

**MATTHIAS GERDTS**

**Einführung in die lineare und nichtlineare Optimierung**

**Universität der Bundeswehr München  
Herbsttrimester 2016**

ADDRESS OF THE AUTHOR:

Matthias Gerdts

Institut für Mathematik und Rechneranwendung

Fakultät für Luft- und Raumfahrttechnik

Universität der Bundeswehr München

Werner-Heisenberg-Weg 39

85577 Neubiberg/München

E-Mail: [matthias.gerdts@unibw.de](mailto:matthias.gerdts@unibw.de)

WWW: [www.unibw.de/lrt1/gerdts](http://www.unibw.de/lrt1/gerdts)

Version: 14. Oktober 2016

Copyright © 2016 by Matthias Gerdts

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
<b>2</b>	<b>Beispiele</b>	<b>8</b>
2.1	Elementare Funktionen und Höhenlinien . . . . .	8
2.2	Lineare und linear-quadratische Optimierungsprobleme . . . . .	12
2.3	Ausgleichs- und Parameteridentifizierungsprobleme . . . . .	16
<b>3</b>	<b>Unrestringierte Optimierung</b>	<b>20</b>
3.1	Notwendige Bedingungen . . . . .	23
3.1.1	Hinreichende Bedingungen für unrestringierte Optimierungsprobleme	29
3.2	Allgemeine Abstiegsverfahren . . . . .	31
3.3	Liniensuche, Armijo-Regel und Schrittweitenbestimmung . . . . .	33
3.4	Newton-Verfahren . . . . .	36
<b>4</b>	<b>Lineare Optimierung</b>	<b>43</b>
4.1	Ecken und zulässige Basislösungen . . . . .	48
4.2	Das primale Simplexverfahren . . . . .	56
4.2.1	Basiswechsel beim Simplexverfahren . . . . .	57
4.2.2	Der Algorithmus . . . . .	61
4.2.3	Updateformeln für das Simplextableau . . . . .	65
4.2.4	Endlichkeit des Simplexverfahrens . . . . .	68
4.3	Dualität, Sensitivität und Parametrische Optimierung . . . . .	72
4.3.1	Dualität . . . . .	72
4.3.2	Sensitivität und Schattenpreise . . . . .	76
<b>5</b>	<b>Restringierte Optimierung</b>	<b>83</b>
5.1	Notwendige Bedingungen für Standard-Optimierungsprobleme mit Gleichungsrestriktionen . . . . .	84
5.2	Notwendige Bedingungen für Standard-Optimierungsprobleme . . . . .	88
5.3	Penalty-Verfahren . . . . .	92
5.4	Multiplikator-Penalty-Verfahren . . . . .	96

<b>A Zusatzmaterial</b>	<b>99</b>
A.1 Abbruchkriterien . . . . .	99
A.2 Berechnung von Ableitungen . . . . .	100
A.3 Globalisierung des Newton-Verfahrens . . . . .	101
<b>B Ausblick</b>	<b>105</b>
<b>Literaturverzeichnis</b>	<b>107</b>

# Vorwort

Die Vorlesung vermittelt eine Einführung in die lineare und nichtlineare Optimierung und richtet sich an Studierende im Bachelorstudiengang Luft- und Raumfahrttechnik.

## Allgemeine Tipps und Hinweise:

- Mathematische Inhalte erschließen sich nicht von selbst und in den seltensten Fällen durch gelegentliches Überfliegen des Vorlesungsskriptes, sondern sie erfordern ein **aktives Mitarbeiten** und **viel Übung**.

Arbeiten Sie deshalb die Vorlesungen (und Übungen) nach!

Auch wenn die wöchentlichen Übungsaufgaben aus Kapazitätsgründen nicht eingesammelt und korrigiert werden können, so ist es zum Verständnis des Stoffes **sehr wichtig**, die Übungsaufgaben **regelmäßig** und **selbstständig** zu bearbeiten.

- Gehen Sie den Vorlesungsstoff und Übungsaufgaben mit Kommilitonen durch. Erklären und diskutieren Sie Definitionen, Sätze und Verfahren in eigenen Worten! Fragen Sie auch die Übungsleiter und den Leiter der Veranstaltung bei Unklarheiten.
- Es ist keine gute Idee, sich erst kurz vor der Prüfung mit dem Vorlesungsstoff zu beschäftigen. Dann ist es zu spät und es bleibt nichts hängen.
- Wichtige Informationen zur Vorlesung, wie z.B. Ansprechpartner, Übungsaufgaben und Zusatzmaterial, finden sich auf der WWW-Seite

<https://www.unibw.de/lrt1/gerdts/lehre/lpnlp>

Diese Seite wird laufend aktualisiert, so dass sie regelmäßig besucht werden sollte.

## Literatur:

- Marti, K. und Gröger, D. *Einführung in die lineare und nichtlineare Optimierung*. Physica-Verlag, Heidelberg, 2000.
- Geiger, C. und Kanzow, C. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer-Verlag, Berlin-Heidelberg-New York, 1999.
- Geiger, C. und Kanzow, C. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer-Verlag, Berlin-Heidelberg-New York, 2002.

- Gerdts, M. und Lempio, F. *Mathematische Optimierungsverfahren des Operations Research*. DeGruyter, Berlin, 2011.

# Kapitel 1

## Einleitung

Optimierungsaufgaben treten in den Wirtschaftswissenschaften (Operations Research), in der Technik und in den Naturwissenschaften in vielfältiger Art und Weise auf. Im Buch [NW99] von Nocedal und Wright findet man die folgende Formulierung:

**People optimize:** Airline companies schedule crews and aircraft to minimize cost. Investors seek to create portfolios that avoid risks while achieving a high rate of return. Manufacturers aim for maximizing efficiency in the design and operation of their production processes.

**Nature optimizes:** Physical systems tend to a state of minimum energy. The molecules in an isolated chemical system react with each other until the total potential energy of their electrons is minimized, Rays of light follow paths that minimize their travel time.

Die Optimierung steht in enger Beziehung zur Modellierung, d.h. Optimierungstechniken werden auf mathematische Modelle angewendet, für die dann gewisse unbekannte Modellparameter oder -funktionen so zu bestimmen sind, dass eine *Zielfunktion* unter vorgegebenen *Nebenbedingungen* minimiert (oder maximiert) wird.

