Lineare und nichtlineare Optimierung

Vorlesung - Teil 1

Sven-Joachim Kimmerle

Physical Software Solutions &
Lehrbeauftragter (extern)
Institut für Angewandte Mathematik und Wissenschaftliches Rechnen (LRT-1)
Universität der Bundeswehr München, Neubiberg/München

Vorlesung HT 2019
Studiengänge BSc. Luft- und Raumfahrttechnik, BSc. ME

Organisatorisches

- 2-stündige Vorlesung mit 1-stündiger Übung:
  - Mo, 15:00-16:30 in 033-3131
  - Mi, 08:00-09:30 in 033-2211 (oder z.B. Di, 08:00-09:30 ?), 14-tägig ab Mi, 09.10.2019
- Vorlesung & Übung:

  Dr. habil. Sven-Joachim Kimmerle

  sven-joachim.kimmerle@unibw.de, Durchwahl 2129

  Sprechzeiten: flexibel nach Vereinbarung, Büro 041-2304
- Voraussetzungen:
  - Ingenieurmathematik 1-3
  - Programmierkenntnisse in einer beliebigen prozeduralen Programmiersprache (z.B. MATLAB) wünschenswert
- Prüfung: mündlich
- Internet:

  www.unibw.de/ingmathe/teaching/linearenichtlineareoptimierung

Empfohlene Literatur

[CG99] Geiger, C. und Kanzow, C.: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer-Verlag, Berlin/Heidelberg/New York, 1999.

[CG02] Geiger, C. und Kanzow, C.: *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer-Verlag, Berlin/Heidelberg/New York, 2002.

[GL11] Gerdts, M. und Lempio, F.: *Mathematische Optimierungsverfahren des Operations Research.* DeGruyter, Berlin, 2011.

[Dr15] Dreves, A.: *Lineare und nichtlineare Optimierung*, Vorlesungsskript, BSc.-Studiengang "Luft- und Raumfahrttechnik", Universität der Bundeswehr München, 2015.

[Ge16] Gerdts, M.: *Einführung in die lineare und nichtlineare Optimierung*, Vorlesungsskript, BSc.-Studiengang "Luft- und Raumfahrttechnik", Universität der Bundeswehr München, 2016.

Inhaltsverzeichnis

Inhaltsverzeichnis

Motivation

- Optimierungsaufgaben finden sich:
  - in den Wirtschaftswissenschaften
  - im Management
  - in der Technik
  - in der Produktion
  - in der Bildverarbeitung
  - in der Natur
  - etc.
- Optimierung steht in engem Zusammenhang zur (mathematischen) Modellierung
- Idee: Zielfunktion wird unter Nebenbedingungen extremal
- Ohne Einschränkung betrachten wir nur Minimierungsprobleme

  Durch Multiplikation der Zielfunktion mit -1 wird ein Maximierungsproblem in ein äquivalentes Minimierungsproblem transformiert

Allgemeines Optimierungsproblem

Problem (Allgemeines Optimierungsproblem (OP))

Minimiere $f(x)$ *unter den Nebenbedingungen* $x \in X$.

Dabei sei $X \subseteq \mathbb{R}^n$ *eine nichtleere Menge und*

$f : X \to \mathbb{R}$ *eine Funktion.*

f heißt **Zielfunktion**.

x heißt **zulässig** für (OP), falls $x \in X$.

X heißt **zulässige Menge** von (OP).

Verschiedene Klassen von Optimierungsproblemen I

Bei einem **unrestringierten Optimierungsproblem (UOP)** gilt $X = \mathbb{R}^n$.

Oft sind die Restriktionen durch (Un-)Gleichungen gegeben:

Problem (Standard-Optimierungsproblem (SOP))

Minimiere $f(x)$ unter den Nebenbedingungen (u. d. N.)

$$\begin{aligned} g_i(x) &\le 0, \quad i = 1, \dots, m, \\ h_j(x) &= 0, \quad j = 1, \dots, p. \end{aligned}$$

bzw. in Vektornotation

$$\begin{aligned} g(x) &\le 0, \\ h(x) &= 0. \end{aligned}$$

Dabei seien $x \in \mathbb{R}^n$ und die Funktionen

$$\begin{aligned} f &: \mathbb{R}^n \to \mathbb{R}, \\ g = (g_1, \dots, g_m)^\top &: \mathbb{R}^n \to \mathbb{R}^m \text{ und} \\ h = (h_1, \dots, h_p)^\top &: \mathbb{R}^n \to \mathbb{R}^p. \end{aligned}$$

Verschiedene Klassen von Optimierungsproblemen II

Wichtiger Spezialfall:

Problem (Lineares Optimierungsproblem (LOP) (in primaler Normalform))

Minimiere $c^\top x$ *u. d. N.*

$$\begin{aligned} x &\geq 0, \\ Ax &= b. \end{aligned}$$

Dabei seien:

$$\begin{aligned} x, c &: \in \mathbb{R}^n, \\ b &: \in \mathbb{R}^p \text{ und} \\ A &: \in \mathbb{R}^{p \times n}. \end{aligned}$$

Ergibt sich aus (SOP) für

$$f(x) = c^\top x, \quad g(x) = -x, \quad h(x) = Ax - b.$$

Verschiedene Klassen von Optimierungsproblemen III

Falls im Optimierungsproblem die Zielfunktion quadratisch ist, z.B.

$$f(x) = \frac{1}{2}(x - x_{soll})^2$$

und ansonsten die Situation von (LOP) vorliegt, dann ist dies ein **linear-quadratisches Optimierungsproblem** (LQOP).

Es gibt weitere Klassen von Optimierungsproblemen (OP):

- Vektoroptimierungsprobleme
- unendlichdimensionale OP
- semi-infinite OP
- ganzzahlige OP
- gemischt-ganzzahlige OP
- nicht-differenzierbare OP
- ...

Extrema: Minima und Maxima

Definition (Minima)

- $\hat{x} \in X$ heißt **globales Minimum** von (OP), falls
$$f(\hat{x}) \leq f(x) \quad \text{für alle } x \in X. \tag{1.1}$$
- $\hat{x} \in X$ heißt **striktes globales Minimum** von (OP), falls in (1.1) “<” für alle $x \in X$ mit Ausnahme von $x = \hat{x}$ gilt.
- $\hat{x} \in X$ heißt **lokales Minimum** von (OP), falls es für ein $\varepsilon > 0$ eine Umgebung
$$U_\varepsilon(\hat{x}) := \{x \in \mathbb{R}^n \mid \|x - \hat{x}\| < \varepsilon\}$$
gibt mit
$$f(\hat{x}) \leq f(x) \quad \text{für alle } x \in X \cap U_\varepsilon(\hat{x}). \tag{1.2}$$
- $\hat{x} \in X$ heißt **striktes globales Minimum** von (OP), falls in (1.2) “<” für alle $x \in X \cap U_\varepsilon(\hat{x})$ mit Ausnahme von $x = \hat{x}$ gilt.

Analog werden die entsprechenden Maxima definiert.

