noch reellwertige Funktionen, also Funktionen, welche als Wertebereich die reellen Zahlen haben, also

$$f : M \rightarrow \mathbb{R}. \quad (12)$$

# Reellwertige und reelle Funktionen

Wir betrachten nur noch reellwertige Funktionen, also Funktionen, welche als Wertebereich die reellen Zahlen haben, also

$$f : M \rightarrow \mathbb{R}. \quad (12)$$

Als ersten Spezialfall betrachten wir Funktionen, welche eine Teilmenge der reellen Zahlen als Definitionsbereich haben, also

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}. \quad (13)$$

# Reellwertige und reelle Funktionen

Wir betrachten nur noch reellwertige Funktionen, also Funktionen, welche als Wertebereich die reellen Zahlen haben, also

$$f : M \rightarrow \mathbb{R}. \quad (12)$$

Als ersten Spezialfall betrachten wir Funktionen, welche eine Teilmenge der reellen Zahlen als Definitionsbereich haben, also

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}. \quad (13)$$

Um Optimierung betreiben zu können, sollten die zu minimierenden Funktionen gewisse Eigenschaften haben. Sie sollten "glatt" genug sein.

# Reellwertige und reelle Funktionen

Wir betrachten nur noch reellwertige Funktionen, also Funktionen, welche als Wertebereich die reellen Zahlen haben, also

$$f : M \rightarrow \mathbb{R}. \quad (12)$$

Als ersten Spezialfall betrachten wir Funktionen, welche eine Teilmenge der reellen Zahlen als Definitionsbereich haben, also

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}. \quad (13)$$

Um Optimierung betreiben zu können, sollten die zu minimierenden Funktionen gewisse Eigenschaften haben. Sie sollten “glatt” genug sein.

Diese “Glätte” wird mathematisch durch die Begriffe der Stetigkeit und der (stetigen) Differenzierbarkeit beschrieben.

# Stetige und differenzierbare Funktionen

## Definition (Stetigkeit reeller Funktionen)

Eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  ist im Punkt  $x \in D$  stetig, wenn

$$\forall \epsilon \in \mathbb{R}_+ \quad \exists \delta \in \mathbb{R}_+ : |\tilde{x} - x| < \delta \quad \Rightarrow \quad |f(\tilde{x}) - f(x)| < \epsilon, \quad \text{wobei } \tilde{x} \in D. \quad (14)$$

Die Funktion heißt stetig, wenn sie in allen Punkten  $x \in D$  stetig ist.

# Stetige und differenzierbare Funktionen

## Definition (Stetigkeit reeller Funktionen)

Eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  ist im Punkt  $x \in D$  stetig, wenn

$$\forall \epsilon \in \mathbb{R}_+ \quad \exists \delta \in \mathbb{R}_+ : |\tilde{x} - x| < \delta \quad \Rightarrow \quad |f(\tilde{x}) - f(x)| < \epsilon, \quad \text{wobei } \tilde{x} \in D. \quad (14)$$

Die Funktion heißt stetig, wenn sie in allen Punkten  $x \in D$  stetig ist.

Anschaulich kann man sich diese Definition als “Wackelimmunität” vorstellen.

# Stetige und differenzierbare Funktionen

## Definition (Stetigkeit reeller Funktionen)

Eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  ist im Punkt  $x \in D$  stetig, wenn

$$\forall \epsilon \in \mathbb{R}_+ \quad \exists \delta \in \mathbb{R}_+ : |\tilde{x} - x| < \delta \quad \Rightarrow \quad |f(\tilde{x}) - f(x)| < \epsilon, \quad \text{wobei } \tilde{x} \in D. \quad (14)$$

Die Funktion heißt stetig, wenn sie in allen Punkten  $x \in D$  stetig ist.

Anschaulich kann man sich diese Definition als “Wackelimmunität” vorstellen.

## Definition (Differenzierbarkeit reeller Funktionen)

Eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  ist im Punkt  $x \in D$  differenzierbar, wenn

$$\frac{df}{dx} := f'(x) := \lim_{\tilde{x} \rightarrow x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} \quad (15)$$

existiert. Die Funktion heißt differenzierbar in  $D$ , wenn sie in allen Punkten  $x \in D$  differenzierbar ist.

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

Man sieht sofort, dass differenzierbare Funktionen stetig sind. Aber nicht jede Ableitung einer differenzierbaren Funktion ist wieder stetig.

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

Man sieht sofort, dass differenzierbare Funktionen stetig sind. Aber nicht jede Ableitung einer differenzierbaren Funktion ist wieder stetig.

Eine differenzierbare Funktion, deren Ableitung  $f' : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$  wieder stetig ist, nennt man (einmal) stetig differenzierbar. Nun kann man fragen, ob diese Ableitung wiederum (stetig) differenzierbar ist.

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

Man sieht sofort, dass differenzierbare Funktionen stetig sind. Aber nicht jede Ableitung einer differenzierbaren Funktion ist wieder stetig.

Eine differenzierbare Funktion, deren Ableitung  $f' : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$  wieder stetig ist, nennt man (einmal) stetig differenzierbar. Nun kann man fragen, ob diese Ableitung wiederum (stetig) differenzierbar ist.

Diese Überlegungen der wiederholten stetigen Differenzierbarkeit reeller Funktionen führen auf die Mengen der  $n$ -fach stetig differenzierbaren Funktionen, in Zeichen  $C^n(D)$ , wobei  $n \in \mathbb{N}$ .

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

Man sieht sofort, dass differenzierbare Funktionen stetig sind. Aber nicht jede Ableitung einer differenzierbaren Funktion ist wieder stetig.

Eine differenzierbare Funktion, deren Ableitung  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$  wieder stetig ist, nennt man (einmal) stetig differenzierbar. Nun kann man fragen, ob diese Ableitung wiederum (stetig) differenzierbar ist.

Diese Überlegungen der wiederholten stetigen Differenzierbarkeit reeller Funktionen führen auf die Mengen der  $n$ -fach stetig differenzierbaren Funktionen, in Zeichen  $C^n(D)$ , wobei  $n \in \mathbb{N}$ .

Für  $f \in C^n(D)$  wird die Funktion der  $k$ ten Ableitung,  $k \leq n$ , bezeichnet mit  $f^{(k)} : D \subset \mathbb{R} \rightarrow \mathbb{R}$ .

# Stetig differenzierbare Funktionen

Die Menge aller stetigen Funktionen bezeichnet man mit  $C^0(D)$ .

Man sieht sofort, dass differenzierbare Funktionen stetig sind. Aber nicht jede Ableitung einer differenzierbaren Funktion ist wieder stetig.

Eine differenzierbare Funktion, deren Ableitung  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$  wieder stetig ist, nennt man (einmal) stetig differenzierbar. Nun kann man fragen, ob diese Ableitung wiederum (stetig) differenzierbar ist.

Diese Überlegungen der wiederholten stetigen Differenzierbarkeit reeller Funktionen führen auf die Mengen der  $n$ -fach stetig differenzierbaren Funktionen, in Zeichen  $C^n(D)$ , wobei  $n \in \mathbb{N}$ .

Für  $f \in C^n(D)$  wird die Funktion der  $k$ ten Ableitung,  $k \leq n$ , bezeichnet mit  $f^{(k)} : D \subset \mathbb{R} \rightarrow \mathbb{R}$ .

Der nächste Schritt ist die Definition der Menge der unendlich oft stetig differenzierbare Funktionen, welche mit  $C^\infty(D)$  bezeichnet wird.

# Taylor-Entwicklung

Wenn eine Funktion hinreichend oft stetig differenzierbar ist, so kann man schauen, wie gut sich die Funktion durch Taylor-Polynome approximieren läßt.

# Taylor-Entwicklung

Wenn eine Funktion hinreichend oft stetig differenzierbar ist, so kann man schauen, wie gut sich die Funktion durch Taylor-Polynome approximieren läßt.

Ein Taylor-Polynom  $T_{x_0, n}$  ist ein Polynom, welches in einem gegebenen Punkt  $x_0$  bis zu dem gegebenen Grad  $n$  mit einer Funktion  $f \in C^n(D)$  und ihren Ableitungen übereinstimmt.

# Taylor-Entwicklung

Wenn eine Funktion hinreichend oft stetig differenzierbar ist, so kann man schauen, wie gut sich die Funktion durch Taylor-Polynome approximieren lässt.

Ein Taylor-Polynom  $T_{x_0,n}$  ist ein Polynom, welches in einem gegebenen Punkt  $x_0$  bis zu dem gegebenen Grad  $n$  mit einer Funktion  $f \in C^n(D)$  und ihren Ableitungen übereinstimmt.

Man sieht sofort, dass dieses Taylor-Polynom  $T_{x_0,n}$  gegeben ist durch

$$T_{x_0,n}(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \quad (16)$$

# Taylor-Entwicklung

Wenn eine Funktion hinreichend oft stetig differenzierbar ist, so kann man schauen, wie gut sich die Funktion durch Taylor-Polynome approximieren läßt.

Ein Taylor-Polynom  $T_{x_0,n}$  ist ein Polynom, welches in einem gegebenen Punkt  $x_0$  bis zu dem gegebenen Grad  $n$  mit einer Funktion  $f \in C^n(D)$  und ihren Ableitungen übereinstimmt.

Man sieht sofort, dass dieses Taylor-Polynom  $T_{x_0,n}$  gegeben ist durch

$$T_{x_0,n}(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \quad (16)$$

Ist die Funktion sogar aus  $C^{n+1}(D)$ , so gilt die folgende Lagrangsche Restgliedformel für den Fehler zwischen der Funktion und der Taylor-Approximation:

$$f(x) - T_{x_0,n}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)^{n+1}, \quad |\xi - x_0| < |x - x_0|. \quad (17)$$

# Analytische Funktionen

Wenn also alle Ableitungen “nett”, ergo, “glatt” sind, dann kann man eine Konvergenz der Folge der Taylor-Polynome gegen die Funktion in einer Umgebung erwarten.

# Analytische Funktionen

Wenn also alle Ableitungen “nett”, ergo, “glatt” sind, dann kann man eine Konvergenz der Folge der Taylor-Polynome gegen die Funktion in einer Umgebung erwarten.

Leider gilt im Allgemeinen nicht für jede Funktion  $f \in C^\infty(D)$ , dass

$$f(x) = \lim_{n \rightarrow \infty} T_{x_0, n}(x). \quad (18)$$

# Analytische Funktionen

Wenn also alle Ableitungen “nett”, ergo, “glatt” sind, dann kann man eine Konvergenz der Folge der Taylor-Polynome gegen die Funktion in einer Umgebung erwarten.

Leider gilt im Allgemeinen nicht für jede Funktion  $f \in C^\infty(D)$ , dass

$$f(x) = \lim_{n \rightarrow \infty} T_{x_0, n}(x). \quad (18)$$

Dieses gilt auch nicht, wenn man die Konvergenz einschränkt auf eine kleinere Umgebung.

# Analytische Funktionen

Wenn also alle Ableitungen “nett”, ergo, “glatt” sind, dann kann man eine Konvergenz der Folge der Taylor-Polynome gegen die Funktion in einer Umgebung erwarten.

Leider gilt im Allgemeinen nicht für jede Funktion  $f \in C^\infty(D)$ , dass

$$f(x) = \lim_{n \rightarrow \infty} T_{x_0, n}(x). \quad (18)$$

Dieses gilt auch nicht, wenn man die Konvergenz einschränkt auf eine kleinere Umgebung.

Ein (klassisches) Gegenbeispiel ist die Funktion

$$g(x) := \begin{cases} 0, & x \leq 0, \\ \exp(-1/x^2), & \text{sonst.} \end{cases} \quad (19)$$

# Analytische Funktionen

Wenn also alle Ableitungen “nett”, ergo, “glatt” sind, dann kann man eine Konvergenz der Folge der Taylor-Polynome gegen die Funktion in einer Umgebung erwarten.

Leider gilt im Allgemeinen nicht für jede Funktion  $f \in C^\infty(D)$ , dass

$$f(x) = \lim_{n \rightarrow \infty} T_{x_0, n}(x). \quad (18)$$

Dieses gilt auch nicht, wenn man die Konvergenz einschränkt auf eine kleinere Umgebung.

Ein (klassisches) Gegenbeispiel ist die Funktion

$$g(x) := \begin{cases} 0, & x \leq 0, \\ \exp(-1/x^2), & \text{sonst.} \end{cases} \quad (19)$$

Allgemeiner kann man viele solcher Gegenbeispiele konstruieren, indem man Funktionen vom Typus  $\exp(-1/p(x))$  für Polynome  $p$  betrachtet.

# Analytische Funktionen

Wenn es für alle  $x_0 \in D$  eine Umgebung gibt, so dass die Taylorreihe

$$T_{x_0, \infty}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \quad (20)$$

gegen die Funktion  $f \in C^\infty(D)$  konvergiert, so nennt man die Funktion **analytisch**, in Zeichen,  $f \in C^\omega(D)$ .

# Analytische Funktionen

Wenn es für alle  $x_0 \in D$  eine Umgebung gibt, so dass die Taylorreihe

$$T_{x_0, \infty}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \quad (20)$$

gegen die Funktion  $f \in C^\infty(D)$  konvergiert, so nennt man die Funktion **analytisch**, in Zeichen,  $f \in C^\omega(D)$ .

Beispiele analytischer Funktionen sind Polynome, die trigonometrischen Funktionen, die Exponentialfunktion und der Logarithmus.

# Analytische Funktionen

Wenn es für alle  $x_0 \in D$  eine Umgebung gibt, so dass die Taylorreihe

$$T_{x_0, \infty}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \quad (20)$$

gegen die Funktion  $f \in C^\infty(D)$  konvergiert, so nennt man die Funktion **analytisch**, in Zeichen,  $f \in C^\omega(D)$ .

Beispiele analytischer Funktionen sind Polynome, die trigonometrischen Funktionen, die Exponentialfunktion und der Logarithmus.

Analytische Funktionen lassen sich also bequem analysieren, da sie in einer kleinen Umgebung mit einer konvergenten Potenzreihe übereinstimmen. Die Ableitungen in einem Punkt bestimmen lokal das Verhalten der Funktion.

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# “Das” Newton-Verfahren?

Es gibt mehrere Verfahren, welche als “das” Newton-Verfahren bekannt sind. Die eigentliche Variante dient der Berechnung von Nullstellen und wurde an einem Beispiel von Newton erdacht und von Raphson auf allgemeinere Polynome erweitert (daher auch “Newton-Raphson-Verfahren” genannt).

# “Das” Newton-Verfahren?

Es gibt mehrere Verfahren, welche als “das” Newton-Verfahren bekannt sind. Die eigentliche Variante dient der Berechnung von Nullstellen und wurde an einem Beispiel von Newton erdacht und von Raphson auf allgemeinere Polynome erweitert (daher auch “Newton-Raphson-Verfahren” genannt).

Da ein Minimum unter gewissen Umständen eine Nullstelle der Ableitung ist, kann man das obengenannte Newton-Verfahren zur Berechnung von Nullstellen der ersten Ableitung und somit indirekt zur Berechnung von Minima heranziehen.