Wir beschränken uns in der Klassifikation von Optimierungsproblemen ohne Beschränkung der Allgemeinheit auf Minimierungsprobleme. Ein Maximierungsproblem wird durch Multiplikation der zu maximierenden Funktion mit  $-1$  in ein äquivalentes Minimierungsproblem transformiert.

### Problem 1.0.1 (Allgemeines Optimierungsproblem (OP))

Finde  $x \in \mathbb{R}^n$ , so dass  $f(x)$  minimal wird unter der Nebenbedingung  $x \in X$ .

In Kurzform:

$$\text{Minimiere } f(x) \quad \text{u.d.N. } x \in X.$$

Darin sei  $X \subseteq \mathbb{R}^n$  eine beliebige nichtleere Menge und  $f : X \rightarrow \mathbb{R}$  eine beliebige Funktion.

Wir führen einige Bezeichnungen ein, die wir immer wieder verwenden werden:

**Definition 1.0.2**

- Die zu minimierende Funktion  $f$  heißt **Zielfunktion**.
- Ein Vektor  $x$  heißt **zulässig** für  $(OP)$ , falls  $x \in X$  gilt.  $X$  heißt **zulässige Menge** von  $(OP)$ .
- $\hat{x} \in X$  heißt **globales Minimum** von  $(OP)$ , falls

$$f(\hat{x}) \leq f(x) \quad \forall x \in X. \quad (1.1)$$

$\hat{x} \in X$  heißt **striktes globales Minimum** von  $(OP)$ , falls in (1.1) „<“ für alle  $x \in X$ ,  $x \neq \hat{x}$  gilt.

- $\hat{x} \in X$  heißt **lokales Minimum** von  $(OP)$ , falls es eine Umgebung

$$U_\varepsilon(\hat{x}) := \{x \in \mathbb{R}^n \mid \|x - \hat{x}\| < \varepsilon\}$$

gibt mit

$$f(\hat{x}) \leq f(x) \quad \forall x \in X \cap U_\varepsilon(\hat{x}). \quad (1.2)$$

$\hat{x} \in X$  heißt **striktes lokales Minimum** von  $(OP)$ , falls in (1.2) „<“ für alle  $x \in X \cap U_\varepsilon(\hat{x})$ ,  $x \neq \hat{x}$  gilt.

Globale bzw. lokale Maxima werden analog definiert. Die Begriffe werden in Abbildung 1.1 erläutert.

Wir betrachten **Spezialfälle**:

Ein unrestringiertes Problem liegt vor, falls  $X = \mathbb{R}^n$  gilt:

**Problem 1.0.3 (Unrestringiertes Optimierungsproblem (UOP))**

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

Darin sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  eine beliebige Funktion.

Häufig wird die Menge  $X$  in  $(OP)$  durch endlich viele Gleichungen und Ungleichungen beschrieben:



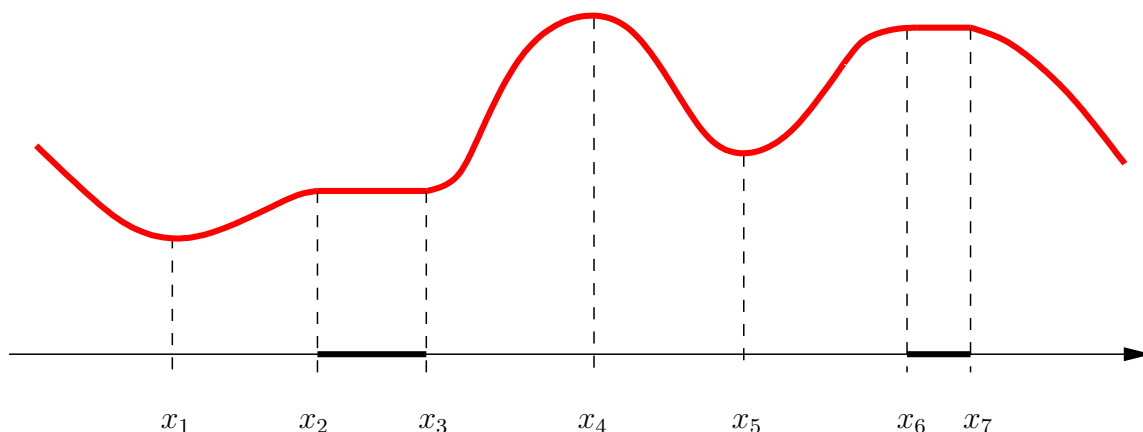


Abbildung 1.1: Lokale und globale Minima und Maxima einer Funktion:  $x_1$ : striktes globales Minimum,  $x_2$ : lokales Maximum,  $x_3$ : lokales Minimum;  $(x_2, x_3)$ : gleichzeitig lokales Minimum und Maximum,  $x_4$ : striktes globales Maximum,  $x_5$ : striktes lokales Minimum,  $x_6, x_7$ : lokale Maxima,  $(x_6, x_7)$ : gleichzeitig lokales Minimum und Maximum.

#### Problem 1.0.4 (Standard Optimierungsproblem (SOP))

Finde  $x \in \mathbb{R}^n$ , so dass  $f(x)$  minimal wird unter den Nebenbedingungen

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

In Kurzform schreiben wir auch:

$$\text{Minimiere } f(x) \quad \text{u.d.N.} \quad g(x) \leq 0, \quad h(x) = 0.$$

Darin seien  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g = (g_1, \dots, g_m)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , und  $h = (h_1, \dots, h_p)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^p$  beliebige Funktionen.

Lineare Optimierungsprobleme stellen einen wichtigen Spezialfall des Standard Optimierungsproblems mit  $f(x) = c^\top x$ ,  $g(x) = -x$ ,  $h(x) = Ax - b$  (oder  $h(x) = b - Ax$ ) dar, wobei  $c \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$  Vektoren und  $A \in \mathbb{R}^{m \times n}$  eine Matrix sind.