Typische Fragestellungen

i) Existenz zulässiger Lösungen?
ii) Existenz von Optimallösungen?
iii) Eindeutigkeit der Optimallösung?
iv) Abhängigkeit der Optimallösungen von Problemdaten?
v) Welche Eigenschaften besitzen Optimallösungen, d.h. welche Bedingungen sind notwendig?
vi) Welche Bedingungen sind hinreichend dafür, dass eine zulässige Lösung eine Optimallösung ist?
vii) Welche Bedingungen sind gleichzeitig notwendig und hinreichend?
viii) Wie gewinnt man Schranken für den Optimalwert bzw. Fehlerabschätzungen für die Optimallösung?
ix) Welche Algorithmen gibt es zur Berechnung einer Optimallösung?
x) Welche numerische Eigenschaften haben diese?

i) - vii) sind eher theoretischer Natur, aber ohne diese Fragestellungen ist die Beantwortung der numerischen Aspekte viii) - x), um die es in dieser VL gehen soll, nicht möglich.

Grund-Annahme

In dieser Vorlesung setzen wir die Existenz von Optimallösungen stets voraus.

Es gilt nämlich

Theorem (Satz von Weierstraß)

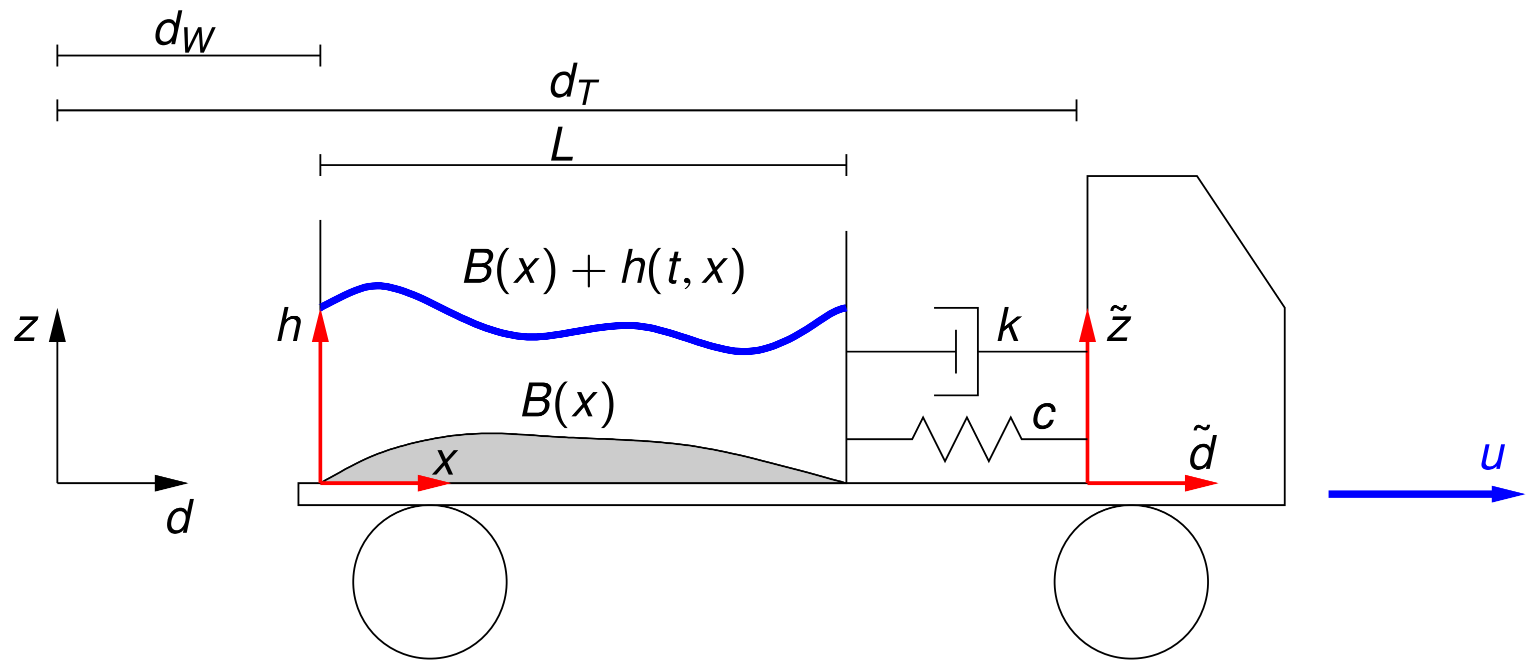
Sei $X \subseteq D \subseteq \mathbb{R}^n$ kompakt (d.h. abgeschlossen und beschränkt) und

$f : D \to \mathbb{R}$ stetig.

Dann nimmt f das Minimum (bzw. Maximum) auf X an.

Beispiele: Ein Optimalsteuerungsproblem [K., Gerdts 2015 -]

Truck with a water container: horizontal control force u, parameter final time T



m_T: mass of truck, m_W: mass of water container, d_T: distance truck, d_W: distance water container, $B(x)$: bottom elevation of container, $h(t,x)$: water level relative to bottom at time t and position x, $v(t,x)$: horizontal water velocity at time t and position x, L: length of water container, c : spring force constant, k: damper force constant

Minimierung mithilfe graphischer Darstellung

Beispiel (Funktion von Himmelblau)

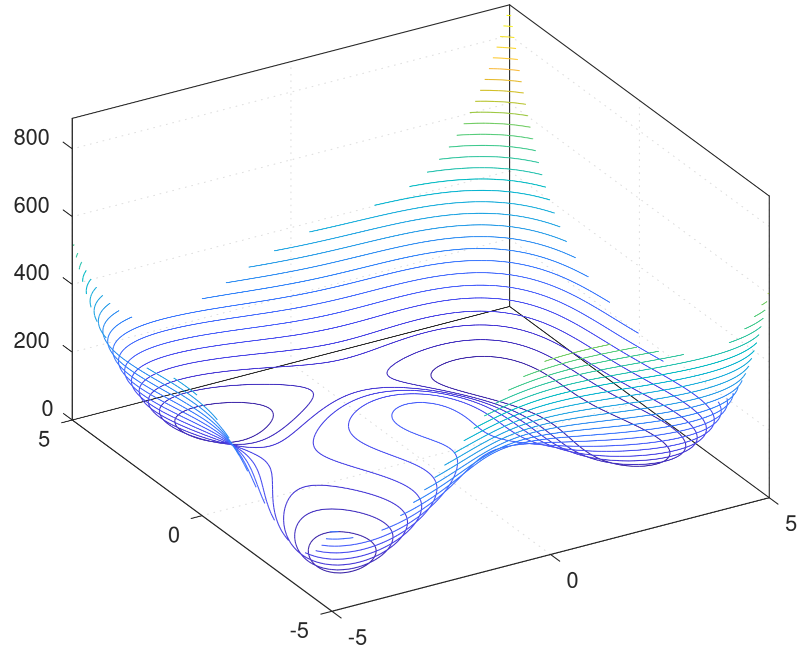
Minimiere

$$f(x,y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

über $(x,y)^\top \in \mathbb{R}^2$.