# “Das” Newton-Verfahren?

Es gibt mehrere Verfahren, welche als “das” Newton-Verfahren bekannt sind. Die eigentliche Variante dient der Berechnung von Nullstellen und wurde an einem Beispiel von Newton erdacht und von Raphson auf allgemeinere Polynome erweitert (daher auch “Newton-Raphson-Verfahren” genannt).

Da ein Minimum unter gewissen Umständen eine Nullstelle der Ableitung ist, kann man das obengenannte Newton-Verfahren zur Berechnung von Nullstellen der ersten Ableitung und somit indirekt zur Berechnung von Minima heranziehen.

Dann gibt es immer noch viele verschiedene Newton-Verfahren zur Minimierung, nämlich für viele verschiedene Funktionentypen  $f : D \subset \mathbb{M} \rightarrow \mathbb{R}$ , abhängig von der jeweiligen Menge  $\mathbb{M}$ . Die Teilmenge  $D$  sei hierbei immer offen.

# “Das” Newton-Verfahren?

Es gibt mehrere Verfahren, welche als “das” Newton-Verfahren bekannt sind. Die eigentliche Variante dient der Berechnung von Nullstellen und wurde an einem Beispiel von Newton erdacht und von Raphson auf allgemeinere Polynome erweitert (daher auch “Newton-Raphson-Verfahren” genannt).

Da ein Minimum unter gewissen Umständen eine Nullstelle der Ableitung ist, kann man das obengenannte Newton-Verfahren zur Berechnung von Nullstellen der ersten Ableitung und somit indirekt zur Berechnung von Minima heranziehen.

Dann gibt es immer noch viele verschiedene Newton-Verfahren zur Minimierung, nämlich für viele verschiedene Funktionentypen  $f : D \subset \mathbb{M} \rightarrow \mathbb{R}$ , abhängig von der jeweiligen Menge  $\mathbb{M}$ . Die Teilmenge  $D$  sei hierbei immer offen.

Wir starten mit dem wahrscheinlich einfachsten Fall:  $\mathbb{M} = \mathbb{R}$ .

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Das Newton-Verfahren zur Nullstellensuche

Ist eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  einmal (stetig) differenzierbar, so kann man die Funktion in der Nähe eines Punktes  $x_0$  durch das Taylor-Polynom vom Grade Eins approximieren:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \quad (21)$$

# Das Newton-Verfahren zur Nullstellensuche

Ist eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  einmal (stetig) differenzierbar, so kann man die Funktion in der Nähe eines Punktes  $x_0$  durch das Taylor-Polynom vom Grade Eins approximieren:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \quad (21)$$

Statt eine Nullstelle von  $f(x)$  zu suchen, suchen wir eine Nullstelle der linearen Approximation:

$$\tilde{x} := x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (22)$$

# Das Newton-Verfahren zur Nullstellensuche

Ist eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  einmal (stetig) differenzierbar, so kann man die Funktion in der Nähe eines Punktes  $x_0$  durch das Taylor-Polynom vom Grade Eins approximieren:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \quad (21)$$

Statt eine Nullstelle von  $f(x)$  zu suchen, suchen wir eine Nullstelle der linearen Approximation:

$$\tilde{x} := x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (22)$$

Jetzt werten wir  $f$  an der Stelle  $\tilde{x}$  aus. Wenn  $f(\tilde{x}) = 0$  gilt, oder zumindest  $f(\tilde{x}) \approx 0$ , brechen wir ab.

# Das Newton-Verfahren zur Nullstellensuche

Ist eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  einmal (stetig) differenzierbar, so kann man die Funktion in der Nähe eines Punktes  $x_0$  durch das Taylor-Polynom vom Grade Eins approximieren:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \quad (21)$$

Statt eine Nullstelle von  $f(x)$  zu suchen, suchen wir eine Nullstelle der linearen Approximation:

$$\tilde{x} := x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (22)$$

Jetzt werten wir  $f$  an der Stelle  $\tilde{x}$  aus. Wenn  $f(\tilde{x}) = 0$  gilt, oder zumindest  $f(\tilde{x}) \approx 0$ , brechen wir ab. Wenn nicht, so ist vielleicht  $x_1 = \tilde{x}$  näher an einer (hypothetischen) Nullstelle.

# Das Newton-Verfahren zur Nullstellensuche

Ist eine Funktion  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  einmal (stetig) differenzierbar, so kann man die Funktion in der Nähe eines Punktes  $x_0$  durch das Taylor-Polynom vom Grade Eins approximieren:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \quad (21)$$

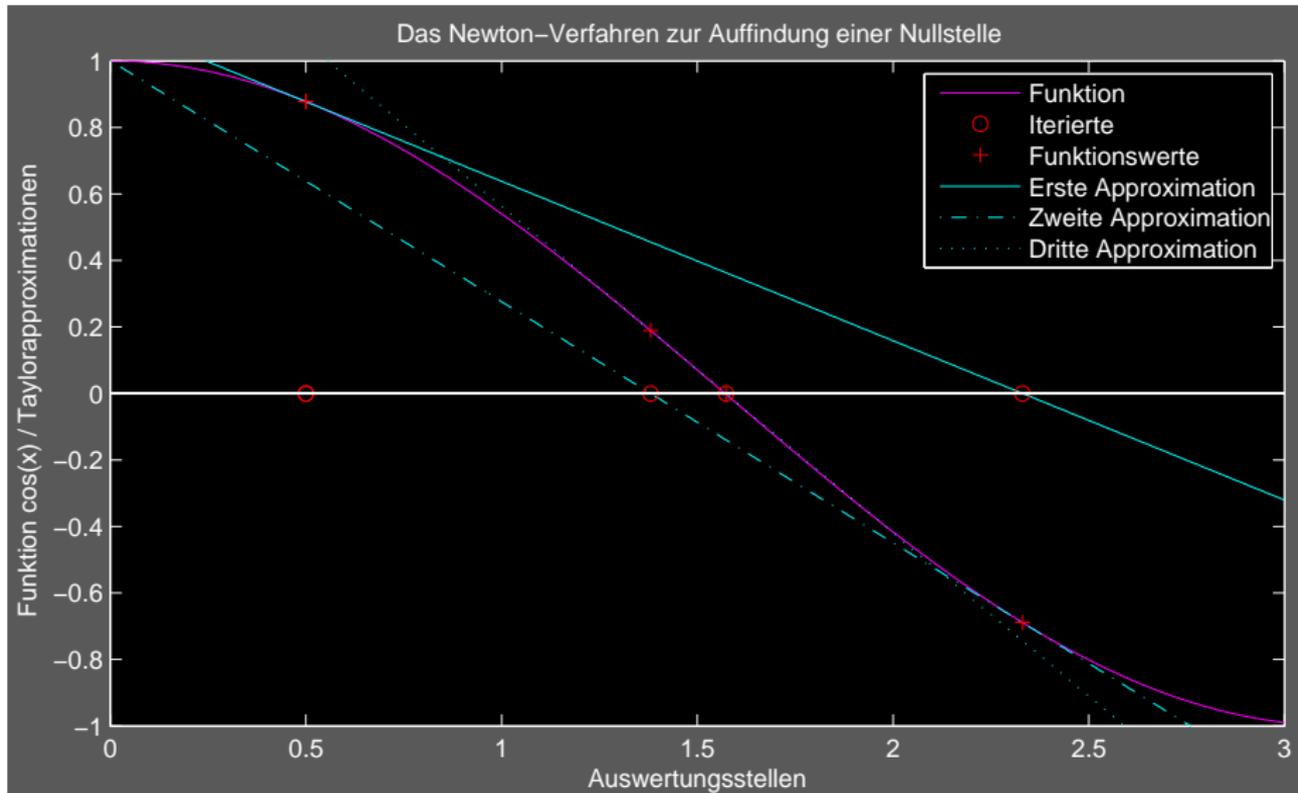
Statt eine Nullstelle von  $f(x)$  zu suchen, suchen wir eine Nullstelle der linearen Approximation:

$$\tilde{x} := x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (22)$$

Jetzt werten wir  $f$  an der Stelle  $\tilde{x}$  aus. Wenn  $f(\tilde{x}) = 0$  gilt, oder zumindest  $f(\tilde{x}) \approx 0$ , brechen wir ab. Wenn nicht, so ist vielleicht  $x_1 = \tilde{x}$  näher an einer (hypothetischen) Nullstelle. Insgesamt iterieren wir gemäß

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \quad (23)$$

# Zwei Beispiele: 1. Kleinste Nullstelle des Kosinus



# Das Newton-Verfahren zur Minimierung

Das Newton-Verfahren zur Minimierung einer (mindestens zweimal differenzierbaren) Funktion ersetzt die Minimierungsaufgabe durch die allgemeinere Aufgabe des Auffindens stationärer Punkte

$$f'(x) = 0. \quad (24)$$

# Das Newton-Verfahren zur Minimierung

Das Newton-Verfahren zur Minimierung einer (mindestens zweimal differenzierbaren) Funktion ersetzt die Minimierungsaufgabe durch die allgemeinere Aufgabe des Auffindens stationärer Punkte

$$f'(x) = 0. \quad (24)$$

Die gegebene Funktion wird also durch das Taylorpolynom zweiter Ordnung approximiert; gesucht wird der stationäre Punkt der dadurch entstehenden quadratischen Funktion  $T_{x_0,2}$ .

# Das Newton-Verfahren zur Minimierung

Das Newton-Verfahren zur Minimierung einer (mindestens zweimal differenzierbaren) Funktion ersetzt die Minimierungsaufgabe durch die allgemeinere Aufgabe des Auffindens stationärer Punkte

$$f'(x) = 0. \quad (24)$$

Die gegebene Funktion wird also durch das Taylorpolynom zweiter Ordnung approximiert; gesucht wird der stationäre Punkt der dadurch entstehenden quadratischen Funktion  $T_{x_0,2}$ .

An dieser Stelle wird der stationäre Punkt der nächsten lokalen quadratischen Approximation durch das Taylorpolynom vom Grade zwei berechnet.

# Das Newton-Verfahren zur Minimierung

Das Newton-Verfahren zur Minimierung einer (mindestens zweimal differenzierbaren) Funktion ersetzt die Minimierungsaufgabe durch die allgemeinere Aufgabe des Auffindens stationärer Punkte

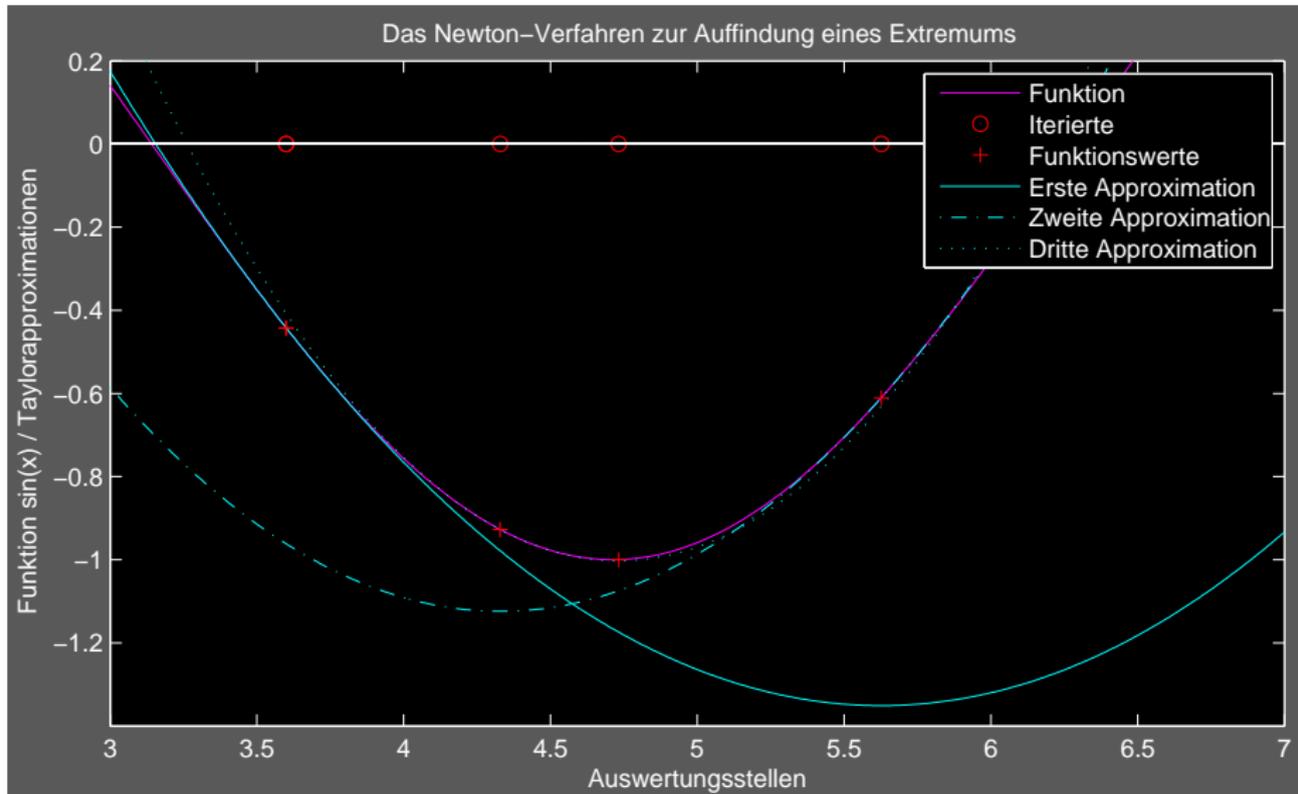
$$f'(x) = 0. \quad (24)$$

Die gegebene Funktion wird also durch das Taylorpolynom zweiter Ordnung approximiert; gesucht wird der stationäre Punkt der dadurch entstehenden quadratischen Funktion  $T_{x_0,2}$ .

An dieser Stelle wird der stationäre Punkt der nächsten lokalen quadratischen Approximation durch das Taylorpolynom vom Grade zwei berechnet, allgemein also

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}. \quad (25)$$

# Zwei Beispiele: 2. Kleinstes Minimum des Sinus



# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

## Theorem (Hinreichende Klassifizierung der Extrema)

*Ein stationärer Punkt ist ein Minimum, wenn die zweite Ableitung im stationären Punkt größer als Null ist, und ein Maximum, wenn die zweite Ableitung kleiner als Null ist.*

# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

## Theorem (Hinreichende Klassifizierung der Extrema)

*Ein stationärer Punkt ist ein Minimum, wenn die zweite Ableitung im stationären Punkt größer als Null ist, und ein Maximum, wenn die zweite Ableitung kleiner als Null ist.*

Wenn die zweite Ableitung gleich Null ist, müssen weitere Ableitungen berechnet werden. Beispiele hierfür sind die Funktionen

# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

## Theorem (Hinreichende Klassifizierung der Extrema)

*Ein stationärer Punkt ist ein Minimum, wenn die zweite Ableitung im stationären Punkt größer als Null ist, und ein Maximum, wenn die zweite Ableitung kleiner als Null ist.*

Wenn die zweite Ableitung gleich Null ist, müssen weitere Ableitungen berechnet werden. Beispiele hierfür sind die Funktionen

- ▶  $f(x) = x^3$ :  $f$  hat in  $\hat{x} = 0$  einen **Sattelpunkt**,

# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

## Theorem (Hinreichende Klassifizierung der Extrema)

*Ein stationärer Punkt ist ein Minimum, wenn die zweite Ableitung im stationären Punkt größer als Null ist, und ein Maximum, wenn die zweite Ableitung kleiner als Null ist.*

Wenn die zweite Ableitung gleich Null ist, müssen weitere Ableitungen berechnet werden. Beispiele hierfür sind die Funktionen

- ▶  $f(x) = x^3$ :  $f$  hat in  $\hat{x} = 0$  einen **Sattelpunkt**,
- ▶  $f(x) = x^4$ :  $f$  hat in  $\hat{x} = 0$  ein **Minimum**,

# Das Newton-Verfahren zur Minimierung

Da das Newton-Verfahren angewendet auf die erste Ableitung (bei Erfolg) “nur” **stationäre Punkte** berechnet, müssen wir nachher noch bestimmen, ob wirklich ein **Minimum** bestimmt wurde. Dazu helfen die folgenden Aussagen.