#### Problem 1.0.5 (Lineares Optimierungsproblem (LOP) (in sogenannter primaler Normalform))

$$\text{Minimiere } c^\top x \quad \text{u.d.N.} \quad Ax = b, \quad x \geq 0.$$

Darin seien  $x, c \in \mathbb{R}^n$  und  $b \in \mathbb{R}^m$  Vektoren und  $A \in \mathbb{R}^{m \times n}$  eine Matrix.

### Bemerkung 1.0.6

Die Darstellung der verschiedenen Optimierungsprobleme ist nicht vollständig. So gibt es beispielsweise auch unendlichdimensionale Optimierungsprobleme, Vektoroptimierungsprobleme, semi-infinite Optimierungsprobleme, ganzzahlige Optimierungsprobleme, nicht-differenzierbare Optimierungsprobleme sowie Mischformen der oben dargestellten Problemformen. ■

In dieser Vorlesung beschäftigen wir uns mit **unrestringierten Optimierungsproblemen (UOP)**, **linearen Optimierungsproblemen (LP)** und **Optimierungsproblemen in Standardform (SOP)**, wobei die auftretenden Funktionen in der Regel mindestens einmal stetig differenzierbar sind.

#### Typische Fragestellungen:

- i) **Existieren** überhaupt **zulässige Lösungen**?
- ii) **Existieren Optimallösungen**?
- iii) Ist die Optimallösung **eindeutig** bestimmt?
- iv) **Wie hängen die Optimallösungen von den Problemdaten ab**?
- v) Welche Eigenschaften besitzen Optimallösungen, mit anderen Worten welche Bedingungen werden von einer Optimallösung **notwendig** erfüllt?
- vi) Welche Bedingungen sind **hinreichend** dafür, dass eine zulässige Lösung optimal ist?
- vii) Welche Bedingungen sind gleichzeitig notwendig und hinreichend für Optimalität, **charakterisieren** also die Optimallösungen?
- viii) Wie gewinnt man aus zulässigen Lösungen Informationen über die Optimallösungen, insbesondere **Einschließungen** für den Optimalwert und **Fehlerabschätzungen** für die Optimallösung?
- ix) Welche **konzeptionellen Algorithmen** zur Berechnung einer Optimallösung stehen zur Verfügung?

x) Welche **numerischen Eigenschaften** besitzen diese Algorithmen (Konvergenz, Konvergenzgeschwindigkeit, Stabilität)?

Die Fragestellungen i) – vii) sind überwiegend theoretischer Natur. Aber ohne ihre Beantwortung ist die Behandlung der numerischen Fragestellungen viii) – x) nicht möglich.

In dieser Vorlesung widmen wir uns schwerpunktmäßig den praktischen Fragestellungen nach der Konstruktion von Verfahren und der Charakterisierung von Optimallösungen. Dabei setzen wir die Existenz von Optimallösungen stets voraus. Die Existenz einer Optimallösung ist dabei zum Beispiel durch den folgenden aus der Analysis bekannten Satz gesichert.

**Satz 1.0.7 (Weierstrass)**

*Sei  $X \subseteq D \subseteq \mathbb{R}^n$  kompakt (also beschränkt und abgeschlossen) und  $f : D \rightarrow \mathbb{R}$  stetig. Dann nimmt  $f$  ihr Minimum (und Maximum) auf  $X$  an.*

# Kapitel 2

## Beispiele

Es werden exemplarisch einige typische Optimierungsaufgaben vorgestellt und diskutiert. Weitere interessante Anwendungen sind in Spellucci [Spe93], Bazarraa [BSS93] und Alt [Alt02] zu finden. Darüber hinaus gibt es unzählige Optimierungsprobleme in Industrie, Wirtschaft und Wissenschaft.

### 2.1 Elementare Funktionen und Höhenlinien

#### Beispiel 2.1.1 (Funktion von Himmelblau)

*Wir wollen die Funktion von Himmelblau*

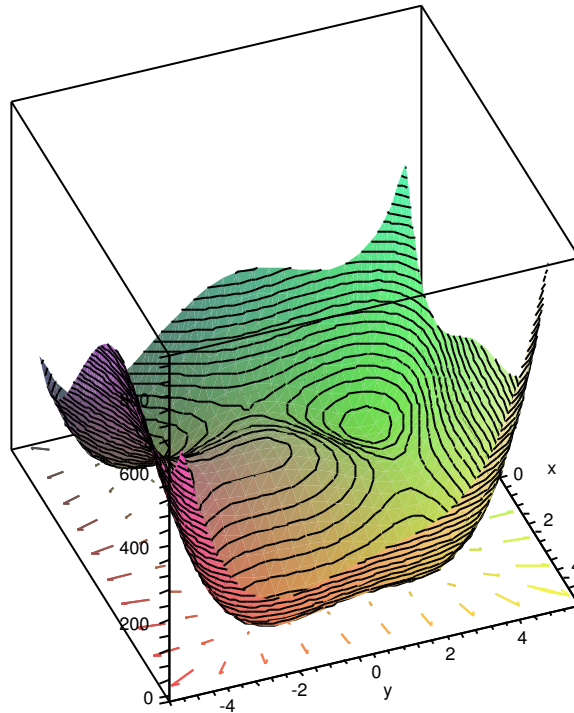
$$f_H(x, y) := (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

*über alle  $(x, y)^T \in \mathbb{R}^2$  minimieren.*

*Um einen besseren Eindruck von der Funktion zu bekommen, stellen wir die Funktion grafisch dar. Dies kann z.B. mit dem Programm MAPLE und dem Befehl*

```
plot3d((x^2+y-11)^2+(x+y^2-7)^2,x=-5..5,y=-5..5,axes=boxed,  
style=PATCHCONTOUR,contours = 40,shading=XYZ);
```

*geschehen:*



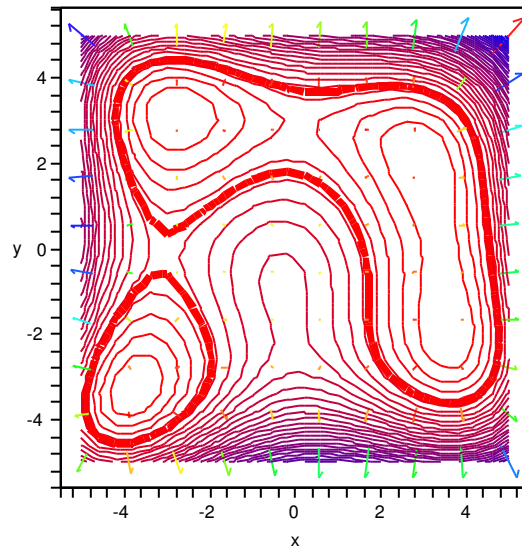
Einen noch besseren Eindruck von den Funktionswerten der Funktion erhalten wir mit dem Befehl

```
contourplot((x^2+y-11)^2+(x+y^2-7)^2,x=-5..5,y=-5..5,contours=80,
  coloring=[red,blue],scaling=constrained,axes=boxed);
```