MATLAB-Code:

```
[x, y] = meshgrid(-5:0.1:5);
z = (x.^2 + y - 11).^2 + (x + y.^2 - 7).^2;
contour3(x, y, z, 30)
```



Höhenlinien

Eine **Höhenlinie (Niveaulinie)** $N_f(c)$ zum Niveau $c \in \mathbb{R}$ für eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist formal definiert als die Menge aller Punkte $x = (x_1, \ldots, x_n)^\top$ für die $f(x) = c$ gilt:

$$N_f(c) := \{x \in \mathbb{R}^n \mid f(x) = c\}, \quad c \in \mathbb{R}.$$

Auf Höhenlinien von f steht der **Gradient**

$$\nabla f(x_1, \ldots, x_n) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_1, \ldots, x_n) \end{pmatrix}$$

senkrecht.

∇f ist ein Spaltenvektor. Er zeigt in die Richtung des steilsten Anstiegs.

Für reellwertige Funktionen mit einem Definitionsbereich in 1D oder 2D, kann die graphische Methode ganz hilfreich sein.

Klassisches lineares Optimierungsproblem I

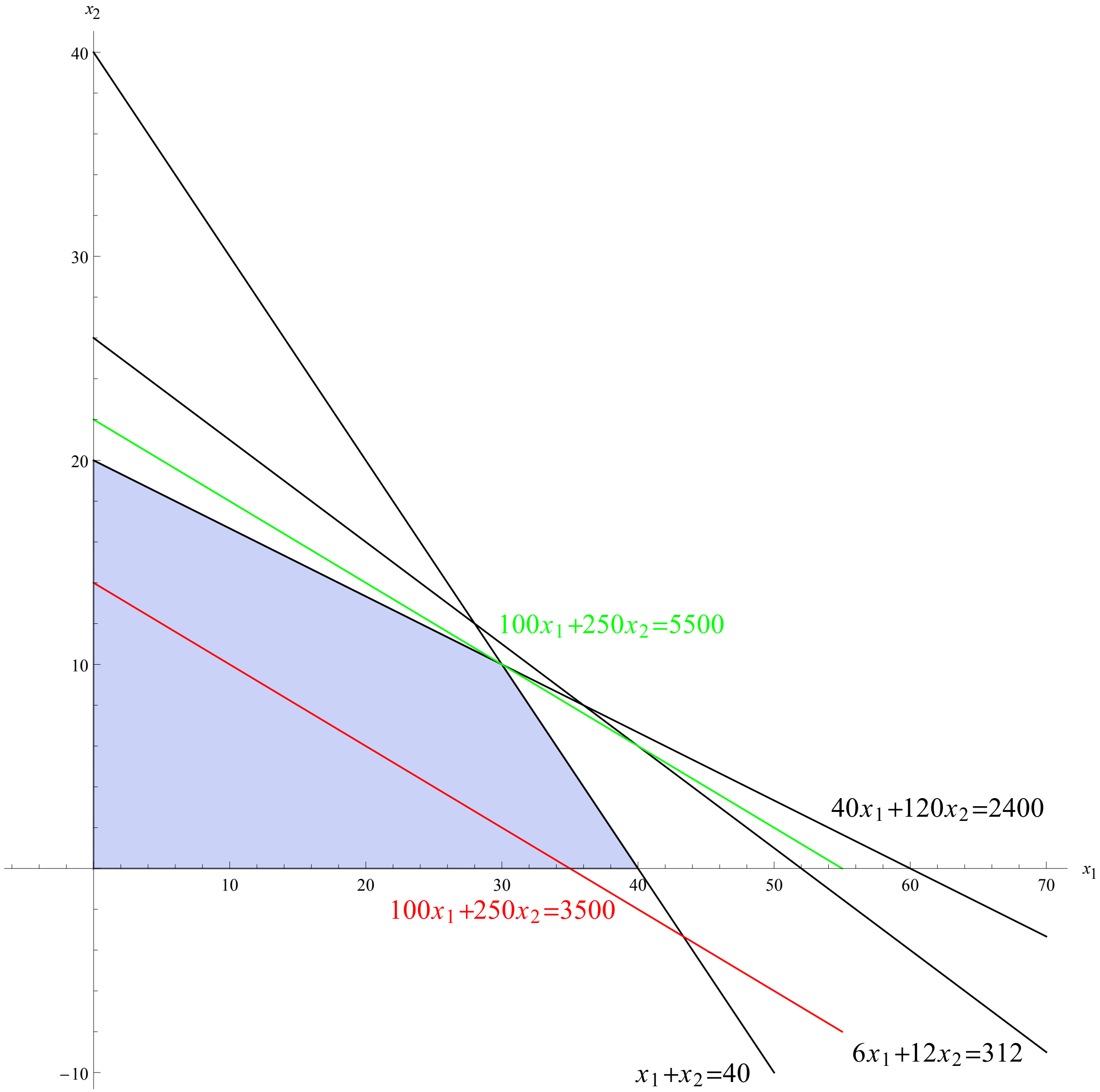
Beispiel (Agrarwirtschaft)

Ein Landwirt bewirtschaftet ein Grundstück von 40 ha Größe mit Zuckerrüben und Weizen.
Er kann hierzu 2400 € und 312 Arbeitstage einsetzen.

Die Anbaukosten betragen bei Rüben 40 €/ha und bei Weizen 120 €/ha.
Für Rüben benötigt er 6 Arbeitstage/ha und für Weizen 12 Arbeitstage/ha.
Der Reingewinn sei bei Rüben 100 €/ha und bei Weizen 250 €/ha.

Anmerkung:
Wir dürfen bereits vorneweg annehmen, dass die zur Verfügung stehenden 2400 € sofort auszugeben sind und nicht anderweitig gespart bzw. angelegt werden können und somit auch nicht in die Zielfunktion eingehen.
Wie wir sehen werden, rechtfertigt die gefundene Optimallösung diese Annahme.