## Theorem (Hinreichende Klassifizierung der Extrema)

*Ein stationärer Punkt ist ein Minimum, wenn die zweite Ableitung im stationären Punkt größer als Null ist, und ein Maximum, wenn die zweite Ableitung kleiner als Null ist.*

Wenn die zweite Ableitung gleich Null ist, müssen weitere Ableitungen berechnet werden. Beispiele hierfür sind die Funktionen

- ▶  $f(x) = x^3$ :  $f$  hat in  $\hat{x} = 0$  einen **Sattelpunkt**,
- ▶  $f(x) = x^4$ :  $f$  hat in  $\hat{x} = 0$  ein **Minimum**,
- ▶  $f(x) = -x^4$ :  $f$  hat in  $\hat{x} = 0$  ein **Maximum**.

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

Theorem (Klassifizierung stationärer Punkte)

*Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$*

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

## Theorem (Klassifizierung stationärer Punkte)

Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$

- ▶ ein Minimum, wenn  $f^{(k)}(\hat{x}) > 0$  und

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

## Theorem (Klassifizierung stationärer Punkte)

Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$

- ▶ ein Minimum, wenn  $f^{(k)}(\hat{x}) > 0$  und
- ▶ ein Maximum, wenn  $f^{(k)}(\hat{x}) < 0$ .

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

## Theorem (Klassifizierung stationärer Punkte)

Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$

- ▶ ein Minimum, wenn  $f^{(k)}(\hat{x}) > 0$  und
- ▶ ein Maximum, wenn  $f^{(k)}(\hat{x}) < 0$ .

Ist  $k$  ungerade, so hat  $f$  bei  $\hat{x}$  einen Sattelpunkt.

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

## Theorem (Klassifizierung stationärer Punkte)

Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$

- ▶ ein Minimum, wenn  $f^{(k)}(\hat{x}) > 0$  und
- ▶ ein Maximum, wenn  $f^{(k)}(\hat{x}) < 0$ .

Ist  $k$  ungerade, so hat  $f$  bei  $\hat{x}$  einen Sattelpunkt.

Diese Klassifizierung hilft auch nur, sofern die Funktion hinreichend oft differenzierbar ist und Ableitungen ungleich Null besitzt, oder die Funktion analytisch ist.

# Das Newton-Verfahren zur Minimierung

Allgemein gilt (für hinreichend oft differenzierbare Funktionen):

## Theorem (Klassifizierung stationärer Punkte)

Seien die Ableitungen  $f^{(j)}(\hat{x})$ ,  $1 \leq j < k$ , gleich Null. Ist  $k$  gerade, so hat  $f$  bei  $\hat{x}$

- ▶ ein Minimum, wenn  $f^{(k)}(\hat{x}) > 0$  und
- ▶ ein Maximum, wenn  $f^{(k)}(\hat{x}) < 0$ .

Ist  $k$  ungerade, so hat  $f$  bei  $\hat{x}$  einen Sattelpunkt.

Diese Klassifizierung hilft auch nur, sofern die Funktion hinreichend oft differenzierbar ist und Ableitungen ungleich Null besitzt, oder die Funktion analytisch ist.

Ist die Funktion nicht genügend oft differenzierbar, so greift man auf das Verhalten der ersten Ableitung in der Umgebung des stationären Punktes zurück.

# Quadratische Konvergenz

Wenn man nah genug an einer Nullstelle der Funktion dran ist, sieht sie, wenn sie “glatt” genug ist, “oft” aus wie eine lineare Funktion. Ab dieser Stelle kann man schnelle Konvergenz erwarten.

# Quadratische Konvergenz

Wenn man nah genug an einer Nullstelle der Funktion dran ist, sieht sie, wenn sie “glatt” genug ist, “oft” aus wie eine lineare Funktion. Ab dieser Stelle kann man schnelle Konvergenz erwarten.

Die quadratische Konvergenz des Newton-Verfahrens zur Nullstellensuche wird im folgenden Satz beschrieben.

# Quadratische Konvergenz

Wenn man nah genug an einer Nullstelle der Funktion dran ist, sieht sie, wenn sie “glatt” genug ist, “oft” aus wie eine lineare Funktion. Ab dieser Stelle kann man schnelle Konvergenz erwarten.

Die quadratische Konvergenz des Newton-Verfahrens zur Nullstellensuche wird im folgenden Satz beschrieben:

## Theorem (Quadratische Konvergenz)

*Sei  $f : I = [a, b] \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und  $\hat{x}$  eine einfache Nullstelle von  $f$  im Innern von  $I$ . Dann gibt es ein  $r > 0$  und eine Konstante  $C \geq 0$ , so dass das Newton Verfahren für alle Startwerte  $x_0 \in I$  mit  $|\hat{x} - x_0| \leq r$  gegen  $\hat{x}$  quadratisch konvergiert,*

$$|\hat{x} - x_{k+1}| \leq C|\hat{x} - x_k|^2. \quad (26)$$

# Quadratische Konvergenz

Wenn man nah genug an einer Nullstelle der Funktion dran ist, sieht sie, wenn sie “glatt” genug ist, “oft” aus wie eine lineare Funktion. Ab dieser Stelle kann man schnelle Konvergenz erwarten.

Die quadratische Konvergenz des Newton-Verfahrens zur Nullstellensuche wird im folgenden Satz beschrieben:

## Theorem (Quadratische Konvergenz)

*Sei  $f : I = [a, b] \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und  $\hat{x}$  eine einfache Nullstelle von  $f$  im Innern von  $I$ . Dann gibt es ein  $r > 0$  und eine Konstante  $C \geq 0$ , so dass das Newton Verfahren für alle Startwerte  $x_0 \in I$  mit  $|\hat{x} - x_0| \leq r$  gegen  $\hat{x}$  quadratisch konvergiert,*

$$|\hat{x} - x_{k+1}| \leq C|\hat{x} - x_k|^2. \quad (26)$$

Da  $f''(\hat{x})$  ungleich Null ist, ist die Konstante  $C > 0$ .

# Konvergenzbereiche

Das Newton-Verfahren konvergiert leider nur lokal gegen eine Nullstelle. Es gibt Startwerte, von denen aus keine Nullstelle berechnet werden kann, und falls die Nullstelle nicht einfach ist, konvergiert es zwar noch in einer geeigneten Umgebung, aber dann nur noch linear.

# Konvergenzbereiche

Das Newton-Verfahren konvergiert leider nur lokal gegen eine Nullstelle. Es gibt Startwerte, von denen aus keine Nullstelle berechnet werden kann, und falls die Nullstelle nicht einfach ist, konvergiert es zwar noch in einer geeigneten Umgebung, aber dann nur noch linear.

Zum Beispiel konvergiert das Newton-Verfahren mit dem Startwert  $x_0 = 1$  gegen die (doppelte) Nullstelle Null der Funktion  $f(x) = x^2$  nur linear, da

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2}{2x_k} = \frac{1}{2}x_k = \left(\frac{1}{2}\right)^{k+1}. \quad (27)$$

# Konvergenzbereiche

Das Newton-Verfahren konvergiert leider nur lokal gegen eine Nullstelle. Es gibt Startwerte, von denen aus keine Nullstelle berechnet werden kann, und falls die Nullstelle nicht einfach ist, konvergiert es zwar noch in einer geeigneten Umgebung, aber dann nur noch linear.

Zum Beispiel konvergiert das Newton-Verfahren mit dem Startwert  $x_0 = 1$  gegen die (doppelte) Nullstelle Null der Funktion  $f(x) = x^2$  nur linear, da

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2}{2x_k} = \frac{1}{2}x_k = \left(\frac{1}{2}\right)^{k+1}. \quad (27)$$

Wenn man die Funktion genau genug kennt, so kann man vorhersagen in welchem Bereich das Newton-Verfahren quadratisch konvergiert.

# Konvergenzbereiche

Das Newton-Verfahren konvergiert leider nur lokal gegen eine Nullstelle. Es gibt Startwerte, von denen aus keine Nullstelle berechnet werden kann, und falls die Nullstelle nicht einfach ist, konvergiert es zwar noch in einer geeigneten Umgebung, aber dann nur noch linear.

Zum Beispiel konvergiert das Newton-Verfahren mit dem Startwert  $x_0 = 1$  gegen die (doppelte) Nullstelle Null der Funktion  $f(x) = x^2$  nur linear, da

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^2}{2x_k} = \frac{1}{2}x_k = \left(\frac{1}{2}\right)^{k+1}. \quad (27)$$

Wenn man die Funktion genau genug kennt, so kann man vorhersagen in welchem Bereich das Newton-Verfahren quadratisch konvergiert.

Dieses kann mittels des Satzes von Kantorovich oder Smales sogenannten Punktabstätzungen (dem Alpha-Test) erfolgen. Beides funktioniert auch höherdimensional und wird später besprochen.

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Komplexe Funktionen

Wir könnten auch auf die Idee kommen, Funktionen zu betrachten, die komplexe Eingabeargumente haben. **Nullstellen** kann man auch suchen, wenn die Funktionswerte **komplex** sind, aber da wir an der **Minimierung** interessiert sind, sollten die Funktionswerte **reell** sein.

# Komplexe Funktionen

Wir könnten auch auf die Idee kommen, Funktionen zu betrachten, die komplexe Eingabeargumente haben. **Nullstellen** kann man auch suchen, wenn die Funktionswerte **komplex** sind, aber da wir an der **Minimierung** interessiert sind, sollten die Funktionswerte **reell** sein.

Es gibt ein paar (noch zu lernende) Resultate über Funktionen von  $\mathbb{C}$  nach  $\mathbb{C}$ . Ein Resultat beruht auf der Definition der komplexen Differenzierbarkeit analog zur reellen Differenzierbarkeit:

# Komplexe Funktionen

Wir könnten auch auf die Idee kommen, Funktionen zu betrachten, die komplexe Eingabeargumente haben. **Nullstellen** kann man auch suchen, wenn die Funktionswerte **komplex** sind, aber da wir an der **Minimierung** interessiert sind, sollten die Funktionswerte **reell** sein.

Es gibt ein paar (noch zu lernende) Resultate über Funktionen von  $\mathbb{C}$  nach  $\mathbb{C}$ . Ein Resultat beruht auf der Definition der komplexen Differenzierbarkeit analog zur reellen Differenzierbarkeit:

## Definition (Differenzierbarkeit komplexer Funktionen)

Eine Funktion  $f : D \subset \mathbb{C} \rightarrow \mathbb{C}$  ist im Punkt  $z \in D$  differenzierbar, wenn

$$\frac{df}{dz} := f'(z) := \lim_{\tilde{z} \rightarrow z} \frac{f(\tilde{z}) - f(z)}{\tilde{z} - z} \quad (28)$$

existiert. Die Funktion heißt differenzierbar in  $D$ , wenn sie in allen Punkten  $z \in D$  differenzierbar ist.

# Komplexe Funktionen

Ist eine Funktion einmal komplex differenzierbar in einer Umgebung, so ist sie unendlich oft differenzierbar in einer Umgebung und stimmt mit ihrer Potenzreihe überein, ist also (komplex) analytisch.

# Komplexe Funktionen

Ist eine Funktion einmal komplex differenzierbar in einer Umgebung, so ist sie unendlich oft differenzierbar in einer Umgebung und stimmt mit ihrer Potenzreihe überein, ist also (komplex) analytisch.

Da für eine Umgebung eines Punktes alle diese Begriffe kollabieren, verwendet man die Bezeichnung “**Holomorphie**” für **komplexe Differenzierbarkeit auf einer Umgebung**.

# Komplexe Funktionen

Ist eine Funktion einmal komplex differenzierbar in einer Umgebung, so ist sie unendlich oft differenzierbar in einer Umgebung und stimmt mit ihrer Potenzreihe überein, ist also (komplex) analytisch.

Da für eine Umgebung eines Punktes alle diese Begriffe kollabieren, verwendet man die Bezeichnung “**Holomorphie**” für **komplexe Differenzierbarkeit auf einer Umgebung**.

Wir betrachten jetzt Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{C}$  als reelle Funktionen  $f : \mathcal{D} \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , indem wir

$$z = x + iy, \quad x, y \in \mathbb{R} \quad \text{und} \quad f(x, y) = u(x, y) + iv(x, y) \quad (29)$$

setzen, wobei

$$u(x, y), v(x, y) \in \mathbb{R} \quad \forall x, y \in \mathbb{R} : z = x + iy \in D. \quad (30)$$

$x$  und  $y$  sind der Real- und Imaginärteil der komplexen **Variablen**  $z$ ,  $u$  und  $v$  der Real- und Imaginärteil der komplexen **Funktion**  $f$ .

# Komplexe Funktionen

Die Funktionen “Real- und Imaginärteil der Funktion  $f$ ” fasst man dann zusammen als die (vektorwertige reelle) Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (31)$$

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}. \quad (32)$$

# Komplexe Funktionen

Die Funktionen “Real- und Imaginärteil der Funktion  $f$ ” fasst man dann zusammen als die (vektorwertige reelle) Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (31)$$

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}. \quad (32)$$

Aufgefasst als vektorwertige reelle Funktionen können wir die komplexe Funktion auch nach den reellen Variablen  $x$  und  $y$  partiell ableiten und auch Kombinationen durch Summation bilden.

# Komplexe Funktionen

Die Funktionen “Real- und Imaginärteil der Funktion  $f$ ” fasst man dann zusammen als die (vektorwertige reelle) Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad (31)$$

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}. \quad (32)$$

Aufgefasst als vektorwertige reelle Funktionen können wir die komplexe Funktion auch nach den reellen Variablen  $x$  und  $y$  partiell ableiten und auch Kombinationen durch Summation bilden.