Dieser Befehl zeichnet die sogenannten **Höhenlinien** oder **Niveaulinien** einer Funktion. Eine **Höhenlinie zum Niveau**  $c \in \mathbb{R}$  für eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  ist formal definiert als die Menge aller Punkte  $x = (x_1, \dots, x_n)^\top$ , die  $f(x) = c$  erfüllen und wird mit  $N_f(c)$  bezeichnet, also

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

Somit besitzt die Funktion  $f$  entlang einer Höhenlinie immer denselben Funktionswert, ist also entlang einer Höhenlinie konstant. Die Abbildung zeigt die Höhenlinien der Funktion von Himmelblau. Die fett gezeichnete Höhenlinie entspricht dem Niveau  $c = 100$ .



Die Pfeile in den beiden Grafiken stellen die **Gradienten von  $f$**  in den jeweiligen Punkten dar. Der Gradient einer Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  an der Stelle  $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$  ist formal definiert als der Spaltenvektor

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

Bekanntlich zeigt der Gradient einer Funktion  $f$  in die **Richtung des steilsten Anstiegs** von  $f$ . Außerdem steht der Gradient **senkrecht** auf den Höhenlinien von  $f$ .

Für die Funktion von Himmelblau ergibt sich speziell

$$\nabla f_H(x, y) = \begin{pmatrix} \frac{\partial f_H}{\partial x}(x, y) \\ \frac{\partial f_H}{\partial y}(x, y) \end{pmatrix} = \begin{pmatrix} 4x(x^2 + y - 11) + 2(x + y^2 - 7) \\ 2(x^2 + y - 11) + 4y(x + y^2 - 7) \end{pmatrix}.$$

Die Gradienten von  $f_H$  können mit dem folgenden Befehl dargestellt werden:

```
gradplot( (x^2+y-11)^2+(x+y^2-7)^2, x=-5..5, y=-5..5, grid=[10,10],
color=(x^2+y-11)^2+(x+y^2-7)^2);
```

Anhand der grafischen Darstellungen läßt sich folgendes ablesen: Die Funktion von Himmelblau besitzt

- 4 lokale Minimalstellen (zugleich global) mit Funktionswert 0
- 4 Sattelpunkte
- ein lokales Maximum in  $(-0.270845, -0.923039)^\top$

Wir werden später sehen, dass der Gradient von  $f$  in jedem dieser Punkte gleich dem Nullvektor ist. ■

### Beispiel 2.1.2 (Optimierungsproblem mit Nebenbedingungen)

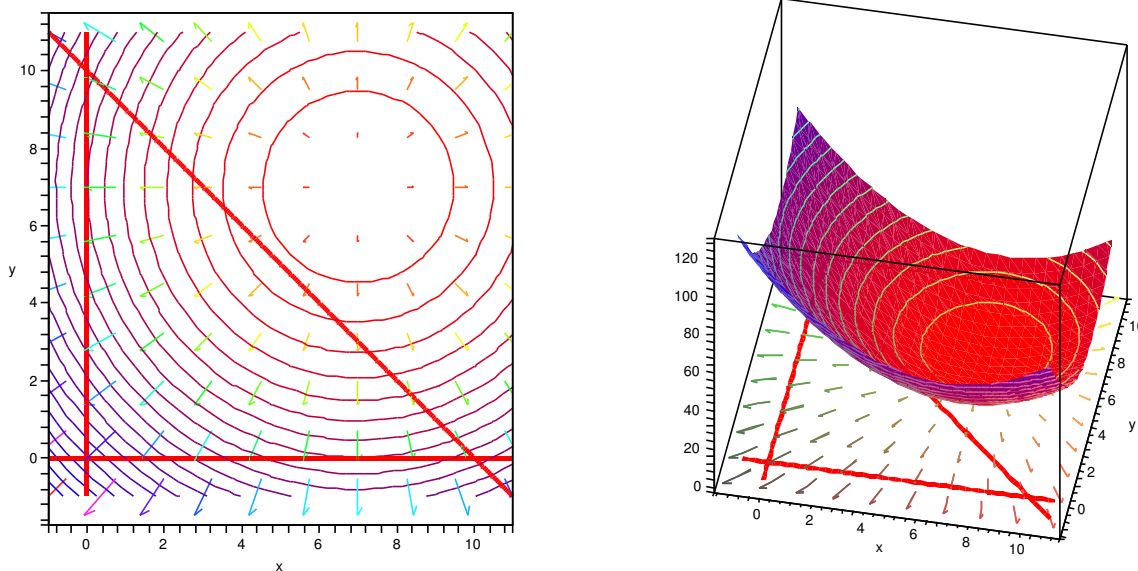
Betrachte das folgende Standard-Optimierungsproblem:

$$\begin{aligned} \text{Minimiere} \quad & f(x, y) = (x - 7)^2 + (y - 7)^2 \\ \text{unter} \quad & x, y \in \mathbb{R}, \\ & x, y \geq 0, \\ & x + y \leq 10. \end{aligned}$$

Zunächst betrachten wir den **zulässigen Bereich**  $X$  des Optimierungsproblems, also alle Punkte  $(x, y)^\top \in \mathbb{R}^2$ , die die Nebenbedingungen

$$x \geq 0, \quad y \geq 0, \quad x + y \leq 10$$

erfüllen. Der Bereich innerhalb des Dreiecks in der nachfolgenden Abbildung ist der zulässige Bereich (zusätzlich sind die Höhenlinien und die Gradienten der Zielfunktion  $f$  eingezeichnet):