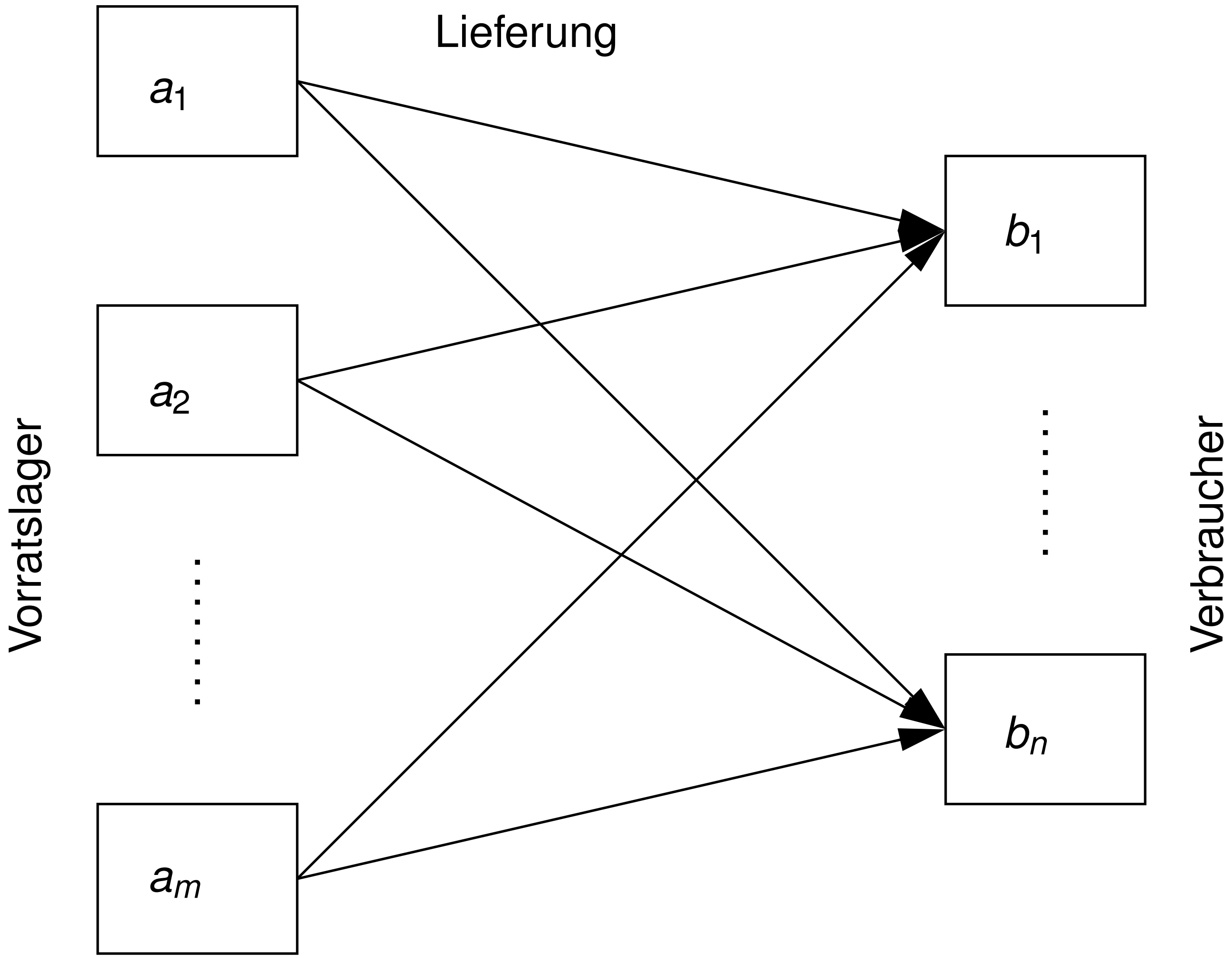
Klassisches lineares Optimierungsproblem II



Transportproblem I

Beispiel (Transportproblem)

Ein Transportunternehmer hat m Vorratslager, aus denen n Verbraucher mit einem Produkt beliefert werden können. Die Lieferkosten von Lager i zu Verbraucher j betragen c_{ij} Einheiten pro Produkteinheit. In Lager i sind a_i Einheiten des Produktes vorrätig. Verbraucher j hat einen Bedarf von b_j Einheiten. Um die Kunden nicht zu verärgern, muss der Lieferant den Bedarf der Kunden befriedigen. Andererseits möchte der Lieferant seine Lieferkosten minimieren.



Transportproblem II

Bezeichnet x_{ij} die Liefermenge von Lager i zu Verbraucher j, so führt das Problem auf das folgende Transportproblem, welches ein spezielles lineares Optimierungsproblem ist:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad \text{u.d.N.}$$

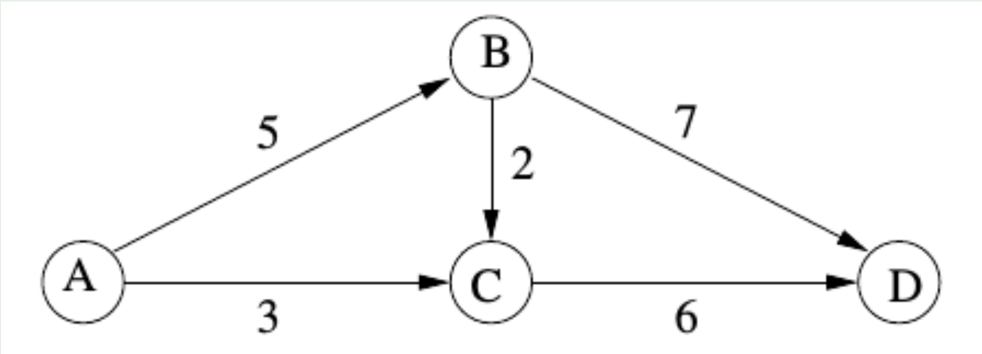
$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &\le a_i, \quad && i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} &\ge b_j, \quad && j = 1, \ldots, n, \\ x_{ij} &\ge 0, \quad && i = 1, \ldots, m,\ j = 1, \ldots, n. \end{aligned}$$

Die erste Nebenbedingung besagt, dass aus Lager i maximal a_i Einheiten abtransportiert werden können.
Die zweite Nebenbedingung besagt, dass der Bedarf b_j befriedigt werden muss.
Die letzte Nebenbedingung verbietet negative Liefermengen.

Graphenproblem

Beispiel (Netzwerkproblem)

Eine Firma möchte möglichst viele Waren von A nach D über das abegebildete Straßennetzwerk transportieren.



Die Zahlen neben den Kanten des Netzwerk-Graphen geben die maximale Kapazität der Kante an.

Wie kann dieses Problem mathematisch modelliert werden?

Portfoliooptimierung I

Beispiel (Portfoliooptimierungsaufgabe nach Markowitz)

Gegeben seien $j = 1, \ldots, n$ mögliche Anlagen. Jede Anlage wirft im nächsten Zeitraum einen Gewinn (bzw. Verlust) R_j ab. R_j ist allerdings nicht bekannt, aber die zugrundeliegende zufällige Verteilung.
Um einerseits den Gewinn zu maximieren und andererseits das Risiko eines Verlusts zu minimieren, wird die Anlagesumme zu prozentualen Anteilen x_j auf die n Anlagen verteilt.
Die Aufgabe eines Portfoliomanagers besteht in der optimalen Zusammenstellung des Portfolios $(x_1, \ldots, x_n)$.

Ein mögliches Ziel ist es, den Erwartungswert des Gewinns R zu maximieren:

$$E(R) = \sum_{j=1} x_j E(R_j), \quad R = \sum_{j=1}^{n} x_j R_j$$

Als Maß für das Risiko kann die Varianz von R betrachtet werden:

$$Var(R) = E(R - E(R))^2 = ER^2 - (ER)^2$$

Portfoliooptimierung II

Ein Kompromiss ist das folgende Optimierungsproblem

$$\min \; -E(R) + \alpha Var(R) \qquad \text{u. d. N.}$$

$$\sum_{j=1}^{n} x_j = 1,$$

$$x_j \geq 0, \quad j = 1, \ldots, n.$$

$\alpha \geq 0$ ist dabei ein noch zu wählender Gewichtungsparameter.