Wenn wir nur Funktionen betrachten, deren Funktionswerte reell sind, so ist der Imaginärteil von  $f$  identisch gleich Null. Also können wir im Fall solcher Funktionen die Aufgabe interpretieren als die Optimierung von Funktionen

$$f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}. \quad (33)$$

# Komplexe Funktionen

Man stellt schnell fest, dass die komplexe Ableitung sich als gewichtete Summe partieller Ableitungen schreiben läßt, es gilt

$$\frac{df}{dz} = \frac{\partial f}{\partial z}, \quad (34)$$

# Komplexe Funktionen

Man stellt schnell fest, dass die komplexe Ableitung sich als gewichtete Summe partieller Ableitungen schreiben läßt, es gilt

$$\frac{df}{dz} = \frac{\partial f}{\partial z}, \quad (34)$$

wobei die **Wirtinger-Ableitung**  $\partial f / \partial z$  nach  $z$  definiert ist als

$$\frac{\partial f}{\partial z} := \frac{1}{2} \left( \frac{\partial f}{\partial x} - i \frac{\partial f}{\partial y} \right), \quad (35)$$

# Komplexe Funktionen

Man stellt schnell fest, dass die komplexe Ableitung sich als gewichtete Summe partieller Ableitungen schreiben lässt, es gilt

$$\frac{df}{dz} = \frac{\partial f}{\partial z}, \quad (34)$$

wobei die **Wirtinger-Ableitung**  $\partial f / \partial z$  nach  $z$  definiert ist als

$$\frac{\partial f}{\partial z} := \frac{1}{2} \left( \frac{\partial f}{\partial x} - i \frac{\partial f}{\partial y} \right), \quad (35)$$

wobei

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \quad (36)$$

# Komplexe Funktionen

Man stellt schnell fest, dass die komplexe Ableitung sich als gewichtete Summe partieller Ableitungen schreiben läßt, es gilt

$$\frac{df}{dz} = \frac{\partial f}{\partial z}, \quad (34)$$

wobei die **Wirtinger-Ableitung**  $\partial f / \partial z$  nach  $z$  definiert ist als

$$\frac{\partial f}{\partial z} := \frac{1}{2} \left( \frac{\partial f}{\partial x} - i \frac{\partial f}{\partial y} \right), \quad (35)$$

wobei

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \quad (36)$$

und

$$\frac{\partial f}{\partial y} = \frac{\partial u}{\partial y} + i \frac{\partial v}{\partial y}. \quad (37)$$

# Komplexe Funktionen

Analog definiert man die **Wirtinger-Ableitung**  $\partial f / \partial \bar{z}$  nach  $\bar{z}$  als

$$\frac{\partial f}{\partial \bar{z}} := \frac{1}{2} \left( \frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right). \quad (38)$$

# Komplexe Funktionen

Analog definiert man die **Wirtinger-Ableitung**  $\partial f / \partial \bar{z}$  nach  $\bar{z}$  als

$$\frac{\partial f}{\partial \bar{z}} := \frac{1}{2} \left( \frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right). \quad (38)$$

Eine Funktion ist exakt dann (komplex) differenzierbar, wenn  $\partial f / \partial \bar{z}$  gleich Null ist.

# Komplexe Funktionen

Analog definiert man die **Wirtinger-Ableitung**  $\partial f / \partial \bar{z}$  nach  $\bar{z}$  als

$$\frac{\partial f}{\partial \bar{z}} := \frac{1}{2} \left( \frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right). \quad (38)$$

Eine Funktion ist exakt dann (komplex) differenzierbar, wenn  $\partial f / \partial \bar{z}$  gleich Null ist. Ausgeschrieben in reeller Form erhält man die sogenannten **Cauchy-Riemannschen Differentialgleichungen**:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \quad (39)$$

# Komplexe Funktionen

Analog definiert man die **Wirtinger-Ableitung**  $\partial f / \partial \bar{z}$  nach  $\bar{z}$  als

$$\frac{\partial f}{\partial \bar{z}} := \frac{1}{2} \left( \frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right). \quad (38)$$

Eine Funktion ist exakt dann (komplex) differenzierbar, wenn  $\partial f / \partial \bar{z}$  gleich Null ist. Ausgeschrieben in reeller Form erhält man die sogenannten

**Cauchy-Riemannschen Differentialgleichungen:**

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \quad (39)$$

Man sieht sofort, dass

$$\overline{\frac{\partial f}{\partial \bar{z}}} = \frac{\partial \bar{f}}{\partial z} \quad (40)$$

gilt, und damit im Falle  $f = \bar{f}$  die Wirtinger-Ableitungen zueinander komplex konjugiert sind.

# Komplexe Funktionen

Das Newton-Verfahren basiert auf Ableitungen. Die reellen Zahlen sind in die komplexen Zahlen eingebettet. Also könnte man auf die Idee kommen, auf  $\mathbb{C}$  holomorphe Funktionen zu betrachten, die als Funktionswerte nur reelle Zahlen haben und die komplexe Differenzierbarkeit zu verwenden.

# Komplexe Funktionen

Das Newton-Verfahren basiert auf Ableitungen. Die reellen Zahlen sind in die komplexen Zahlen eingebettet. Also könnte man auf die Idee kommen, auf  $\mathbb{C}$  holomorphe Funktionen zu betrachten, die als Funktionswerte nur reelle Zahlen haben und die komplexe Differenzierbarkeit zu verwenden.

Leider folgt aus der Holomorphie von  $f$

$$\frac{\partial f}{\partial \bar{z}} = 0, \quad (41)$$

# Komplexe Funktionen

Das Newton-Verfahren basiert auf Ableitungen. Die reellen Zahlen sind in die komplexen Zahlen eingebettet. Also könnte man auf die Idee kommen, auf  $\mathbb{C}$  holomorphe Funktionen zu betrachten, die als Funktionswerte nur reelle Zahlen haben und die komplexe Differenzierbarkeit zu verwenden.

Leider folgt aus der Holomorphie von  $f$

$$\frac{\partial f}{\partial \bar{z}} = 0, \quad (41)$$

und aus  $f = \bar{f}$  mit der Beziehung zwischen den beiden Wirtinger-Ableitungen sofort

$$\frac{\partial f}{\partial z} = \frac{\partial \bar{f}}{\partial z} = \overline{\frac{\partial f}{\partial \bar{z}}} = \bar{0} = 0. \quad (42)$$

# Komplexe Funktionen

Das Newton-Verfahren basiert auf Ableitungen. Die reellen Zahlen sind in die komplexen Zahlen eingebettet. Also könnte man auf die Idee kommen, auf  $\mathbb{C}$  holomorphe Funktionen zu betrachten, die als Funktionswerte nur reelle Zahlen haben und die komplexe Differenzierbarkeit zu verwenden.

Leider folgt aus der Holomorphie von  $f$

$$\frac{\partial f}{\partial \bar{z}} = 0, \quad (41)$$

und aus  $f = \bar{f}$  mit der Beziehung zwischen den beiden Wirtinger-Ableitungen sofort

$$\frac{\partial f}{\partial z} = \frac{\partial \bar{f}}{\partial z} = \overline{\frac{\partial f}{\partial \bar{z}}} = \bar{0} = 0. \quad (42)$$

Alle holomorphen Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{C}$ , welche nur in die reellen Zahlen abbilden, sind also konstant,  $f(z) = c \in \mathbb{R}$ .

# Komplexe Funktionen

Konstante Funktionen haben keine echten Minima und sind überhaupt vom mathematischen Aspekt her ziemlich langweilig.

# Komplexe Funktionen

Konstante Funktionen haben keine echten Minima und sind überhaupt vom mathematischen Aspekt her ziemlich langweilig.

Daher werden wir nur noch Funktionen mit komplexen Argumenten betrachten, die in die reellen Zahlen abbilden und reell-analytisch sind, und diese dann als Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  betrachten.

# Komplexe Funktionen

Konstante Funktionen haben keine echten Minima und sind überhaupt vom mathematischen Aspekt her ziemlich langweilig.

Daher werden wir nur noch Funktionen mit komplexen Argumenten betrachten, die in die reellen Zahlen abbilden und reell-analytisch sind, und diese dann als Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  betrachten.

Die Frage, die damit noch bleibt: Was ist nun mit der Optimierung von Funktionen vom  $\mathbb{R}^2$  nach  $\mathbb{R}$ ?

# Komplexe Funktionen

Konstante Funktionen haben keine echten Minima und sind überhaupt vom mathematischen Aspekt her ziemlich langweilig.

Daher werden wir nur noch Funktionen mit komplexen Argumenten betrachten, die in die reellen Zahlen abbilden und reell-analytisch sind, und diese dann als Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  betrachten.

Die Frage, die damit noch bleibt: Was ist nun mit der Optimierung von Funktionen vom  $\mathbb{R}^2$  nach  $\mathbb{R}$ ?

Zum Beispiel hat die Funktion

$$f(x, y) = 3 \cdot x^2 + 5 \cdot y^2 + 1 \quad (43)$$

im  $\mathbb{R}^2$  das eindeutige Minimum  $(x, y) = (0, 0)$  mit dem Funktionswert 1.

# Komplexe Funktionen

Konstante Funktionen haben keine echten Minima und sind überhaupt vom mathematischen Aspekt her ziemlich langweilig.

Daher werden wir nur noch Funktionen mit komplexen Argumenten betrachten, die in die reellen Zahlen abbilden und reell-analytisch sind, und diese dann als Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  betrachten.

Die Frage, die damit noch bleibt: Was ist nun mit der Optimierung von Funktionen vom  $\mathbb{R}^2$  nach  $\mathbb{R}$ ?

Zum Beispiel hat die Funktion

$$f(x, y) = 3 \cdot x^2 + 5 \cdot y^2 + 1 \quad (43)$$

im  $\mathbb{R}^2$  das eindeutige Minimum  $(x, y) = (0, 0)$  mit dem Funktionswert 1.

Wie läßt sich dieses Minimum mittels eines Newton-Verfahrens von einem Startwert ausgehend berechnen?

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Vektorwertige reelle Funktionen

Wir haben bereits gesehen, dass die Optimierung von (zweifach reell differenzierbaren oder reell-analytischen) Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{R}$  äquivalent ist zur Optimierung von Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ .

# Vektorwertige reelle Funktionen

Wir haben bereits gesehen, dass die Optimierung von (zweifach reell differenzierbaren oder reell-analytischen) Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{R}$  äquivalent ist zur Optimierung von Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ .

Wir fragen gleich allgemeiner: Wie optimieren wir Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ?

# Vektorwertige reelle Funktionen

Wir haben bereits gesehen, dass die Optimierung von (zweifach reell differenzierbaren oder reell-analytischen) Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{R}$  äquivalent ist zur Optimierung von Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ .

Wir fragen gleich allgemeiner: Wie optimieren wir Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ?

Um diese Frage beantworten zu können, müssen wir die Begriffe der (stetigen) Differenzierbarkeit und (lokalen) quadratischen Approximation einführen.

# Vektorwertige reelle Funktionen

Wir haben bereits gesehen, dass die Optimierung von (zweifach reell differenzierbaren oder reell-analytischen) Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{R}$  äquivalent ist zur Optimierung von Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ .

Wir fragen gleich allgemeiner: Wie optimieren wir Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ?

Um diese Frage beantworten zu können, müssen wir die Begriffe der (stetigen) Differenzierbarkeit und (lokalen) quadratischen Approximation einführen.

Dazu schauen wir uns als Beispiel die Funktion

$$f(x) = x_1^2 + x_2^2 \quad (44)$$

an.

# Vektorwertige reelle Funktionen

Wir haben bereits gesehen, dass die Optimierung von (zweifach reell differenzierbaren oder reell-analytischen) Funktionen  $f : D \subset \mathbb{C} \rightarrow \mathbb{R}$  äquivalent ist zur Optimierung von Funktionen  $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ .

Wir fragen gleich allgemeiner: Wie optimieren wir Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ ?

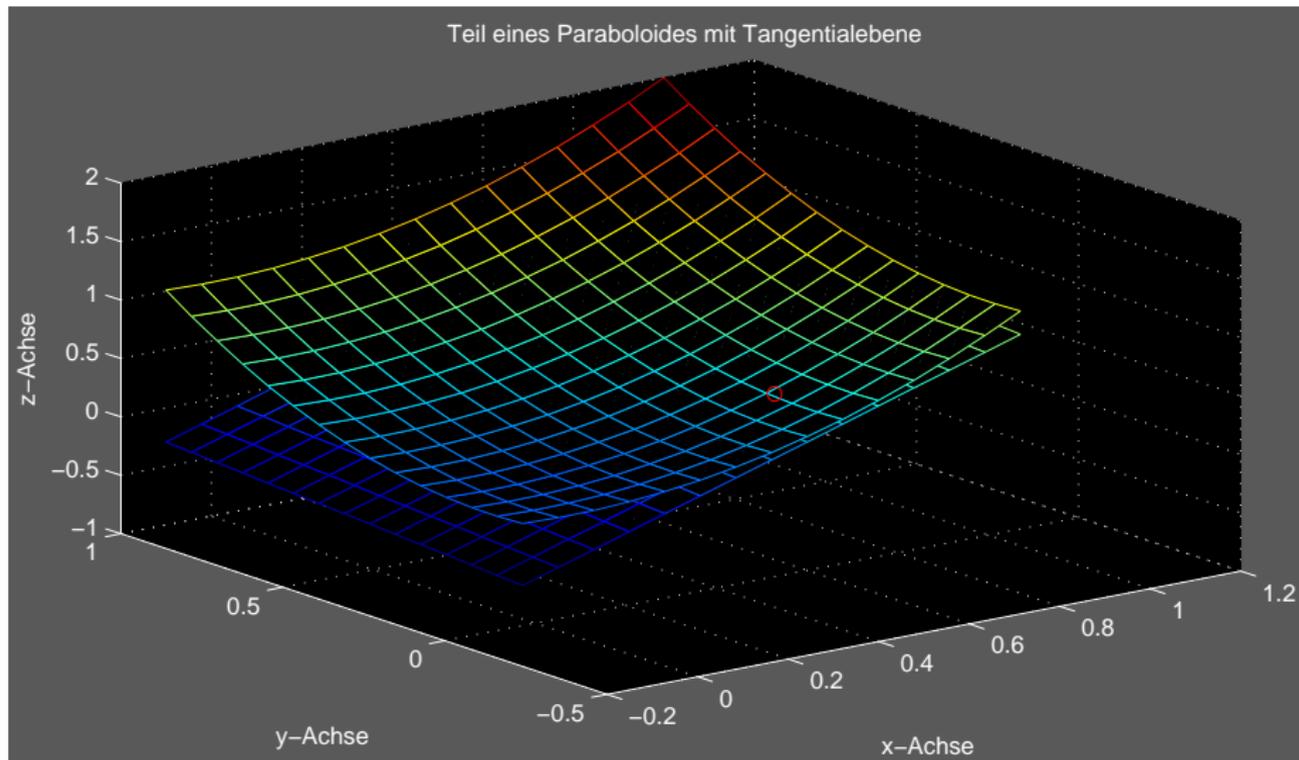
Um diese Frage beantworten zu können, müssen wir die Begriffe der (stetigen) Differenzierbarkeit und (lokalen) quadratischen Approximation einführen.

Dazu schauen wir uns als Beispiel die Funktion

$$f(x) = x_1^2 + x_2^2 \quad (44)$$

an. Diese Funktion ist ein elliptisches Paraboloid (eine gedrehte Parabel), welches im Punkte 0 ein Minimum hat.