Es ist leicht zu sehen, dass die Zielfunktion  $f$  im Punkt  $(7, 7)^\top$  ein globales Minimum mit Funktionswert 0 besitzt. Wegen  $7+7 > 10$  ist dieser Punkt jedoch nicht zulässig (die dritte Nebenbedingung  $x+y \leq 10$  ist verletzt) und somit keine Lösung des Optimierungsproblems! Die tatsächliche Lösung des Problems liegt im Punkt  $(x, y)^\top = (5, 5)^\top$  auf dem Rand des zulässigen Bereiches. ■

## 2.2 Lineare und linear-quadratische Optimierungsprobleme

### Beispiel 2.2.1 (Lineares Optimierungsproblem)

Ein Landwirt bewirtschaftet ein Grundstück von 40 Hektar Größe mit Zuckerrüben und Weizen. Er kann hierzu 2400 Euro und 312 Arbeitstage einsetzen. Pro Hektar betragen seine Anbaukosten bei Rüben 40 Euro und bei Weizen 120 Euro. Für Rüben benötigt er 6 Arbeitstage, für Weizen 12 Arbeitstage pro Hektar. Der Reingewinn bei Rüben sei 100 Euro pro Hektar, bei Weizen sei er 250 Euro pro Hektar.

*Mathematische Formulierung:* Wir bezeichnen mit  $x_1$  die Fläche, die mit Rüben bepflanzt wird und mit  $x_2$  die Fläche, die mit Weizen bepflanzt wird. Der zu maximierende Gewinn lautet

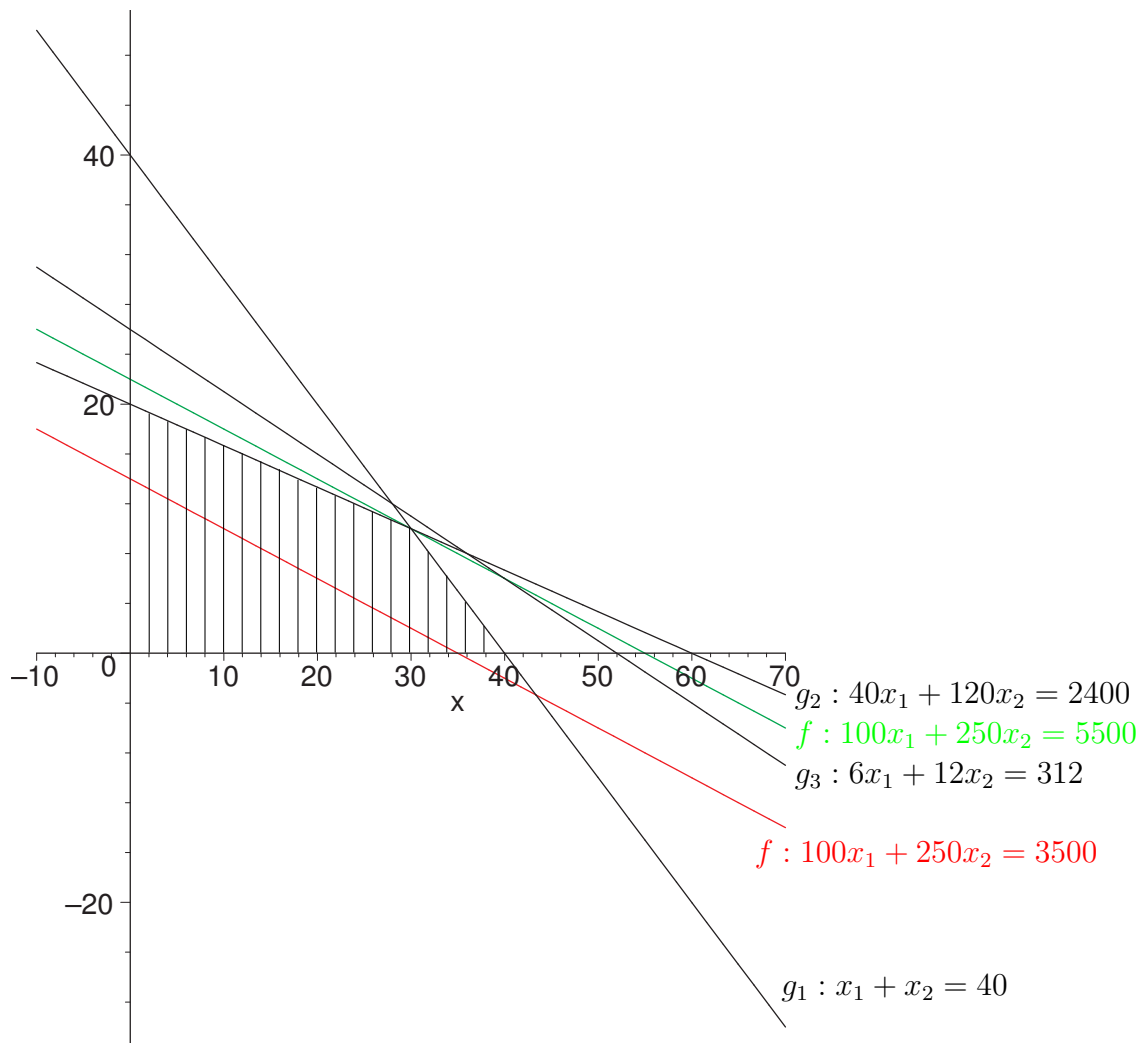
$$f(x_1, x_2) = 100x_1 + 250x_2.$$

Aus der Aufgabenstellung lassen sich folgende Beschränkungen ableiten:

Grundstücksgröße:	$g_1(x_1, x_2) := x_1 + x_2 \leq 40$
Geld:	$g_2(x_1, x_2) := 40x_1 + 120x_2 \leq 2400$
Arbeitstage:	$g_3(x_1, x_2) := 6x_1 + 12x_2 \leq 312$
keine negativen Flächen:	$x_1, x_2 \geq 0$