Desto kleiner α, desto größer das Risiko.

Ausgleichsrechnung und Parameteridentifikation

... später ausführlich

Inhaltsverzeichnis

Unrestringierte Optimierung

Wir betrachten hier

Problem (Unrestringiertes Optimierungsproblem (UOP))

Minimiere $f(x)$ *u. d. N.* $x \in \mathbb{R}^n$.

Dabei sei $f : \mathbb{R}^n \to \mathbb{R}$ *eine mindestens einmal stetig differenzierbare Funktion.*

Ziele:
Notwendige Bedingungen
Hinreichende Bedingungen

Wünschenswert:
Notwendige und hinreichende Bedingungen

Dann:
Numerische Verfahren (meist basierend auf notwendigen Bedingungen)

Rekapitulation I

- Der **Gradient** einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ an der Stelle $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ ist definiert als Spaltenvektor

$$\nabla f(x_1, \dots, x_n) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_1, \dots, x_n) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_1, \dots, x_n) \end{pmatrix}.$$

  Im Fall $n = 1$ ist der Gradient die 1. Ableitung von f an der Stelle x, kurz $f'(x)$.

  Anschaulich beschreibt der Gradient den Anstieg.

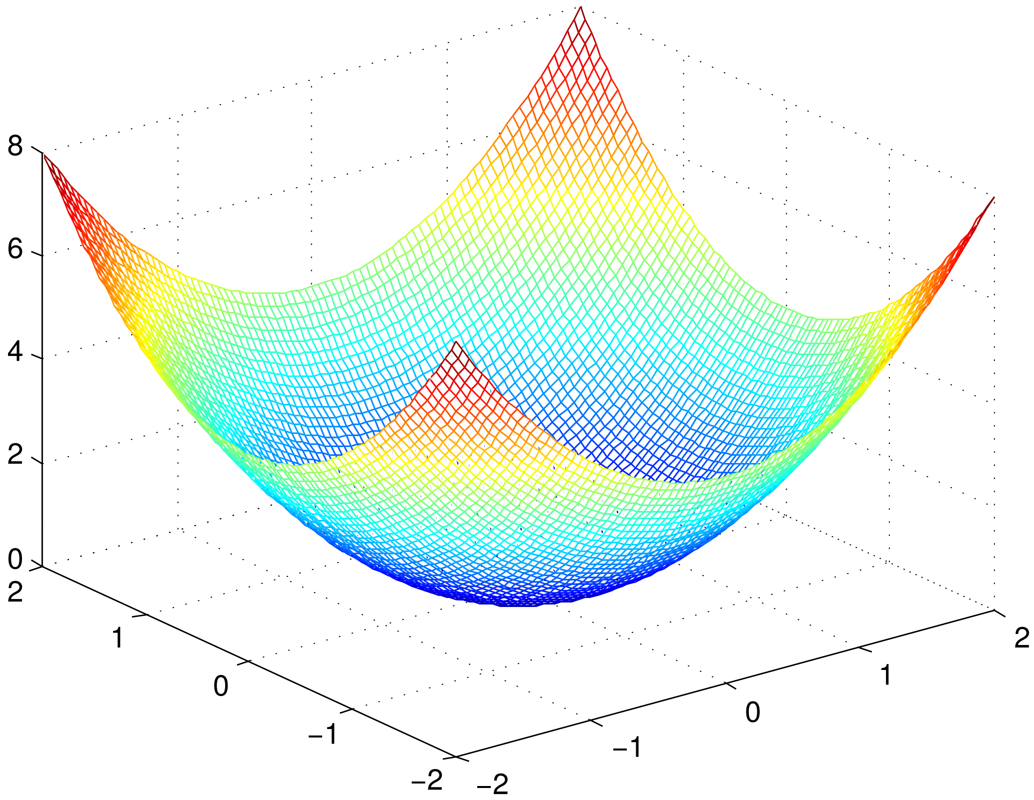
- Die **Hessematrix** einer Funktion f an der Stelle x ist definiert durch

$$H_f(x_1, \dots, x_n) = H_f(x) = H(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}.$$

  Im Fall $n = 1$ ist die Hessematrix die 2. Ableitung von f an der Stelle x, kurz $f''(x)$.

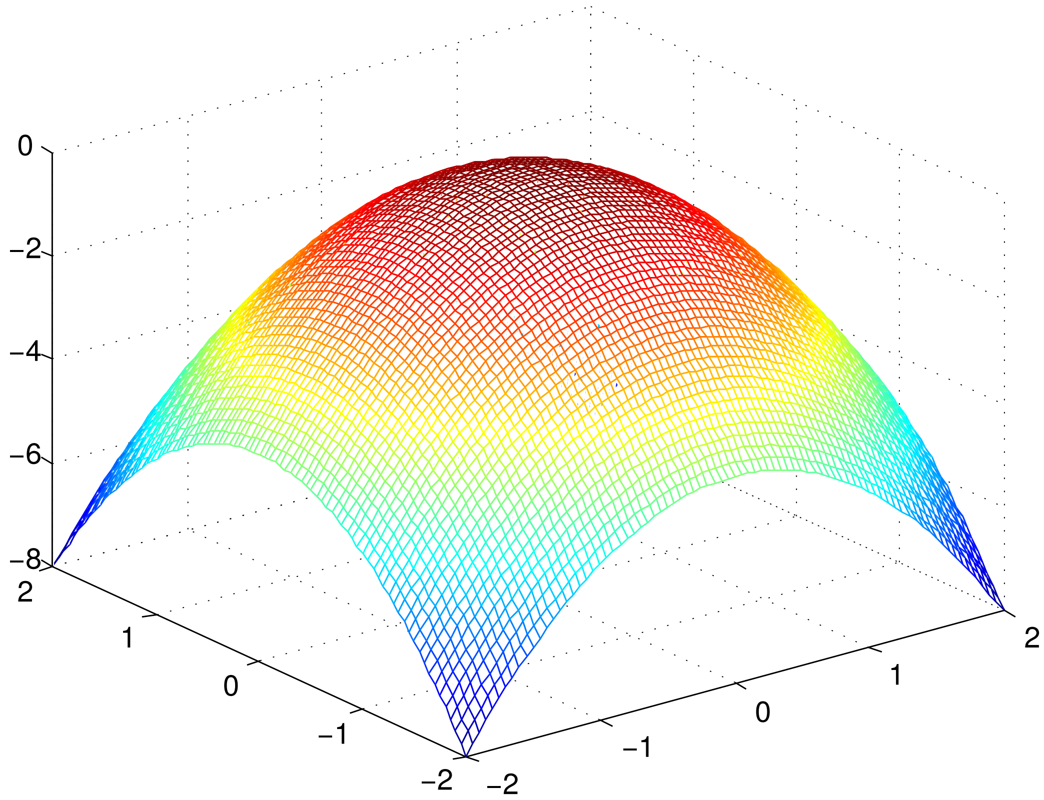
  Anschaulich beschreibt die Hessematrix die lokale Krümmung einer Funktion.