# Vektorwertige reelle Funktionen



# Vektorwertige reelle Funktionen

Im Bild zu sehen ist auch der Punkt  $(x_0, y_0) = (0.1, 0.6)$  und die Ebene

$$T(x, y) := f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (45)$$

$$= 0.1^2 + 0.6^2 + 2 \cdot 0.1 \cdot (x - 0.1) + 2 \cdot 0.6 \cdot (y - 0.6). \quad (46)$$

# Vektorwertige reelle Funktionen

Im Bild zu sehen ist auch der Punkt  $(x_0, y_0) = (0.1, 0.6)$  und die Ebene

$$T(x, y) := f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (45)$$

$$= 0.1^2 + 0.6^2 + 2 \cdot 0.1 \cdot (x - 0.1) + 2 \cdot 0.6 \cdot (y - 0.6). \quad (46)$$

Man sieht schnell, dass

$$T(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \quad (47)$$

# Vektorwertige reelle Funktionen

Im Bild zu sehen ist auch der Punkt  $(x_0, y_0) = (0.1, 0.6)$  und die Ebene

$$T(x, y) := f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (45)$$

$$= 0.1^2 + 0.6^2 + 2 \cdot 0.1 \cdot (x - 0.1) + 2 \cdot 0.6 \cdot (y - 0.6). \quad (46)$$

Man sieht schnell, dass

$$T(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \quad (47)$$

wobei der Zeilenvektor  $\nabla f(x_0, y_0)$  definiert ist als

$$\nabla f(x_0, y_0) = \left( \frac{\partial f}{\partial x}(x_0, y_0) \quad \frac{\partial f}{\partial y}(x_0, y_0) \right). \quad (48)$$

# Vektorwertige reelle Funktionen

Im Bild zu sehen ist auch der Punkt  $(x_0, y_0) = (0.1, 0.6)$  und die Ebene

$$T(x, y) := f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (45)$$

$$= 0.1^2 + 0.6^2 + 2 \cdot 0.1 \cdot (x - 0.1) + 2 \cdot 0.6 \cdot (y - 0.6). \quad (46)$$

Man sieht schnell, dass

$$T(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \quad (47)$$

wobei der Zeilenvektor  $\nabla f(x_0, y_0)$  definiert ist als

$$\nabla f(x_0, y_0) = \left( \frac{\partial f}{\partial x}(x_0, y_0) \quad \frac{\partial f}{\partial y}(x_0, y_0) \right). \quad (48)$$

Die Ebene  $T$  liegt am Punkt  $(x_0, y_0)$  in jede Richtung tangential zum Paraboloid, daher auch "Tangentialebene".

# Vektorwertige reelle Funktionen

Im Bild zu sehen ist auch der Punkt  $(x_0, y_0) = (0.1, 0.6)$  und die Ebene

$$T(x, y) := f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \quad (45)$$

$$= 0.1^2 + 0.6^2 + 2 \cdot 0.1 \cdot (x - 0.1) + 2 \cdot 0.6 \cdot (y - 0.6). \quad (46)$$

Man sieht schnell, dass

$$T(x, y) = f(x_0, y_0) + \nabla f(x_0, y_0) \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \quad (47)$$

wobei der Zeilenvektor  $\nabla f(x_0, y_0)$  definiert ist als

$$\nabla f(x_0, y_0) = \left( \frac{\partial f}{\partial x}(x_0, y_0) \quad \frac{\partial f}{\partial y}(x_0, y_0) \right). \quad (48)$$

Die Ebene  $T$  liegt am Punkt  $(x_0, y_0)$  in jede Richtung tangential zum Paraboloid, daher auch "Tangentialebene". Beweis: durch die Ableitung von Koeffizientenfunktionen von im Paraboloid liegenden Kurven.

# Vektorwertige reelle Funktionen

Der so definierte Zeilenvektor  $\nabla f$  ist das vektorielle Analogon zur Ableitung, allgemein gilt:

# Vektorwertige reelle Funktionen

Der so definierte Zeilenvektor  $\nabla f$  ist das vektorielle Analogon zur Ableitung, allgemein gilt:

## Theorem (Differenzierbarkeit vektorieller Funktionen)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  nach allen Variablen  $x_1$  bis  $x_n$  in einer Umgebung von  $\hat{x}$  (partiell) stetig differenzierbar. Dann ist der Gradient  $\nabla f$  von  $f$ ,

$$\nabla f(\hat{x}) := \left( \frac{\partial f}{\partial x_1}(\hat{x}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\hat{x}) \right), \quad (49)$$

die erste Ableitung von  $f$  im Punkt  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)^T$ .

# Vektorwertige reelle Funktionen

Der so definierte Zeilenvektor  $\nabla f$  ist das vektorielle Analogon zur Ableitung, allgemein gilt:

## Theorem (Differenzierbarkeit vektorieller Funktionen)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  nach allen Variablen  $x_1$  bis  $x_n$  in einer Umgebung von  $\hat{x}$  (partiell) stetig differenzierbar. Dann ist der Gradient  $\nabla f$  von  $f$ ,

$$\nabla f(\hat{x}) := \left( \frac{\partial f}{\partial x_1}(\hat{x}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\hat{x}) \right), \quad (49)$$

die erste Ableitung von  $f$  im Punkt  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)^T$ .

Die Tangentialebene ist die Entsprechung der Taylor-Entwicklung der Funktion vom Grad Eins.

# Vektorwertige reelle Funktionen

Der so definierte Zeilenvektor  $\nabla f$  ist das vektorielle Analogon zur Ableitung, allgemein gilt:

## Theorem (Differenzierbarkeit vektorieller Funktionen)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  nach allen Variablen  $x_1$  bis  $x_n$  in einer Umgebung von  $\hat{x}$  (partiell) stetig differenzierbar. Dann ist der Gradient  $\nabla f$  von  $f$ ,

$$\nabla f(\hat{x}) := \left( \frac{\partial f}{\partial x_1}(\hat{x}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\hat{x}) \right), \quad (49)$$

die erste Ableitung von  $f$  im Punkt  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)^T$ .

Die Tangentialebene ist die Entsprechung der Taylor-Entwicklung der Funktion vom Grad Eins. Daraus kann man ablesen, dass der Gradient die Richtung des lokal steilsten Anstiegs angibt, also der negative Gradient lokal die Richtung des stärksten Abstieges angibt.

# Vektorwertige reelle Funktionen

Damit kann man stationäre Punkte (einschließlich Minima) charakterisieren. Wenn es keine Richtung des stärksten An- oder Abstiegs gibt, ist die Funktion stationär in dem Punkt.

# Vektorwertige reelle Funktionen

Damit kann man stationäre Punkte (einschließlich Minima) charakterisieren. Wenn es keine Richtung des stärksten An- oder Abstiegs gibt, ist die Funktion stationär in dem Punkt.

Im Beispiel  $f(x) = x^T x = x_1^2 + x_2^2$  ist der Gradient ausgewertet an Null identisch gleich Null, da

$$\nabla f(x_1, x_2) = 2 \begin{pmatrix} x_1 & x_2 \end{pmatrix} = 2x^T. \quad (50)$$

Dieses Beispiel verallgemeinert eine Parabel auf mehrere Dimensionen.

# Vektorwertige reelle Funktionen

Damit kann man stationäre Punkte (einschließlich Minima) charakterisieren. Wenn es keine Richtung des stärksten An- oder Abstiegs gibt, ist die Funktion stationär in dem Punkt.

Im Beispiel  $f(x) = x^T x = x_1^2 + x_2^2$  ist der Gradient ausgewertet an Null identisch gleich Null, da

$$\nabla f(x_1, x_2) = 2 \begin{pmatrix} x_1 & x_2 \end{pmatrix} = 2x^T. \quad (50)$$

Dieses Beispiel verallgemeinert eine Parabel auf mehrere Dimensionen.

Da die Funktion also “quadratisch” ist, ist sie ihre eigene quadratische Approximation. Allgemein ist eine quadratische Funktion in mehreren Veränderlichen gegeben als

$$Q(x) = x^T A x + b^T x + c, \quad (51)$$

wobei  $x \in \mathbb{R}^n$  der Spaltenvektor der Variablen ist,  $b, c \in \mathbb{R}$  Spaltenvektoren mit Konstanten und  $A \in \mathbb{R}^{(n,n)}$  eine symmetrisch wählbare **Matrix** ist.

# Vektorwertige reelle Funktionen

Wie kann man nun Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  lokal quadratisch, also mittels

$$Q(x) = x^T A x + b^T x + c, \quad (52)$$

approximieren?

# Vektorwertige reelle Funktionen

Wie kann man nun Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  lokal quadratisch, also mittels

$$Q(x) = x^T A x + b^T x + c, \quad (52)$$

approximieren?

Wie im Eindimensionalen verwenden wir die Idee der Taylor-Approximation, also die Übereinstimmung in den ersten Ableitungen. Die erste Ableitung haben wir bereits definiert. Die zweite Ableitung muss also eine symmetrische **Matrix** sein.

# Vektorwertige reelle Funktionen

Wie kann man nun Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  lokal quadratisch, also mittels

$$Q(x) = x^T A x + b^T x + c, \quad (52)$$

approximieren?

Wie im Eindimensionalen verwenden wir die Idee der Taylor-Approximation, also die Übereinstimmung in den ersten Ableitungen. Die erste Ableitung haben wir bereits definiert. Die zweite Ableitung muss also eine symmetrische **Matrix** sein.

Wenn man sich veranschaulicht, wie man von der Funktion zum Gradienten kommt, nämlich durch partielle Ableitungen der Funktion nach **jeder** Variablen, so ist einleuchtend, dass gute Kandidaten für die Einträge der Matrix der zweiten Ableitung die Ableitungen **jeder** erhaltenen partiellen Ableitung nach **jeder** Variablen sind.

# Vektorwertige reelle Funktionen

## Theorem (Hessematrix)

Die zweite Ableitung einer Funktion  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , die sogenannte Hessematrix, bezeichnet mit  $\nabla^2 f$ , existiert, wenn alle angegebenen partiellen Ableitungen existieren und stetig sind, und ist gegeben als

$$\nabla^2 f := \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \quad (53)$$

# Vektorwertige reelle Funktionen

## Theorem (Hessematrix)

Die zweite Ableitung einer Funktion  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , die sogenannte Hessematrix, bezeichnet mit  $\nabla^2 f$ , existiert, wenn alle angegebenen partiellen Ableitungen existieren und stetig sind, und ist gegeben als

$$\nabla^2 f := \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \quad (53)$$

Das Beispiel  $f(x) = x^T x = \sum_{i=1}^n x_i^2$  hat also im Punkt  $x_0 = e = (1 \ \cdots \ 1)^T$  den Gradienten (die erste Ableitung)

$$\nabla f(x) = 2x^T \quad \Rightarrow \quad \nabla f(e) = 2e^T \quad (54)$$

und die Hessematrix (die zweite Ableitung;  $E = (\delta_{ij})$  ist die Einheitsmatrix)

$$\nabla^2 f(x) = 2E \quad \Rightarrow \quad \nabla^2 f(e) = 2E. \quad (55)$$

# Vektorwertige reelle Funktionen

Dass die Hessematrix unter den angegebenen Voraussetzungen symmetrisch ist, folgt aus dem

# Vektorwertige reelle Funktionen

Dass die Hessematrix unter den angegebenen Voraussetzungen symmetrisch ist, folgt aus dem

## Theorem (Satz von Schwarz)

Die Funktion  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  besitze in einer Umgebung von  $\hat{x} \in \mathbb{R}^n$  die partiellen Ableitungen

$$\frac{\partial f}{\partial x_i}, \quad \frac{\partial f}{\partial x_j} \quad \text{und} \quad \frac{\partial^2 f}{\partial x_j \partial x_i}, \quad (56)$$

wobei die letzte partielle Ableitung stetig sei. Dann existiert auch

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}. \quad (57)$$

# Vektorwertige reelle Funktionen

Die Funktion aus dem Beispiel läßt sich schreiben als

$$f(e) + \nabla f(e)(x - e) + \frac{1}{2}(x - e)^T \nabla^2 f(e)(x - e) = \\ e^T e + 2e^T(x - e) + (x - e)^T(x - e) = x^T x = f(x). \quad (58)$$

# Vektorwertige reelle Funktionen

Die Funktion aus dem Beispiel läßt sich schreiben als

$$f(e) + \nabla f(e)(x - e) + \frac{1}{2}(x - e)^T \nabla^2 f(e)(x - e) = e^T e + 2e^T(x - e) + (x - e)^T(x - e) = x^T x = f(x). \quad (58)$$

Allgemein gilt nur

$$f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0), \quad (59)$$

welches die mehrdimensionale Taylor-Approximation vom Grad 2 der Funktion  $f$  darstellt.

# Vektorwertige reelle Funktionen

Die Funktion aus dem Beispiel läßt sich schreiben als

$$f(e) + \nabla f(e)(x - e) + \frac{1}{2}(x - e)^T \nabla^2 f(e)(x - e) = \\ e^T e + 2e^T(x - e) + (x - e)^T(x - e) = x^T x = f(x). \quad (58)$$

Allgemein gilt nur

$$f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0), \quad (59)$$

welches die mehrdimensionale Taylor-Approximation vom Grad 2 der Funktion  $f$  darstellt.

Die erste Ableitung von  $f$ , also der Gradient, den wir auf Null bringen wollen, wird approximiert durch

$$\nabla f(x) \approx \nabla f(x_0) + (x - x_0)^T \nabla^2 f(x_0). \quad (60)$$

# Vektorwertige reelle Funktionen

Das Newton-Verfahren zur Minimierung basiert auf der Nullsetzung der rechten Seite, also Auffindung eines  $\tilde{x}$ , so dass

$$\nabla f(x_0) + (\tilde{x} - x_0)^T \nabla^2 f(x_0) = o_n^T. \quad (61)$$

# Vektorwertige reelle Funktionen

Das Newton-Verfahren zur Minimierung basiert auf der Nullsetzung der rechten Seite, also Auffindung eines  $\tilde{x}$ , so dass

$$\nabla f(x_0) + (\tilde{x} - x_0)^T \nabla^2 f(x_0) = o_n^T. \quad (61)$$

Nullsetzung bedeutet die Lösung eines linearen Gleichungssystems,

$$\tilde{x} = x_0 - (\nabla^2 f(x_0))^{-1} \nabla f(x_0)^T, \quad (62)$$

welches nur funktioniert, wenn die Hessematrix invertierbar ist.

# Vektorwertige reelle Funktionen

Das Newton-Verfahren zur Minimierung basiert auf der Nullsetzung der rechten Seite, also Auffindung eines  $\tilde{x}$ , so dass

$$\nabla f(x_0) + (\tilde{x} - x_0)^T \nabla^2 f(x_0) = o_n^T. \quad (61)$$

Nullsetzung bedeutet die Lösung eines linearen Gleichungssystems,

$$\tilde{x} = x_0 - (\nabla^2 f(x_0))^{-1} \nabla f(x_0)^T, \quad (62)$$

welches nur funktioniert, wenn die Hessematrix invertierbar ist.

Das Newton-Verfahren iteriert gemäß

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)^T. \quad (63)$$

# Vektorwertige reelle Funktionen

Das Newton-Verfahren zur Minimierung basiert auf der Nullsetzung der rechten Seite, also Auffindung eines  $\tilde{x}$ , so dass

$$\nabla f(x_0) + (\tilde{x} - x_0)^T \nabla^2 f(x_0) = o_n^T. \quad (61)$$

Nullsetzung bedeutet die Lösung eines linearen Gleichungssystems,

$$\tilde{x} = x_0 - (\nabla^2 f(x_0))^{-1} \nabla f(x_0)^T, \quad (62)$$

welches nur funktioniert, wenn die Hessematrix invertierbar ist.