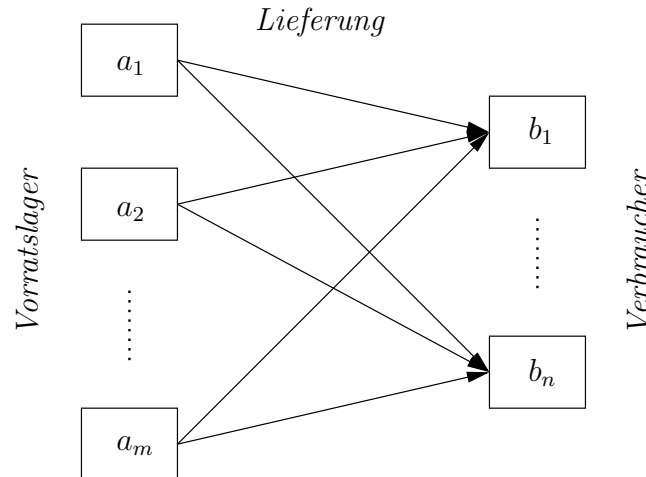
Der zulässige Bereich des Optimierungsproblems ist durch den schraffierten Bereich in der folgenden Abbildung gegeben. Die rote Gerade stellt die Höhenlinie der Zielfunktion zum Niveau 3500, die grüne diejenige zum Niveau 5500 dar.





### Beispiel 2.2.2 (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.



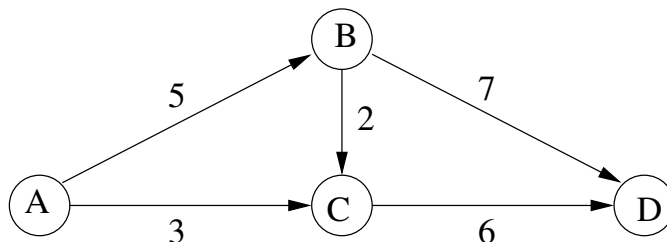
Bezeichnet  $x_{ij}$  die Liefermenge von Lager  $i$  zu Verbraucher  $j$ , so führt das Problem auf das folgende Transportproblem, welches ein spezielles lineares Optimierungsproblem ist:

$$\begin{aligned} &\text{Minimiere} && \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \\ &\text{unter} && \sum_{j=1}^n x_{ij} \leq a_i, \quad i = 1, \dots, m, \\ &&& \sum_{i=1}^m x_{ij} \geq b_j, \quad j = 1, \dots, n, \\ &&& x_{ij} \geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \end{aligned}$$

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. ■

### Beispiel 2.2.3 (Netzwerkproblem)

Eine Firma möchte so viele Waren wie möglich von Stadt A zu Stadt D über das abgebildete Straßennetzwerk transportieren, wobei die Zahlen neben den Kanten des Netzwerks die maximale Kapazität der jeweiligen Kante angeben.



Wie kann dieses Problem mathematisch modelliert werden?

Es bezeichne

$$V = \{A, B, C, D\}$$

die Menge der Knoten des Netzwerks, welche den Städten im Netzwerk entsprechen. Weiter bezeichne

$$E = \{(A, B), (A, C), (B, C), (B, D), (C, D)\}$$

die Menge der Kanten im Netzwerk, welche den Verbindungsstraßen zwischen jeweils zwei Städten entsprechen. Für jede Kante  $(i, j) \in E$  bezeichne  $x_{ij}$  die tatsächlich transportierte Menge entlang der Kante  $(i, j)$  und  $u_{ij}$  die maximale Kapazität der Kante. Dann muss die Kapazitätsbeschränkung

$$0 \leq x_{ij} \leq u_{ij} \quad ((i, j) \in E)$$

gelten. Desweiteren ist es sinnvoll anzunehmen, dass in den Städten  $B$  und  $C$  keine Waren produziert werden und auch keine Waren abhanden kommen, so dass die Erhaltungsgleichungen „Abfluss - Zufluss = 0“ in  $B$  und  $C$  gelten müssen:

$$x_{BD} + x_{BC} - x_{AB} = 0,$$

$$x_{CD} - x_{AC} - x_{BC} = 0.$$

Da auf Grund der Erhaltungsgleichungen unterwegs keine Waren verschwinden können und keine zusätzlichen Waren generiert werden, besteht die Aufgabe nun darin, die den Startknoten verlassende Warenmenge zu maximieren (dies ist dieselbe Warenmenge, die den Knoten  $D$  erreicht):

$$\begin{array}{ll} \text{Maximiere} & x_{AB} + x_{AC} \\ \text{u.d.N.} & x_{BD} + x_{BC} - x_{AB} = 0, \\ & x_{CD} - x_{AC} - x_{BC} = 0, \\ & 0 \leq x_{ij} \leq u_{ij} \quad ((i, j) \in E). \end{array}$$

■

#### Beispiel 2.2.4 (Portfoliooptimierung)

Wir betrachten das Beispiel einer Portfoliooptimierungsaufgabe nach Markowitz. Gegeben seien  $j = 1, \dots, n$  mögliche Anlagen (z.B. Aktien, Fonds, Optionen, Wertpapiere). Jede Anlage wirft im nächsten Zeitintervall einen Gewinn (oder Verlust)  $R_j$  ab. Leider ist  $R_j$  in der Regel nicht bekannt, sondern zufällig verteilt. Um einerseits den Gewinn zu maximieren und andererseits das Risiko eines Verlusts zu minimieren, wird die Anlagesumme zu Anteilen  $x_j$  auf die Anlagen  $j = 1, \dots, n$  verteilt und die anteiligen Anlagen werden in einem Portfolio zusammengefaßt.

Die Aufgabe eines Portfoliomanagers besteht in der optimalen Zusammensetzung eines solchen Portfolios, d.h. die Anteile  $x_j$  der jeweiligen Anlagen, die in das Portfolio übernommen werden sollen, müssen in einem gewissen Sinne optimal bestimmt werden. Ein

mögliches Ziel ist es, den erwarteten Gewinn

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

zu maximieren ( $E$  bezeichnet den Erwartungswert). Jedoch ist ein hoher Gewinn in der Regel nur mit riskanten Anlagen möglich, so dass auch das Risiko eines Verlusts steigt. Als Maß für das Risiko kann die Varianz des Gewinns dienen:

$$\text{Var}(R) = E(R - E(R))^2 = E\left(\sum_{j=1}^n x_j (R_j - E(R_j))\right)^2$$