Rekapitulation II - Hessematrix I



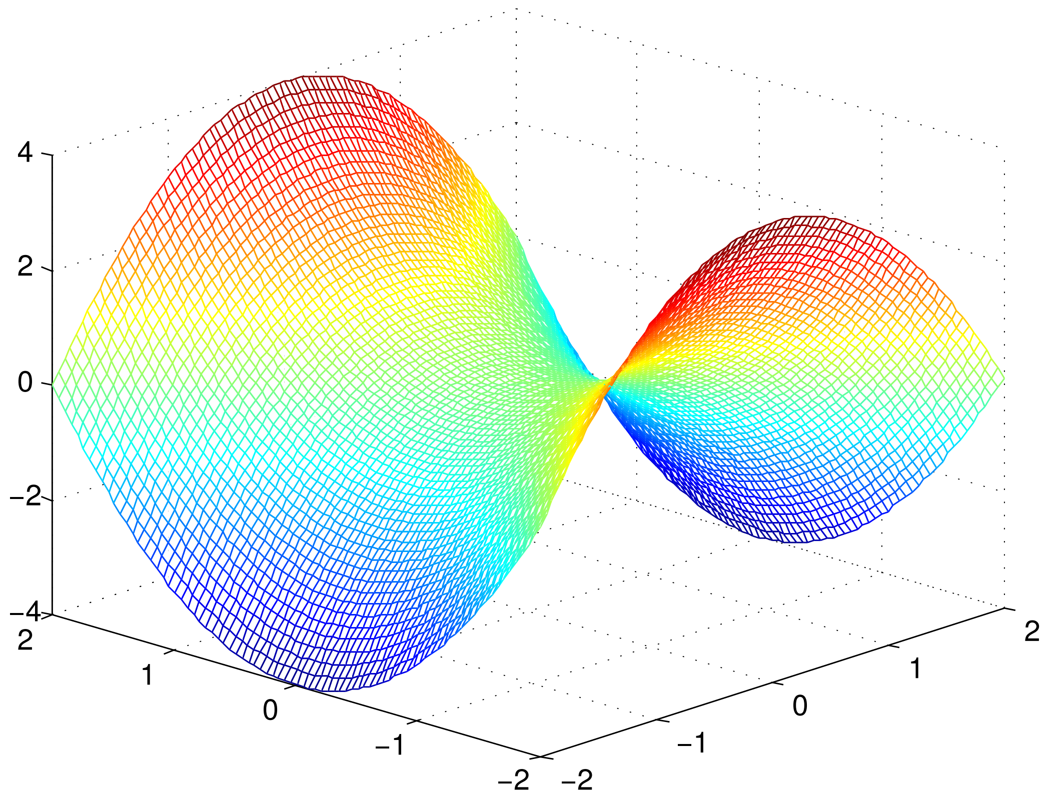
$$f(x,y) = x^2 + y^2, \qquad H_f(\hat{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Rekapitulation II - Hessematrix II



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

Rekapitulation II - Hessematrix III



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

Rekapitulation III

Richtungsableitung

Die Richtungsableitung $f'(x; d)$ einer stetig differenzierbaren Funktion $f : \mathbb{R}^n \to \mathbb{R}$ an der Stelle $x \in \mathbb{R}^n$ in Richtung $d \in \mathbb{R}^n$ ist definiert durch

$$f'(x; d) := \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} = \nabla f(x)^\top d.$$

Rekapitulation IV.I - Taylorentwicklung

Taylorentwicklung

- Ist f stetig differenzierbar, dann kann f in einer Umgebung von $\hat{x}$ durch eine **affin-lineare Funktion** approximiert werden, da

$$f(x) = f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}) + o(\|x - \hat{x}\|),$$

wobei

$$lim_{x \to \hat{x}} \frac{o(\|x - \hat{x}\|)}{\|x - \hat{x}\|} = 0.$$

bzw. alternativ mit $t \in \mathbb{R}$, $d \in \mathbb{R}^n$, $\xi_t \in (0, t)$

$$f(\hat{x} + td) = f(\hat{x}) + t \nabla f(\hat{x} + \xi_t d)^\top d.$$

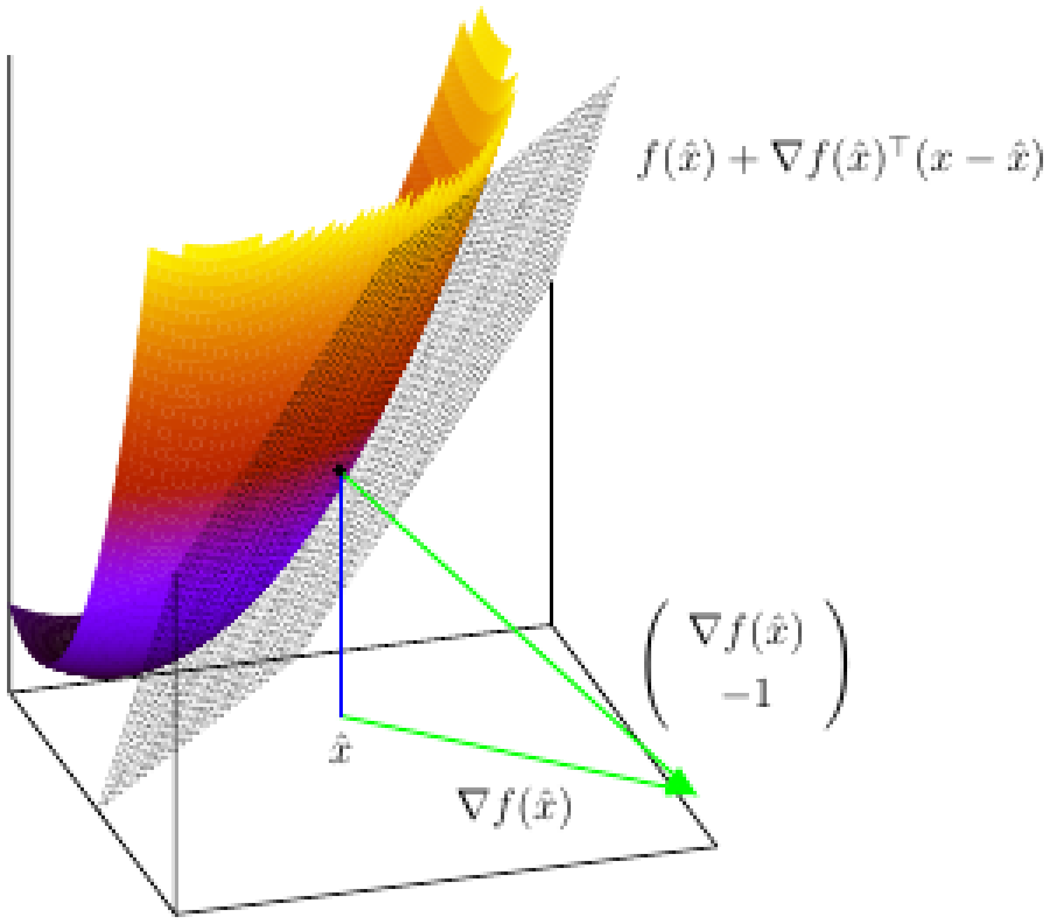
- Ist f 2mal stetig differenzierbar, dann kann f in einer Umgebung von $\hat{x}$ durch eine quadratische Funktion approximiert werden, da

$$f(x) = f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}) + \frac{1}{2}(x - \hat{x})^\top H_f(\hat{x})(x - \hat{x}) + o(\|x - \hat{x}\|^2).$$

bzw. alternativ

$$f(\hat{x} + td) = f(\hat{x}) + t \nabla f(\hat{x})^\top d + \frac{t^2}{2} d^\top H_f(\hat{x} + \xi_t d) d.$$