Das Newton-Verfahren iteriert gemäß

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)^T. \quad (63)$$

Man kann lokale quadratische Konvergenz unter gewissen Voraussetzungen (Differenzierbarkeit; reguläre Hessematrix in der Nullstelle) beweisen.

# Vektorwertige reelle Funktionen

Aussagen über die Konvergenz und etwaige stationäre Punkte (eigentlich in der ursprünglichen Form über Nullstellen) machen die beiden folgenden Sätze.

# Vektorwertige reelle Funktionen

Aussagen über die Konvergenz und etwaige stationäre Punkte (eigentlich in der ursprünglichen Form über Nullstellen) machen die beiden folgenden Sätze.

Die Norm  $\|x\|$  eines Vektors ist seine Euklidische Länge, gegeben durch die (im Zweidimensionalen bekannte) Formel

$$\|x\| := \sqrt{\sum_{i=1}^n x_i^2}. \quad (64)$$

# Vektorwertige reelle Funktionen

Aussagen über die Konvergenz und etwaige stationäre Punkte (eigentlich in der ursprünglichen Form über Nullstellen) machen die beiden folgenden Sätze.

Die Norm  $\|x\|$  eines Vektors ist seine Euklidische Länge, gegeben durch die (im Zweidimensionalen bekannte) Formel

$$\|x\| := \sqrt{\sum_{i=1}^n x_i^2}. \quad (64)$$

Die Norm  $\|A\|$  einer Matrix ist die maximale Verlängerung, die ein Vektor durch die Multiplikation mit dieser Matrix erfährt,

$$\|A\| := \max_{x \in \mathbb{R}^n \setminus \{o_n\}} \frac{\|Ax\|}{\|x\|}. \quad (65)$$

# Vektorwertige reelle Funktionen

Wir benötigen eine mathematische Version des Begriffes der Umgebung, genauer, den Begriff des offenen und abgeschlossenen Balls um einen Punkt im  $\mathbb{R}^n$  mit Radius  $r > 0$ .

# Vektorwertige reelle Funktionen

Wir benötigen eine mathematische Version des Begriffes der Umgebung, genauer, den Begriff des offenen und abgeschlossenen Balls um einen Punkt im  $\mathbb{R}^n$  mit Radius  $r > 0$ .

## Definition (Bälle im $\mathbb{R}^n$ )

Sei  $x \in \mathbb{R}^n$ . Der offene Ball mit Radius  $r > 0$  um  $x$  ist definiert als die Menge

$$B_r(x) := \{y \in \mathbb{R}^n : \|x - y\| < r\}. \quad (66)$$

Der abgeschlossene Ball mit Radius  $r > 0$  um  $x$  ist analog definiert als die Menge

$$\bar{B}_r(x) := \{y \in \mathbb{R}^n : \|x - y\| \leq r\}. \quad (67)$$

# Vektorwertige reelle Funktionen

## Theorem (Satz von Kantorovich)

Sei  $D$  konvex,  $x_0 \in D \subset \mathbb{R}^n$  und  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar, so dass die Hessematrix  $\nabla^2 f(x_0)$  invertierbar ist. Sei

$$h_0 = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0), \quad x_1 = x_0 + h_0. \quad (68)$$

Wenn  $B_{\|h_0\|}(x_1) \subset D$  und die zweite Ableitung eine sogenannte Lipschitz-Bedingung

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\| \quad \forall x, y \in D \quad (69)$$

erfüllt, so konvergiert unter der Voraussetzung

$$L\|h_0\| \|(\nabla^2 f(x_0))^{-1}\| \leq \frac{1}{2} \quad (70)$$

das Newton-Verfahren gegen den eindeutigen stationären Punkt  $\hat{x} \in \bar{B}_{\|h_0\|}(x_1)$ .

# Vektorwertige reelle Funktionen

## Theorem (Alpha-Test von Smale)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  *analytisch*,  $x_0 \in D$ . Definiere

$$\alpha(f, x) := \beta(f, x) \cdot \gamma(f, x), \quad (71)$$

$$\beta(f, x) := \|(\nabla^2 f(x))^{-1} \nabla f(x)\|, \quad (72)$$

$$\gamma(f, x) := \sup_{k \geq 3} \left\| (\nabla^2 f(x))^{-1} \frac{\nabla^k f(x)}{(k-1)!} \right\|^{1/(k-2)}. \quad (73)$$

Wenn  $\alpha(f, x_0) < \alpha_0 \approx 0.130707$  konvergiert das Newton-Verfahren und es gilt

$$\|x_{k+1} - x_k\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|x_1 - x_0\|. \quad (74)$$

# Vektorwertige reelle Funktionen

## Theorem (Alpha-Test von Smale)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  *analytisch*,  $x_0 \in D$ . Definiere

$$\alpha(f, x) := \beta(f, x) \cdot \gamma(f, x), \quad (71)$$

$$\beta(f, x) := \|(\nabla^2 f(x))^{-1} \nabla f(x)\|, \quad (72)$$

$$\gamma(f, x) := \sup_{k \geq 3} \left\| (\nabla^2 f(x))^{-1} \frac{\nabla^k f(x)}{(k-1)!} \right\|^{1/(k-2)}. \quad (73)$$

Wenn  $\alpha(f, x_0) < \alpha_0 \approx 0.130707$  konvergiert das Newton-Verfahren und es gilt

$$\|x_{k+1} - x_k\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|x_1 - x_0\|. \quad (74)$$

Damit ist das Problem, eine Lipschitzkonstante der zweiten Ableitung auf einem Gebiet zu bestimmen (Kantorovich), ersetzt worden durch Formel (73) (Smale).

# Vektorwertige reelle Funktionen

## Theorem (Alpha-Test von Smale)

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  *analytisch*,  $x_0 \in D$ . Definiere

$$\alpha(f, x) := \beta(f, x) \cdot \gamma(f, x), \quad (71)$$

$$\beta(f, x) := \|(\nabla^2 f(x))^{-1} \nabla f(x)\|, \quad (72)$$

$$\gamma(f, x) := \sup_{k \geq 3} \left\| (\nabla^2 f(x))^{-1} \frac{\nabla^k f(x)}{(k-1)!} \right\|^{1/(k-2)}. \quad (73)$$

Wenn  $\alpha(f, x_0) < \alpha_0 \approx 0.130707$  konvergiert das Newton-Verfahren und es gilt

$$\|x_{k+1} - x_k\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|x_1 - x_0\|. \quad (74)$$

Damit ist das Problem, eine Lipschitzkonstante der zweiten Ableitung auf einem Gebiet zu bestimmen (Kantorovich), ersetzt worden durch Formel (73) (Smale). Alles reine Geschmacksache.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

Man kann aber zeigen, dass es immer eine orthogonale Matrix  $Q$  gibt, so dass die Matrix

$$\Lambda = Q^T A Q \quad (75)$$

diagonal ist.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

Man kann aber zeigen, dass es immer eine orthogonale Matrix  $Q$  gibt, so dass die Matrix

$$\Lambda = Q^T A Q \quad (75)$$

diagonal ist.

**Orthogonal** bedeutet, dass

$$Q^T Q = E, \quad (76)$$

also dass die “Inverse” Matrix die Transponierte ist.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

Man kann aber zeigen, dass es immer eine orthogonale Matrix  $Q$  gibt, so dass die Matrix

$$\Lambda = Q^T A Q \quad (75)$$

diagonal ist.

**Orthogonal** bedeutet, dass

$$Q^T Q = E, \quad (76)$$

also dass die “Inverse” Matrix die Transponierte ist.

Die Transformation  $A \rightarrow Q^T A Q$  ist gleichbedeutend mit einem Basiswechsel, also der Einführung neuer Variablen durch Kombinationen alter.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

Man kann aber zeigen, dass es immer eine orthogonale Matrix  $Q$  gibt, so dass die Matrix

$$\Lambda = Q^T A Q \quad (75)$$

diagonal ist.

**Orthogonal** bedeutet, dass

$$Q^T Q = E, \quad (76)$$

also dass die “Inverse” Matrix die Transponierte ist.

Die Transformation  $A \rightarrow Q^T A Q$  ist gleichbedeutend mit einem Basiswechsel, also der Einführung neuer Variablen durch Kombinationen alter. Der orthogonale Basiswechsel ändert weder Längen noch Winkel.

# Vektorwertige reelle Funktionen

Ob ein stationärer Punkt denn nun ein Minimum ist, läßt sich oft an der zweiten Ableitung erkennen. Dazu fehlt leider noch etwas Wissen über die Eigenwerte und Eigenvektoren einer symmetrischen reellen Matrix, wie der Hessematrix.

Man kann aber zeigen, dass es immer eine orthogonale Matrix  $Q$  gibt, so dass die Matrix

$$\Lambda = Q^T A Q \quad (75)$$

diagonal ist.

**Orthogonal** bedeutet, dass

$$Q^T Q = E, \quad (76)$$

also dass die “Inverse” Matrix die Transponierte ist.

Die Transformation  $A \rightarrow Q^T A Q$  ist gleichbedeutend mit einem Basiswechsel, also der Einführung neuer Variablen durch Kombinationen alter. Der orthogonale Basiswechsel ändert weder Längen noch Winkel. Der stationäre Punkt ist ein Minimum, wenn alle Diagonalelemente  $\{\lambda_i\}_{i=1}^n$  von  $\Lambda$  größer als Null sind.

# Übersicht

Zielsetzung

Einordnung

Abgrenzung

Mengen, Minima & Funktionen

Mengen

Minima

Funktionen

Das Newton-Verfahren

Vorbemerkungen

$$f : D \subset \mathbb{R} \rightarrow \mathbb{R}$$

$$f : D \subset \mathbb{C} \rightarrow \mathbb{R}$$

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

$$f : D \subset M \rightarrow \mathbb{R}$$

# Grenzen des $\mathbb{R}^n$

Manche natürlich auftretenden Fragestellungen lassen sich nicht in “Vektorräume”  $\mathbb{R}^n$  packen. Eine solche Minimierungsaufgabe ist mit der Frage nach dem tiefsten Punkt der Erde (bekanntermassen irgendwo im Mariannengraben) gegeben.

# Grenzen des $\mathbb{R}^n$

Manche natürlich auftretenden Fragestellungen lassen sich nicht in “Vektorräume”  $\mathbb{R}^n$  packen. Eine solche Minimierungsaufgabe ist mit der Frage nach dem tiefsten Punkt der Erde (bekanntermassen irgendwo im Mariannengraben) gegeben.

Die Aufgabe kann hier interpretiert werden als: Minimiere die Funktion “Höhe über Normalnull” über der Einheitskugel. Das Ergebnis ist die Richtung in einem gegebenen Koordinatensystem auf der Einheitskugel.

# Grenzen des $\mathbb{R}^n$

Manche natürlich auftretenden Fragestellungen lassen sich nicht in “Vektorräume”  $\mathbb{R}^n$  packen. Eine solche Minimierungsaufgabe ist mit der Frage nach dem tiefsten Punkt der Erde (bekanntermassen irgendwo im Mariannengraben) gegeben.

Die Aufgabe kann hier interpretiert werden als: Minimiere die Funktion “Höhe über Normalnull” über der Einheitskugel. Das Ergebnis ist die Richtung in einem gegebenen Koordinatensystem auf der Einheitskugel.

Lokal sieht die Erdoberfläche (idealisiert, jedenfalls hier im Norden) aus wie eine Ebene,  $S \approx \mathbb{R}^2$ .

# Grenzen des $\mathbb{R}^n$

Manche natürlich auftretenden Fragestellungen lassen sich nicht in “Vektorräume”  $\mathbb{R}^n$  packen. Eine solche Minimierungsaufgabe ist mit der Frage nach dem tiefsten Punkt der Erde (bekanntermassen irgendwo im Mariannengraben) gegeben.

Die Aufgabe kann hier interpretiert werden als: Minimiere die Funktion “Höhe über Normalnull” über der Einheitskugel. Das Ergebnis ist die Richtung in einem gegebenen Koordinatensystem auf der Einheitskugel.

Lokal sieht die Erdoberfläche (idealisiert, jedenfalls hier im Norden) aus wie eine Ebene,  $S \approx \mathbb{R}^2$ . Allgemeinere Objekte, die lokal (in einer sogenannten “Karte”) wie der  $\mathbb{R}^n$  aussehen, wobei die “Karten” aus einem das ganze Objekt überdeckenden “Atlas” stammen und die “Kartenwechsel” “glatt” vonstatten gehen, nennt man eine ( $n$ -dimensionale reelle) Mannigfaltigkeit.

# Grenzen des $\mathbb{R}^n$

Manche natürlich auftretenden Fragestellungen lassen sich nicht in “Vektorräume”  $\mathbb{R}^n$  packen. Eine solche Minimierungsaufgabe ist mit der Frage nach dem tiefsten Punkt der Erde (bekanntermassen irgendwo im Mariannengraben) gegeben.

Die Aufgabe kann hier interpretiert werden als: Minimiere die Funktion “Höhe über Normalnull” über der Einheitskugel. Das Ergebnis ist die Richtung in einem gegebenen Koordinatensystem auf der Einheitskugel.

Lokal sieht die Erdoberfläche (idealisiert, jedenfalls hier im Norden) aus wie eine Ebene,  $S \approx \mathbb{R}^2$ . Allgemeinere Objekte, die lokal (in einer sogenannten “Karte”) wie der  $\mathbb{R}^n$  aussehen, wobei die “Karten” aus einem das ganze Objekt überdeckenden “Atlas” stammen und die “Kartenwechsel” “glatt” vonstatten gehen, nennt man eine ( $n$ -dimensionale reelle) Mannigfaltigkeit. Meistens wird von den Karten auch eine gewisse “Glätte” verlangt.

# Reelle differenzierbare Mannigfaltigkeiten

Reelle differenzierbare Mannigfaltigkeiten sind sehr allgemeine Objekte, auf denen immer noch Optimierung betrieben werden kann. Dazu muss erst das mathematische Vokabular aufgebaut werden.

# Reelle differenzierbare Mannigfaltigkeiten

Reelle differenzierbare Mannigfaltigkeiten sind sehr allgemeine Objekte, auf denen immer noch Optimierung betrieben werden kann. Dazu muss erst das mathematische Vokabular aufgebaut werden.