Ein Kompromiss zwischen hohem Gewinn und geringem Risiko kann durch Lösen des folgenden Optimierungsproblems erreicht werden:

$$\begin{aligned} \min \quad & -\sum_{j=1}^n x_j E(R_j) + \alpha E\left(\sum_{j=1}^n x_j (R_j - E(R_j))\right)^2 \\ \text{unter} \quad & \sum_{j=1}^n x_j = 1, \quad x_j \geq 0, \quad j = 1, \dots, n. \end{aligned}$$

Hierin bezeichnet  $\alpha > 0$  einen Gewichtungparameter, mit dem die Risikobereitschaft gesteuert werden kann. Mit  $\alpha = 0$  wird der Varianzterm in der Zielfunktion eliminiert, so dass nur noch der Gewinn maximiert wird. Dies entspricht einer hohen Risikobereitschaft. Mit wachsendem  $\alpha$  wird der Varianzterm stärker gewichtet und die Risikobereitschaft sinkt. ■

### 2.3 Ausgleichs- und Parameteridentifizierungsprobleme

Ein Experiment liefert die Messpunkte  $(t_i, y_i)$ ,  $i = 1, \dots, m$ . Der dem Experiment zu Grunde liegende Vorgang werde durch die Funktion  $f(t, p)$  modelliert, die einen funktionalen Zusammenhang zwischen den Messstellen  $t_i$  und den Messwerten  $y_i$  herstellt. Allerdings hängt die Funktion auch noch vom unbekanntem Parameter  $p \in \mathbb{R}^{n_p}$  ab. In der Praxis sind die Messwerte verrauscht bzw. fehlerbehaftet, so dass es in der Regel keinen Parameter  $p$  gibt, der die Messpunkte exakt reproduziert. Daher wird versucht, die Messpunkte so gut wie möglich zu approximieren, indem

$$\frac{1}{2} \sum_{i=1}^m (y_i - f(t_i, p))^2$$

bezüglich  $p$  minimiert wird. Häufig sind zusätzlich noch Nebenbedingungen an den Parameter  $p$  gegeben.

Das resultierende Optimierungsproblem ist ein spezielles Least-Squares Problem. Ein allgemeines Least-Squares Problem lautet

**Least-Squares-Problem:**

Finde  $x \in \mathbb{R}^n$ , so dass

$$\frac{1}{2} \|\Phi(x)\|_2^2 = \frac{1}{2} \sum_{i=1}^q \Phi_i(x)^2$$

minimal wird unter den Nebenbedingungen

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

**Beispiel 2.3.1**

*Die Messpunkte*

$t_i$	$y_i$
-0.7882416043	0.396878358
-0.6056336413	0.418410056
-0.3976460600	0.627676951
-0.2144255029	0.821174784
-0.0107919623	0.962155739
0.1997798535	1.303597193
0.3741472164	1.362401309
0.5955672872	1.470902326
0.7899671852	1.528415842
0.9997213026	1.510113124

sollen durch die Funktion

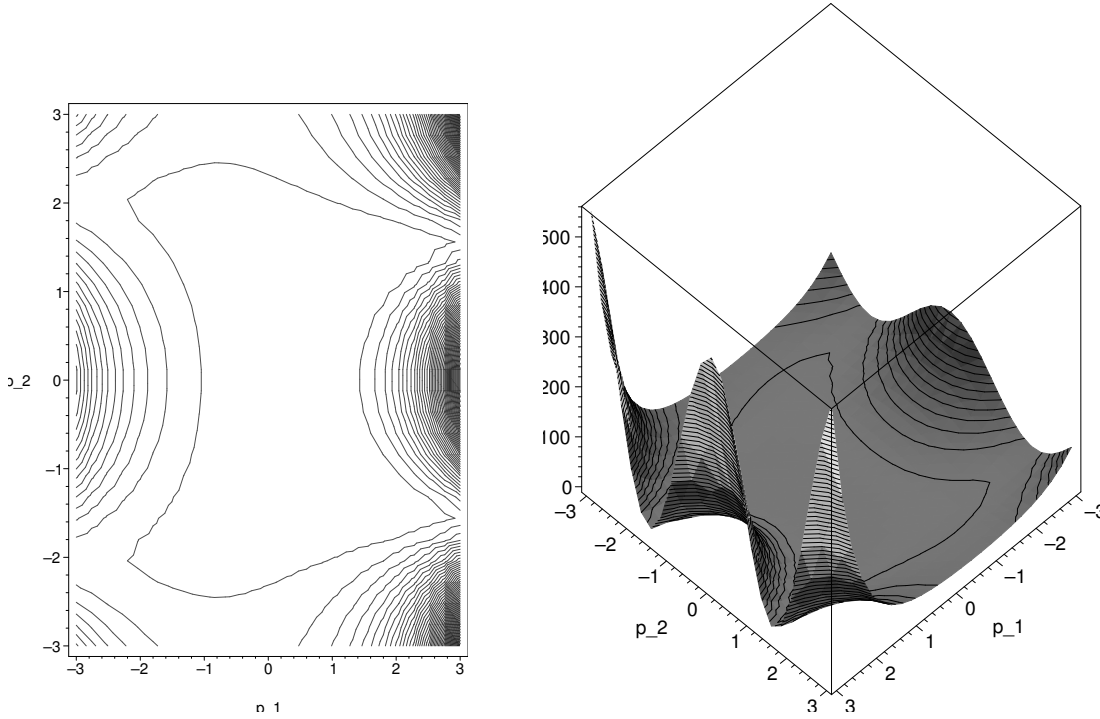
$$f(t, p_1, p_2) = \exp(p_1 t) \cos(p_2 t)$$

wiedergegeben werden.

Die Abbildungen zeigen die zu minimierende Funktion

$$g(p_1, p_2) = \frac{1}{2} \sum_{i=1}^{10} (y_i - f(t_i, p_1, p_2))^2$$

und deren Höhenlinien.



Die Fehlerfunktion zeigt, dass es Bereiche mit sehr steilen Flanken und sehr flache Täler gibt.