Rekapitulation IV.II - Affin-lineare Approximation in 2D: Tangentialebene



Notwendige/hinreichende Bedingungen: Notw. Bed. 1. Ordnung

Theorem (Notwendige Bedingung 1. Ordnung)

Sei $f : \mathbb{R}^n \to \mathbb{R}$ *stetig differenzierbar und* $\hat{x} \in \mathbb{R}^n$ *ein lokales Minimum von* f*.*

Dann gilt

$$\nabla f(\hat{x}) = 0.$$

Definition (Stationärer Punkt)

Jeder Punkt $x \in \mathbb{R}^n$ mit $\nabla f(x) = 0$ heißt **stationärer Punkt** von $f : \mathbb{R}^n \to \mathbb{R}$.

Stationäre Punkte sind nicht automatisch (lokale) Minima, sondern nur potentielle Kandidaten!

Die meisten numerischen Verfahren versuchen stationäre Punkte zu approximieren.

Notwendige Bedingung zweiter Ordnung

Theorem (Notwendige Bedingung 2. Ordnung)

Sei $f : \mathbb{R}^n \to \mathbb{R}$ 2mal stetig differenzierbar und $\hat{x} \in \mathbb{R}^n$ ein lokales Minimum von f.

Dann ist die Hessematrix

$$H_f(\hat{x})$$

positiv semidefinit.

Im Fall $n = 1$ gilt

$$f''(\hat{x}) \geq 0.$$

Ebenfalls findet man hierdurch nur potentielle Kandidaten für ein (lokales) Minimum.

Hinreichende Bedingung zweiter Ordnung

Theorem (Hinreichende Bedingung 2. Ordnung)

Sei $f : \mathbb{R}^n \to \mathbb{R}$ 2mal stetig differenzierbar und $\hat{x} \in \mathbb{R}^n$ ein stationärer Punkt von f mit positiv definiter Hessematrix.

Dann ist $\hat{x}$ ein striktes lokales Minimum von f.

Im Fall $n = 1$ wird vorausgesetzt, dass

$$f''(\hat{x}) > 0.$$

Damit können wir entscheiden, ob ein Kandidat $\hat{x}$ der notwendige Bedingungen (1. und 2. Ordnung) erfüllt, in der Tat ein lokales Minimum ist.

Abstiegsrichtung

Definition (Abstiegsrichtung)

Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $x \in \mathbb{R}^n$.

$d \in \mathbb{R}^n$ heißt **Abstiegsrichtung von f in x**, falls es ein $\bar{\alpha} > 0$ gibt mit

$$f(x + \alpha d) < f(x) \quad \text{für alle } 0 < \alpha \leq \bar{\alpha}.$$

Lemma (Hilfssatz 3.1.2)

Sei $f : \mathbb{R} \to \mathbb{R}$ stetig differenzierbar in x.

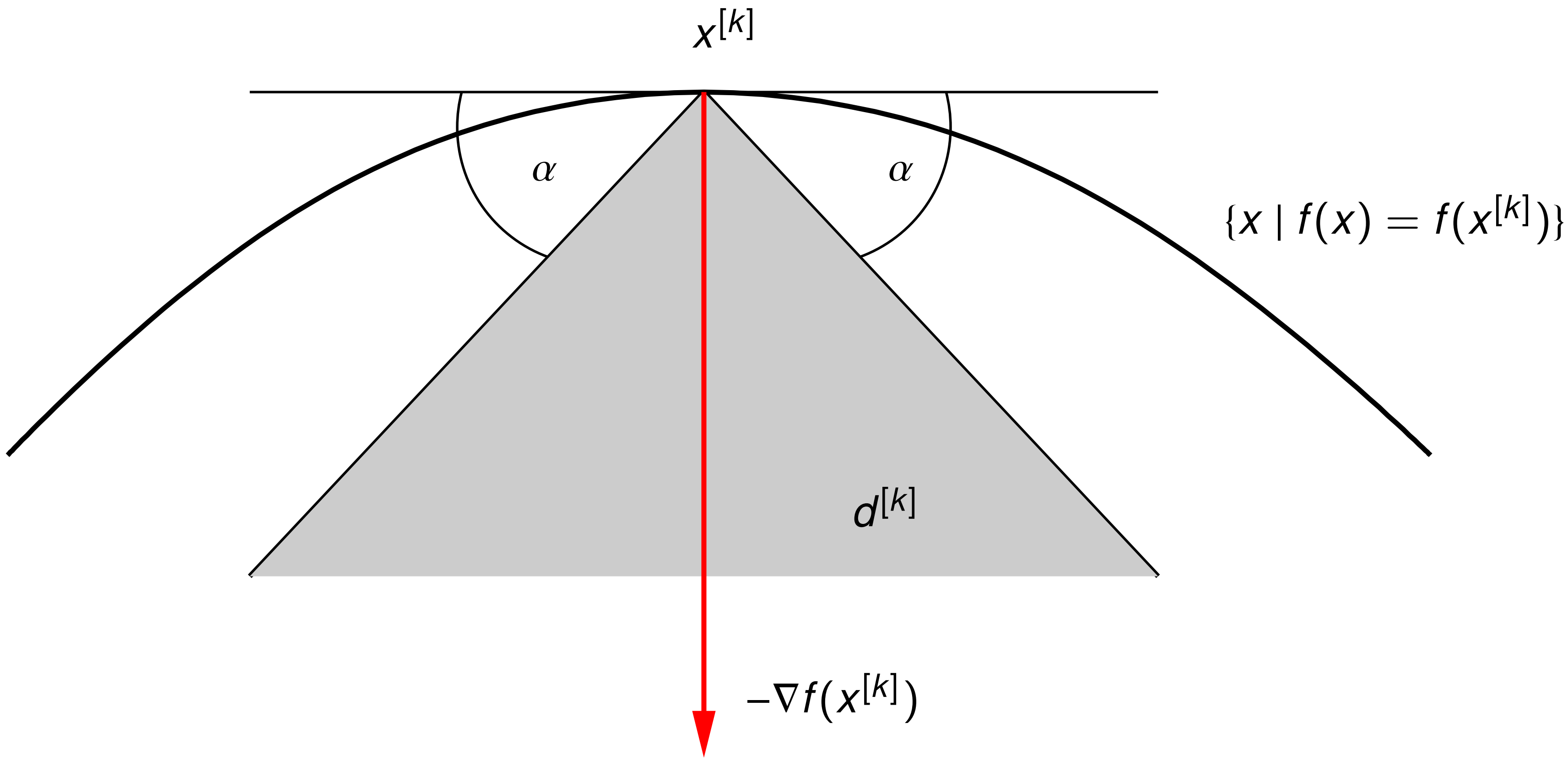
Gilt $\nabla f(x)^\top d < 0$, so ist d eine Abstiegsrichtung von f in x.