## Definition (Differenzierbare reelle Mannigfaltigkeit)

Eine differenzierbare reelle Mannigfaltigkeit ist eine Menge  $M$  mit einer Familie ( $\alpha$  ist aus einer Indexmenge) von injektiven Abbildungen

$$\mathfrak{r}_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow M \quad (77)$$

von offenen Mengen  $U_\alpha$  des  $\mathbb{R}^n$  nach  $M$ , so dass gilt:

$$\cup_\alpha \mathfrak{r}_\alpha(U_\alpha) = M, \quad (78)$$

$$\forall \alpha, \beta : \mathfrak{r}_\alpha(U_\alpha) \cap \mathfrak{r}_\beta(U_\beta) = W \neq \emptyset \Rightarrow \mathfrak{r}_\beta^{-1} \circ \mathfrak{r}_\alpha \in C^\omega(W), \quad (79)$$

$$\{\mathfrak{r}_\alpha, U_\alpha\} \text{ ist maximal bzgl. (78) und (79).} \quad (80)$$

# Reelle differenzierbare Mannigfaltigkeiten

Die injektiven Abbildungen nennt man die Koordinatenabbildungen oder auch Karten. Die Familie  $\{x_\alpha, U_\alpha\}$  aller Karten nennt man einen Atlas. Bedingung (80) sorgt dafür, dass der Atlas “maximal” ist, was eine rein technische Bedingung ist.

# Reelle differenzierbare Mannigfaltigkeiten

Die injektiven Abbildungen nennt man die Koordinatenabbildungen oder auch Karten. Die Familie  $\{\chi_\alpha, U_\alpha\}$  aller Karten nennt man einen Atlas.

Bedingung (80) sorgt dafür, dass der Atlas “maximal” ist, was eine rein technische Bedingung ist.

Bedingung (79) ist die obengenannte Bedingung an den “Kartenwechsel”, wenn man feststellt, dass man eine neue Karte nehmen muss, sollten die gemeinsamen Gegebenheiten auch noch “ähnlich” aussehen (was bei der Optimierung immens wichtig ist).

# Reelle differenzierbare Mannigfaltigkeiten

Die injektiven Abbildungen nennt man die Koordinatenabbildungen oder auch Karten. Die Familie  $\{\mathfrak{x}_\alpha, U_\alpha\}$  aller Karten nennt man einen Atlas.

Bedingung (80) sorgt dafür, dass der Atlas “maximal” ist, was eine rein technische Bedingung ist.

Bedingung (79) ist die obengenannte Bedingung an den “Kartenwechsel”, wenn man feststellt, dass man eine neue Karte nehmen muss, sollten die gemeinsamen Gegebenheiten auch noch “ähnlich” aussehen (was bei der Optimierung immens wichtig ist).

Um optimieren zu können, insbesondere um “ein” Newton-Verfahren für Funktionen  $f : D \subset M \rightarrow \mathbb{R}$  zu definieren, benötigen wir Funktionen und deren Ableitungen. Funktionen werden analog als spezielle Teilmenge des kartesischen Produktes definiert.

# Reelle differenzierbare Mannigfaltigkeiten

Die injektiven Abbildungen nennt man die Koordinatenabbildungen oder auch Karten. Die Familie  $\{\mathfrak{x}_\alpha, U_\alpha\}$  aller Karten nennt man einen Atlas.

Bedingung (80) sorgt dafür, dass der Atlas “maximal” ist, was eine rein technische Bedingung ist.

Bedingung (79) ist die obengenannte Bedingung an den “Kartenwechsel”, wenn man feststellt, dass man eine neue Karte nehmen muss, sollten die gemeinsamen Gegebenheiten auch noch “ähnlich” aussehen (was bei der Optimierung immens wichtig ist).

Um optimieren zu können, insbesondere um “ein” Newton-Verfahren für Funktionen  $f : D \subset M \rightarrow \mathbb{R}$  zu definieren, benötigen wir Funktionen und deren Ableitungen. Funktionen werden analog als spezielle Teilmenge des kartesischen Produktes definiert.

Ableitungen haben wir bis dato nur für den  $\mathbb{R}^n$  definiert. Lokal entspricht eine Mannigfaltigkeit ja dem  $\mathbb{R}^n$ . Dieser Sachverhalt wird verwendet, um Ableitungen zu definieren.

# Reelle differenzierbare Mannigfaltigkeiten

Zuerst muss gefragt werden, was denn die Ableitung beschreiben soll. Eine naheliegende Idee, wenn man an die Zweidimensionale Sphäre denkt, ist es, alle Tangentenvektoren an einem Punkt als mögliche Ableitungen zu betrachten.

# Reelle differenzierbare Mannigfaltigkeiten

Zuerst muss gefragt werden, was denn die Ableitung beschreiben soll. Eine naheliegende Idee, wenn man an die Zweidimensionale Sphäre denkt, ist es, alle Tangentenvektoren an einem Punkt als mögliche Ableitungen zu betrachten.

Das wird motiviert durch die Ableitung einer in der Mannigfaltigkeit verlaufenden Kurve, welche ja dann bei einer solchen in einen  $\mathbb{R}^n$  “eingebetteten” Mannigfaltigkeit auf die Ableitung von Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  hinausläuft.

# Reelle differenzierbare Mannigfaltigkeiten

Zuerst muss gefragt werden, was denn die Ableitung beschreiben soll. Eine naheliegende Idee, wenn man an die Zweidimensionale Sphäre denkt, ist es, alle Tangentenvektoren an einem Punkt als mögliche Ableitungen zu betrachten.

Das wird motiviert durch die Ableitung einer in der Mannigfaltigkeit verlaufenden Kurve, welche ja dann bei einer solchen in einen  $\mathbb{R}^n$  "eingebetteten" Mannigfaltigkeit auf die Ableitung von Funktionen  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  hinausläuft.

Dazu hat man allgemein dann nicht nur eine Tangentialebene an einer Mannigfaltigkeit, sondern einen Tangentialraum. Diesen werden wir noch genauer definieren. Zuerst definieren wir die Differentiation auf einer Mannigfaltigkeit.

# Reelle differenzierbare Mannigfaltigkeiten

## Definition (Differenzierbarkeit auf Mannigfaltigkeiten)

Sei  $M$  eine reelle differenzierbare Mannigfaltigkeit,  $f : D \subset M \rightarrow \mathbb{R}$ . Die Funktion  $f$  heißt in  $p \in D \subset M$  differenzierbar, wenn für ein offenes Intervall  $I = (a, b)$  mit  $a < f(p) < b$  eine Karte  $\chi : U \subset \mathbb{R}^n \rightarrow M$  um  $p$  mit  $f(\chi(U)) \subset I$  existiert und die Abbildung

$$f \circ \chi : U \subset \mathbb{R}^n \rightarrow \mathbb{R} \quad (81)$$

bei  $\chi^{-1}(p)$  differenzierbar ist. Die Funktion heißt differenzierbar, wenn sie für alle Punkte differenzierbar ist.

# Reelle differenzierbare Mannigfaltigkeiten

## Definition (Differenzierbarkeit auf Mannigfaltigkeiten)

Sei  $M$  eine reelle differenzierbare Mannigfaltigkeit,  $f : D \subset M \rightarrow \mathbb{R}$ . Die Funktion  $f$  heißt in  $p \in D \subset M$  differenzierbar, wenn für ein offenes Intervall  $I = (a, b)$  mit  $a < f(p) < b$  eine Karte  $\chi : U \subset \mathbb{R}^n \rightarrow M$  um  $p$  mit  $f(\chi(U)) \subset I$  existiert und die Abbildung

$$f \circ \chi : U \subset \mathbb{R}^n \rightarrow \mathbb{R} \quad (81)$$

bei  $\chi^{-1}(p)$  differenzierbar ist. Die Funktion heißt differenzierbar, wenn sie für alle Punkte differenzierbar ist.

Man muss sich jetzt erst mal klar machen, dass diese Definition der Differenzierbarkeit nicht von der Wahl der Karte abhängt.

# Reelle differenzierbare Mannigfaltigkeiten

## Definition (Differenzierbarkeit auf Mannigfaltigkeiten)

Sei  $M$  eine reelle differenzierbare Mannigfaltigkeit,  $f : D \subset M \rightarrow \mathbb{R}$ . Die Funktion  $f$  heißt in  $p \in D \subset M$  differenzierbar, wenn für ein offenes Intervall  $I = (a, b)$  mit  $a < f(p) < b$  eine Karte  $\chi : U \subset \mathbb{R}^n \rightarrow M$  um  $p$  mit  $f(\chi(U)) \subset I$  existiert und die Abbildung

$$f \circ \chi : U \subset \mathbb{R}^n \rightarrow \mathbb{R} \quad (81)$$

bei  $\chi^{-1}(p)$  differenzierbar ist. Die Funktion heißt differenzierbar, wenn sie für alle Punkte differenzierbar ist.

Man muss sich jetzt erst mal klar machen, dass diese Definition der Differenzierbarkeit nicht von der Wahl der Karte abhängt.

Jetzt können wir fragen, was für Objekte die Ableitungen denn sind, und dazu definieren wir den Tangentialraum.

# Reelle differenzierbare Mannigfaltigkeiten

## Definition (Tangentialraum an $p \in M$ )

Sei  $M$  eine differenzierbare Mannigfaltigkeit. Eine differenzierbare Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  wird eine (differenzierbare) Kurve in  $M$  genannt. Wir nehmen an, dass  $p = \alpha(0) \in M$ , und sei  $\mathcal{D}$  die Menge aller in  $p$  differenzierbaren Funktionen. **Der Tangentenvektor zu der Kurve  $\alpha$**  bei  $t = 0$  ist eine Funktion  $\alpha'(0) : \mathcal{D} \rightarrow \mathbb{R}$  gegeben durch

$$\alpha'(0)f := \left. \frac{df \circ \alpha}{dt} \right|_{t=0}, \quad f \in \mathcal{D}. \quad (82)$$

Ein **Tangentenvektor an  $p$**  ist der Tangentenvektor bei  $t = 0$  irgendeiner Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  mit  $\alpha(0) = p$ . Die Menge *aller* Tangentenvektoren zu  $M$  an  $p$ , der **Tangentialraum an  $p$** , wird mit  $T_p M$  bezeichnet.

# Reelle differenzierbare Mannigfaltigkeiten

## Definition (Tangentialraum an $p \in M$ )

Sei  $M$  eine differenzierbare Mannigfaltigkeit. Eine differenzierbare Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  wird eine (differenzierbare) Kurve in  $M$  genannt. Wir nehmen an, dass  $p = \alpha(0) \in M$ , und sei  $\mathcal{D}$  die Menge aller in  $p$  differenzierbaren Funktionen. **Der Tangentenvektor zu der Kurve  $\alpha$**  bei  $t = 0$  ist eine Funktion  $\alpha'(0) : \mathcal{D} \rightarrow \mathbb{R}$  gegeben durch

$$\alpha'(0)f := \left. \frac{df \circ \alpha}{dt} \right|_{t=0}, \quad f \in \mathcal{D}. \quad (82)$$

Ein **Tangentenvektor an  $p$**  ist der Tangentenvektor bei  $t = 0$  irgendeiner Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  mit  $\alpha(0) = p$ . Die Menge *aller* Tangentenvektoren zu  $M$  an  $p$ , der **Tangentialraum an  $p$** , wird mit  $T_p M$  bezeichnet.

Um die erste Ableitung zu beschreiben, reicht diese Definition aus. Aber was ist mit der zweiten Ableitung?

# Reelle differenzierbare Mannigfaltigkeiten

Die erste Ableitung läßt sich als Tangentenvektor deuten. Wenn wir eine Parametrisierung  $\mathfrak{x} : U \subset \mathbb{R}^n \rightarrow M$  um  $p = \mathfrak{x}(0)$  wählen, so können wir die Funktionen  $f$  und  $\alpha$  in dieser Parametrisierung durch

$$f \circ \mathfrak{x}(q) = f(x_1, \dots, x_n), \quad q = (x_1, \dots, x_n) \in U \quad (83)$$

und

$$\mathfrak{x}^{-1} \circ \alpha(t) = (x_1(t), \dots, x_n(t)) \quad (84)$$

ausdrücken.

# Reelle differenzierbare Mannigfaltigkeiten

Die erste Ableitung läßt sich als Tangentenvektor deuten. Wenn wir eine Parametrisierung  $\mathfrak{x} : U \subset \mathbb{R}^n \rightarrow M$  um  $p = \mathfrak{x}(0)$  wählen, so können wir die Funktionen  $f$  und  $\alpha$  in dieser Parametrisierung durch

$$f \circ \mathfrak{x}(q) = f(x_1, \dots, x_n), \quad q = (x_1, \dots, x_n) \in U \quad (83)$$

und

$$\mathfrak{x}^{-1} \circ \alpha(t) = (x_1(t), \dots, x_n(t)) \quad (84)$$

ausdrücken.

Wenn wir also  $f$  auf die Kurve  $\alpha$  einschränken, erhalten wir

$$\alpha'(0)f = \frac{d}{dt}(f \circ \alpha) \Big|_{t=0} = \frac{d}{dt}f(x_1(t), \dots, x_n(t)) \Big|_{t=0} \quad (85)$$

$$= \sum_{i=1}^n x'_i(0) \frac{\partial f}{\partial x_i} \Big|_{t=0} = \left( \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0 \right) f. \quad (86)$$

# Reelle differenzierbare Mannigfaltigkeiten

Also können wir den Tangentenvektor  $\alpha'(0)$  darstellen als

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0 \quad (87)$$

# Reelle differenzierbare Mannigfaltigkeiten

Also können wir den Tangentenvektor  $\alpha'(0)$  darstellen als

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0 \quad (87)$$

Die Größen  $\left( \frac{\partial}{\partial x_i} \right)_0$  sind die Tangentenvektoren zum Punkt  $p$  der “Koordinatenkurven”

$$x_i \rightarrow \mathfrak{x}(0, \dots, 0, x_i, 0, \dots, 0). \quad (88)$$

# Reelle differenzierbare Mannigfaltigkeiten

Also können wir den Tangentenvektor  $\alpha'(0)$  darstellen als

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0 \quad (87)$$

Die Größen  $\left( \frac{\partial}{\partial x_i} \right)_0$  sind die Tangentenvektoren zum Punkt  $p$  der “Koordinatenkurven”

$$x_i \rightarrow \mathfrak{x}(0, \dots, 0, x_i, 0, \dots, 0). \quad (88)$$

Damit rechnet man leicht nach, dass die Wahl einer Parametrisierung

$$\mathfrak{x} : U \rightarrow M \quad (89)$$

eine Basis

$$\left\{ \left( \frac{\partial}{\partial x_1} \right)_0, \dots, \left( \frac{\partial}{\partial x_n} \right)_0 \right\} \quad (90)$$

des Vektorraumes  $T_p M$  auswählt.

# Reelle differenzierbare Mannigfaltigkeiten

Jetzt können wir die Ableitung einer Funktion von einer Mannigfaltigkeit in die reellen Zahlen definieren und berechnen:

# Reelle differenzierbare Mannigfaltigkeiten

Jetzt können wir die Ableitung einer Funktion von einer Mannigfaltigkeit in die reellen Zahlen definieren und berechnen:

## Definition (Differential auf einer Mannigfaltigkeit)

Sei  $f : M \rightarrow \mathbb{R}$  eine differenzierbare Abbildung. Für jedes  $p \in M$  und jedes  $v \in T_p M$  wähle man eine differenzierbare Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  mit  $\alpha(0) = p$  und  $\alpha'(0) = v$ . Setze  $\beta = f \circ \alpha$ . Die durch

$$df_p(v) := \beta'(0) \quad (91)$$

definierte Abbildung  $df_p : T_p M \rightarrow \mathbb{R}$  ist eine lineare Abbildung, die nicht von der Wahl von  $\alpha$  abhängt.