Winkelbedingung

Geometrische Interpretation des Lemma:

Winkel zwischen Gradient und Abstiegsrichtung zwischen 90° und 270°

Für effiziente Algorithmen kommt sogar nur der folgende graue Kegel für $d^{[k]}$ in Frage (Winkelbedingung):



Abstiegsverfahren

Ziel: Verfahren zur Minimierung von $f \in C^1(\mathbb{R}^n)$

Idee: Kombiniere Abstiegsrichtungen mit einer Schrittweitensteuerung:

Algorithmus (*Allgemeines Abstiegsverfahren*)

(i) *Wähle Startpunkt $x^{[0]} \in \mathbb{R}^n$.*
Setze $i = 0$.

(ii) *Falls Abruchkriterium erfüllt, STOP.*

(iii) *Berechne Abstiegsrichtung $d^{[i]}$ und*
Schrittweite $\alpha_i > 0$, so dass

$$f(x^{[i]} + \alpha_i d^{[i]}) < f(x^{[i]}).$$

Setze $x^{[i+1]} = x^{[i]} + \alpha_i d^{[i]}$.

(iv) *Setze $i := i + 1$ und*
gehe zu (ii).

Beispiele für Abstiegsverfahren

- Gradientenverfahren (Verfahren des steilsten Abstiegs):

$$d^{[i]} := -\nabla f(x^{[i]})$$

Abstiegsrichtung in nicht stationr̈en Punkten.

Naheliegend, aber nicht die effizienteste Wahl.

- Newtonverfahren:

$$d^{[i]} := -H_f(x^{[i]})^{-1}\,\nabla f(x^{[i]})$$

Falls $H_f(x^{[i]})$ positiv definit, dann $d^{[i]}$ Abstiegsrichtung in nicht stationären Punkten.

Vergleich von Abstiegsverfahren

Liniensuche zur Schrittweitenbestimmung entlang Linie

$$\varphi(\alpha) := f(x + \alpha d)$$

Beispiel (Bsp. 3.2.2)

Minimiere $f(x_1, x_2) := x_1^2 + 10x_2^2$.

Exakte Liniensuche zur Bestimmung der Schrittweite:

$$\alpha := \mathit{argmin}\{\varphi(\alpha) \,:\, \alpha > 0\} = \mathit{argmin}\{f(x + \alpha d) \,:\, \alpha > 0\}$$

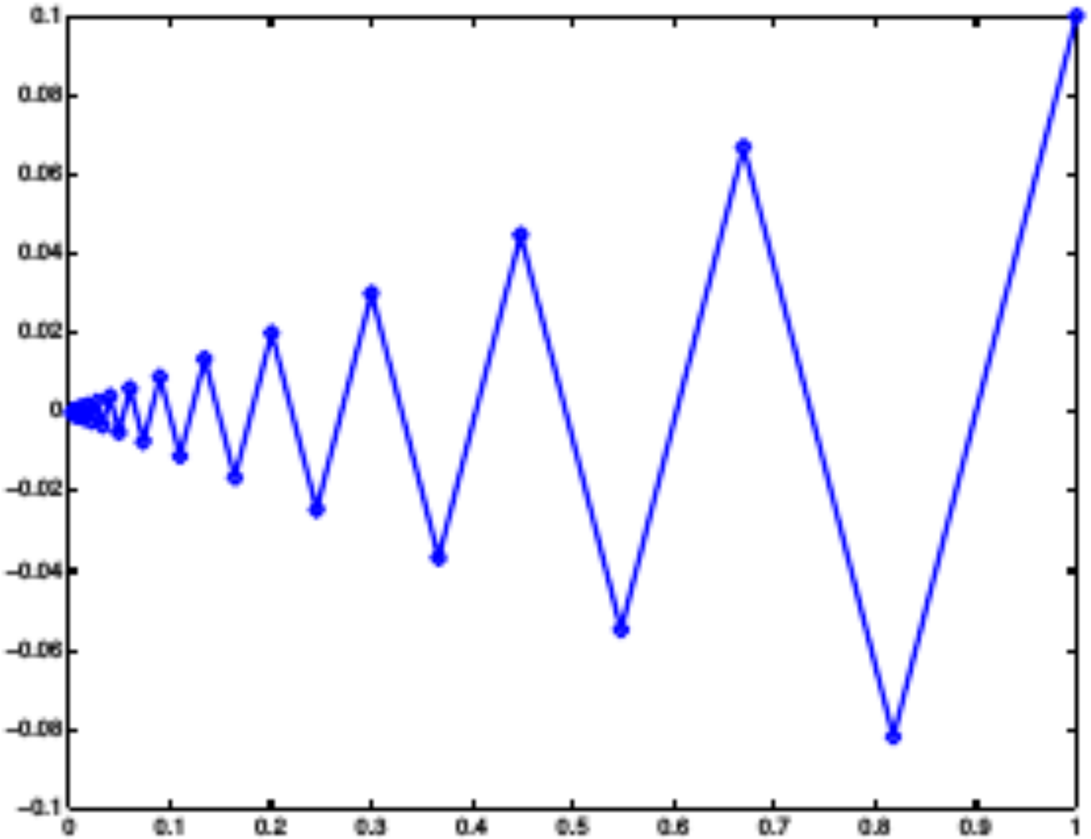
Abbruchkriterium $\|\nabla f(x)\|_2 \leq 10^{-5}$

Startvektor $x^{[0]} := (1, 0.1)^\top$

Gradientenverfahren benötigt (unskaliert) 63 Iterationen.

Newtoverfahren benötigt 1 Iteration (und löst exakt).

Konvergenz des Gradientenverfahren im Vergleich



Schrittweitenbestimmung

Algorithmus (*Armijo-Regel*)

(i) *Wähle* $\beta \in (0, 1)$, $\sigma \in (0, 1)$.
Setze $\alpha := 1$.

(ii) *Falls*

$$\varphi(\alpha) \leq \varphi(0) + \sigma\alpha\varphi'(0),$$

setze $\alpha_i := \alpha$ *und STOP.*

(iii) *Setze* $\alpha := \beta\alpha$.
Gehe zu (ii).

Die Armijo-Regel ist wohldefiniert.

In der Praxis typischerweise: $\beta = 0.9$, $\sigma = 0.01$

Intervalle, die Armijo-Bedingung erfüllen