# Reelle differenzierbare Mannigfaltigkeiten

Jetzt können wir die Ableitung einer Funktion von einer Mannigfaltigkeit in die reellen Zahlen definieren und berechnen:

## Definition (Differential auf einer Mannigfaltigkeit)

Sei  $f : M \rightarrow \mathbb{R}$  eine differenzierbare Abbildung. Für jedes  $p \in M$  und jedes  $v \in T_p M$  wähle man eine differenzierbare Kurve  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  mit  $\alpha(0) = p$  und  $\alpha'(0) = v$ . Setze  $\beta = f \circ \alpha$ . Die durch

$$df_p(v) := \beta'(0) \quad (91)$$

definierte Abbildung  $df_p : T_p M \rightarrow \mathbb{R}$  ist eine lineare Abbildung, die nicht von der Wahl von  $\alpha$  abhängt.

Jetzt wissen wir, wie die erste Ableitung zu berechnen ist. In den vorherigen Fällen war es dann nicht so schwer, die zweite Ableitung als Ableitung der ersten Ableitung aufzufassen.

# Reelle differenzierbare Mannigfaltigkeiten

Für  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  war die Ableitung im “glatten” Fall wieder eine Funktion  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$ , also kam nichts neues hinzu.

# Reelle differenzierbare Mannigfaltigkeiten

Für  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  war die Ableitung im “glatten” Fall wieder eine Funktion  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$ , also kam nichts neues hinzu.

Für den Fall  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  war die Ableitung auch eine Funktion  $\nabla f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , dieses Mal ein Zeilenvektor, der den Spaltenvektor der Variablen wieder auf eine reelle Zahl abbildet. Um die zweite Ableitung zu berechnen, mussten alle Komponenten nach allen Variablen partiell differenziert werden, das Ergebnis war die Hessematrix.

# Reelle differenzierbare Mannigfaltigkeiten

Für  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  war die Ableitung im “glatten” Fall wieder eine Funktion  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$ , also kam nichts neues hinzu.

Für den Fall  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  war die Ableitung auch eine Funktion  $\nabla f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , dieses Mal ein Zeilenvektor, der den Spaltenvektor der Variablen wieder auf eine reelle Zahl abbildet. Um die zweite Ableitung zu berechnen, mussten alle Komponenten nach allen Variablen partiell differenziert werden, das Ergebnis war die Hessematrix.

Für den Fall  $f : M \rightarrow \mathbb{R}$  ist die Ableitung **im Punkt  $p \in M$**  eine Abbildung  $df_p : T_p M \rightarrow \mathbb{R}$ . Um hier eine zweite Ableitung berechnen zu können, also die Änderung der ersten Ableitung wenn man  $p$  variiert, muss irgendwie auch die Änderung in  $T_p M$  erfasst werden ...

# Reelle differenzierbare Mannigfaltigkeiten

Für  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  war die Ableitung im “glatten” Fall wieder eine Funktion  $f' : D \subset \mathbb{R} \rightarrow \mathbb{R}$ , also kam nichts neues hinzu.

Für den Fall  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  war die Ableitung auch eine Funktion  $\nabla f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ , dieses Mal ein Zeilenvektor, der den Spaltenvektor der Variablen wieder auf eine reelle Zahl abbildet. Um die zweite Ableitung zu berechnen, mussten alle Komponenten nach allen Variablen partiell differenziert werden, das Ergebnis war die Hessematrix.

Für den Fall  $f : M \rightarrow \mathbb{R}$  ist die Ableitung **im Punkt  $p \in M$**  eine Abbildung  $df_p : T_p M \rightarrow \mathbb{R}$ . Um hier eine zweite Ableitung berechnen zu können, also die Änderung der ersten Ableitung wenn man  $p$  variiert, muss irgendwie auch die Änderung in  $T_p M$  erfasst werden ...

Um dieses bewerkstelligen zu können, benötigen wir Vektorfelder, das Tangentialbündel sowie die Begriffe der Länge und des Winkels auf einer Mannigfaltigkeit.

# Reelle differenzierbare Mannigfaltigkeiten

Man kann alle Tangentialräume  $T_p M$  zu einem einzigen Objekt zusammenfassen und dann zeigen, dass es ebenfalls eine Mannigfaltigkeit der doppelten Dimension  $2n$  darstellt:

# Reelle differenzierbare Mannigfaltigkeiten

Man kann alle Tangentialräume  $T_p M$  zu einem einzigen Objekt zusammenfassen und dann zeigen, dass es ebenfalls eine Mannigfaltigkeit der doppelten Dimension  $2n$  darstellt:

## Lemma (Tangentialbündel)

Das Tangentialbündel  $TM$  einer reellen Mannigfaltigkeit  $M$  der Dimension  $n$  wird definiert als

$$TM := \{(p, v) : p \in M, v \in T_p M\} \quad (92)$$

und ist eine Mannigfaltigkeit der Dimension  $2n$ .

# Reelle differenzierbare Mannigfaltigkeiten

Man kann alle Tangentialräume  $T_p M$  zu einem einzigen Objekt zusammenfassen und dann zeigen, dass es ebenfalls eine Mannigfaltigkeit der doppelten Dimension  $2n$  darstellt:

## Lemma (Tangentialbündel)

*Das Tangentialbündel  $TM$  einer reellen Mannigfaltigkeit  $M$  der Dimension  $n$  wird definiert als*

$$TM := \{(p, v) : p \in M, v \in T_p M\} \quad (92)$$

*und ist eine Mannigfaltigkeit der Dimension  $2n$ .*

Die Ableitung einer Funktion auf einer Mannigfaltigkeit liefert für jeden Punkt  $p \in M$  einen Tangentenvektor, also ein Objekt aus dem an  $p$  "angekoppelten" Tangentialraum. Allgemein bezeichnet man eine Auswahl je eines Tangentialvektors aus jedem Tangentialraum als Schnitt und die resultierende Menge von Tangentenvektoren als Vektorfeld.

# Reelle differenzierbare Mannigfaltigkeiten

Die Frage nach der zweiten Ableitung einer Funktion auf einer Mannigfaltigkeit ist also die Frage nach der Ableitung eines Vektorfeldes.

# Reelle differenzierbare Mannigfaltigkeiten

Die Frage nach der zweiten Ableitung einer Funktion auf einer Mannigfaltigkeit ist also die Frage nach der Ableitung eines Vektorfeldes.

Um die Änderung eines Vektorfeldes mit  $p$  zu erfassen, muss man irgendwie einen Zusammenhang zwischen Tangentenvektoren verschiedener Tangentialräume  $T_p M$  herstellen.

# Reelle differenzierbare Mannigfaltigkeiten

Die Frage nach der zweiten Ableitung einer Funktion auf einer Mannigfaltigkeit ist also die Frage nach der Ableitung eines Vektorfeldes.

Um die Änderung eines Vektorfeldes mit  $p$  zu erfassen, muss man irgendwie einen Zusammenhang zwischen Tangentenvektoren verschiedener Tangentialräume  $T_p M$  herstellen.

Die Menge aller glatten Schnitte, also glatten Vektorfelder  $X$ , bezeichnen wir mit  $\mathcal{X}(M) = C^\infty(M, TM)$ . In einer Parametrisierung  $\mathfrak{x} : U \subset \mathbb{R}^n \rightarrow M$  gilt

$$X(p) = \sum_{i=1}^n a_i(p) \frac{\partial}{\partial x_i}, \quad (93)$$

Differenzierbarkeit folgt aus der Differenzierbarkeit der  $a_i$ .

# Reelle differenzierbare Mannigfaltigkeiten

Die Frage nach der zweiten Ableitung einer Funktion auf einer Mannigfaltigkeit ist also die Frage nach der Ableitung eines Vektorfeldes.

Um die Änderung eines Vektorfeldes mit  $p$  zu erfassen, muss man irgendwie einen Zusammenhang zwischen Tangentenvektoren verschiedener Tangentialräume  $T_p M$  herstellen.

Die Menge aller glatten Schnitte, also glatten Vektorfelder  $X$ , bezeichnen wir mit  $\mathcal{X}(M) = C^\infty(M, TM)$ . In einer Parametrisierung  $\mathfrak{x} : U \subset \mathbb{R}^n \rightarrow M$  gilt

$$X(p) = \sum_{i=1}^n a_i(p) \frac{\partial}{\partial x_i}, \quad (93)$$

Differenzierbarkeit folgt aus der Differenzierbarkeit der  $a_i$ .

Gleichung (93) ist eine Abbildung  $X : \mathcal{D} \rightarrow \mathcal{F}$  differenzierbarer Funktionen auf Funktionen:

$$(Xf)(p) = \sum_{i=1}^n a_i(p) \frac{\partial f}{\partial x_i}(p). \quad (94)$$

# Reelle differenzierbare Mannigfaltigkeiten

Wir waren auf der Suche nach dem Zusammenhang. Solch ein Zusammenhang heißt “affiner Zusammenhang”. Ein affiner Zusammenhang hat viele Gemeinsamkeiten mit dem Differentiationsoperator:

# Reelle differenzierbare Mannigfaltigkeiten

Wir waren auf der Suche nach dem Zusammenhang. Solch ein Zusammenhang heißt “affiner Zusammenhang”. Ein affiner Zusammenhang hat viele Gemeinsamkeiten mit dem Differentiationsoperator:

## Definition

Ein affiner Zusammenhang auf  $M$  ist eine bilineare Abbildung

$$\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \rightarrow \mathcal{X}(M), \quad (95)$$

$$\nabla : (X, Y) \rightarrow \nabla_X Y, \quad (96)$$

so dass für alle  $f, g \in C^\infty(M, \mathbb{R})$  und alle Vektorfelder  $X, Y, Z$  auf  $M$

$$\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z, \quad (\text{Linearität Argument Eins})$$

$$\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z, \quad (\text{Linearität Argument Zwei})$$

$$\nabla_X(fY) = f\nabla_X Y + X(f)Y. \quad (\text{Leibniz-Regel})$$

# Reelle differenzierbare Mannigfaltigkeiten

Eine Riemannsche Mannigfaltigkeit ist eine reelle differenzierbare Mannigfaltigkeit mit einer Riemannschen Metrik.

# Reelle differenzierbare Mannigfaltigkeiten

Eine Riemannsche Mannigfaltigkeit ist eine reelle differenzierbare Mannigfaltigkeit mit einer Riemannschen Metrik.

## Definition

Eine Riemannsche Metrik  $g$  auf einer reellen differenzierbaren Mannigfaltigkeit  $M$  ist ein inneres Produkt  $\langle \cdot, \cdot \rangle_p$  (eine symmetrische, bilineare, positive definite Form) auf dem Tangentialraum  $T_p M$ , welches differenzierbar in dem folgenden Sinne variiert:

Wenn  $\mathfrak{x} : U \subset \mathbb{R}^n \rightarrow M$  ein Koordinatensystem um  $p \in M$  mit  $\mathfrak{x}(x_1, \dots, x_n) = q \in \mathfrak{x}(U)$  und  $\frac{\partial}{\partial x_i}(q) = dx_q(0, \dots, 1, \dots, 0)$  ist, dann ist

$$\left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q = g_{ij}(x_1, \dots, x_n) \quad (97)$$

eine differenzierbare Funktion auf  $U$ .

# Reelle differenzierbare Mannigfaltigkeiten

Ein affiner Zusammenhang kann aus der Wirkung auf eine Basis  $X_i$  erschlossen werden. Die Koeffizienten  $\Gamma_{ij}^k$  werden Christoffel-Symbole genannt:

$$\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k. \quad (98)$$

# Reelle differenzierbare Mannigfaltigkeiten

Ein affiner Zusammenhang kann aus der Wirkung auf eine Basis  $X_i$  erschlossen werden. Die Koeffizienten  $\Gamma_{ij}^k$  werden Christoffel-Symbole genannt:

$$\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k. \quad (98)$$

Zu einer solchen Metrik existiert auch immer etwas, was die Rolle einer Differentiation übernimmt, nämlich der sogenannte Levi-Civita oder Riemannsche Zusammenhang.

# Reelle differenzierbare Mannigfaltigkeiten

Ein affiner Zusammenhang kann aus der Wirkung auf eine Basis  $X_i$  erschlossen werden. Die Koeffizienten  $\Gamma_{ij}^k$  werden Christoffel-Symbole genannt:

$$\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k. \quad (98)$$

Zu einer solchen Metrik existiert auch immer etwas, was die Rolle einer Differentiation übernimmt, nämlich der sogenannte Levi-Civita oder Riemannsche Zusammenhang.

## Theorem (Levi-Civita Zusammenhang)

*Es gibt einen eindeutigen mit der Metrik kompatiblen und symmetrischen affinen Zusammenhang, welcher in Koordinaten durch*

$$\sum_{\ell=1}^n \Gamma_{ij}^{\ell} g_{\ell k} = \frac{1}{2} \left( \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right) \quad (99)$$

*gegeben ist.*

# Reelle differenzierbare Mannigfaltigkeiten

Viele Optimierungsaussagen lassen sich auf reelle Riemannsche Mannigfaltigkeiten übertragen. Dazu gehören das Newton-Verfahren zur Minimierung einer reellwertigen Funktion inklusive einem angepassten Satz von Kantorovich und Smales Alpha-Test.

# Reelle differenzierbare Mannigfaltigkeiten

Viele Optimierungsaussagen lassen sich auf reelle Riemannsche Mannigfaltigkeiten übertragen. Dazu gehören das Newton-Verfahren zur Minimierung einer reellwertigen Funktion inklusive einem angepassten Satz von Kantorovich und Smales Alpha-Test.

Leider ist nicht mehr genug Zeit um dieses interessante Gebiet hinreichend zu vertiefen . . .

# Reelle differenzierbare Mannigfaltigkeiten

Viele Optimierungsaussagen lassen sich auf reelle Riemannsche Mannigfaltigkeiten übertragen. Dazu gehören das Newton-Verfahren zur Minimierung einer reellwertigen Funktion inklusive einem angepassten Satz von Kantorovich und Smales Alpha-Test.

Leider ist nicht mehr genug Zeit um dieses interessante Gebiet hinreichend zu vertiefen . . .

Danke für die Aufmerksamkeit ; -)

# Reelle differenzierbare Mannigfaltigkeiten

Viele Optimierungsaussagen lassen sich auf reelle Riemannsche Mannigfaltigkeiten übertragen. Dazu gehören das Newton-Verfahren zur Minimierung einer reellwertigen Funktion inklusive einem angepassten Satz von Kantorovich und Smales Alpha-Test.

Leider ist nicht mehr genug Zeit um dieses interessante Gebiet hinreichend zu vertiefen . . .

Danke für die Aufmerksamkeit ; -)

War nett, mal wieder die Erstsemester verschaukeln zu dürfen ; -)

Immer wieder gerne geschehen ; -)



Königsberger

Analysis 1. Zweite Auflage.

Springer, Berlin, 1992.



Königsberger

Analysis 2. Zweite Auflage.

Springer, Berlin, 1993.



Ebbinghaus et al.

Zahlen. Dritte Auflage.

Springer, Berlin, 1992.



Manfredo Perdigão do Carmo

Riemannian Geometry. Second Printing.

Birkhäuser, Boston, 1993.