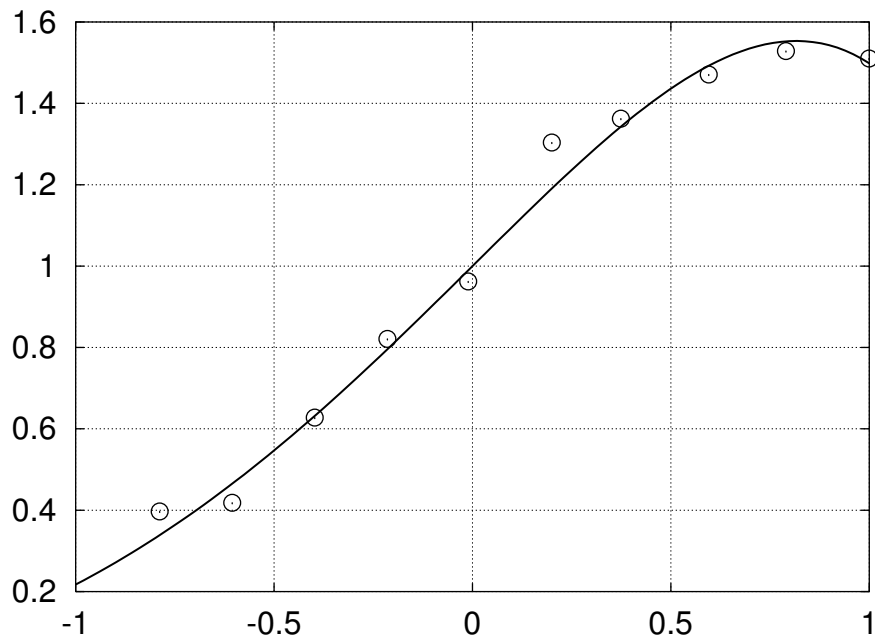
Anwendung eines SQP-Verfahrens liefert das folgende Resultat.

```
----- SQP VERSION 1.1 (C) Matthias Gerdts, University of Bayreuth, 2004 -----
NUMBER OF VARIABLES      :      2
NUMBER OF CONSTRAINTS    :      0
METHOD                   : SEQUENTIAL QUADRATIC PROGRAMMING (SQP)
MERIT FUNCTION           : AUGMENTED LAGRANGIAN
MULTIPLIER UPDATE RULE   : SCHITTOKOWSKI OR
OPTIMALITY TOLERANCE     : 0.149E-07
FEASIBILITY TOLERANCE    : 0.100E-11
LINE SEARCH PARAMETER    : SIGMA= 0.100E+00 BETA= 0.900E+00
MAXIMUM NUMBER OF ITERATIONS : 10000
INFINITY                 : 0.100E+21
ROUND OFF TOLERANCE      : 0.300E-12
REAL WORK SPACE PROVIDED : 5338   NEEDED : 50
INTEGER WORK SPACE PROVIDED : 82   NEEDED : 6
-----
```

ITER	QPIT	ALPHA	OBJ	NB	KKT	PEN	IDI	DELTA	RDELTA	F/G
0	0	0.0000E+00	0.6954883660193656E+01	0.0000E+00	0.4961E+01	0.0000E+00	0.0000E+00	0.0000E+00	0.1000E+01	1/ 0
1	1	0.4305E+00	0.4967522120517841E+01	0.0000E+00	0.6195E+01	0.0000E+00	0.5608E+01	0.0000E+00	0.1000E+01	10/ 0
2	1	0.1000E+01	0.3151863029012962E+00	0.0000E+00	0.2952E+01	0.0000E+00	0.2958E+01	0.0000E+00	0.1000E+01	11/ 0
3	1	0.2059E+00	0.1828168391624115E+00	0.0000E+00	0.1755E+01	0.0000E+00	0.1609E+01	0.0000E+00	0.1000E+01	27/ 0
4	1	0.1000E+01	0.1275778706056140E-01	0.0000E+00	0.1589E+00	0.0000E+00	0.2781E+00	0.0000E+00	0.1000E+01	28/ 0
5	1	0.1000E+01	0.1158079716928466E-01	0.0000E+00	0.7956E-01	0.0000E+00	0.8876E-02	0.0000E+00	0.1000E+01	29/ 0
6	1	0.1000E+01	0.1096782064438834E-01	0.0000E+00	0.1468E-01	0.0000E+00	0.1593E-01	0.0000E+00	0.1000E+01	30/ 0
7	1	0.1000E+01	0.1074391932113511E-01	0.0000E+00	0.1579E-01	0.0000E+00	0.1675E-01	0.0000E+00	0.1000E+01	31/ 0
8	1	0.1000E+01	0.1067396922629297E-01	0.0000E+00	0.3616E-02	0.0000E+00	0.1358E-01	0.0000E+00	0.1000E+01	32/ 0
9	1	0.1000E+01	0.1067272182372846E-01	0.0000E+00	0.7891E-03	0.0000E+00	0.9467E-03	0.0000E+00	0.1000E+01	33/ 0
10	1	0.1000E+01	0.1067267346498150E-01	0.0000E+00	0.5059E-04	0.0000E+00	0.1286E-03	0.0000E+00	0.1000E+01	34/ 0
11	1	0.1000E+01	0.1067267301930280E-01	0.0000E+00	0.1533E-05	0.0000E+00	0.3190E-04	0.0000E+00	0.1000E+01	35/ 0
12	1	0.1000E+01	0.1067267301842276E-01	0.0000E+00	0.7265E-07	0.0000E+00	0.1620E-05	0.0000E+00	0.1000E+01	36/ 0
13	1	0.1000E+01	0.1067267301842218E-01	0.0000E+00	0.9460E-09	0.0000E+00	0.2440E-07	0.0000E+00	0.1000E+01	37/ 0

```
KKT CONDITIONS SATISFIED (IER= 0)!
SOLUTION:
OBJ = 0.1067267301842218E-01
KKT = 0.9460120590228484E-09
CDN = 0.0000000000000000E+00
X =
0.9656009650544685E+00
-0.9636591123058328E+00
```

Schließlich ergibt sich die Funktion  $f$  aus den identifizierten Parametern:



■

# Kapitel 3

## Unrestringierte Optimierung

Gegenstand dieses Kapitels sind die theoretische Untersuchung und die Entwicklung numerischer Verfahren für das

### Problem 3.0.2 (Unrestringiertes Optimierungsproblem (UOP))

$$\text{Minimiere } f(x) \text{ unter } x \in \mathbb{R}^n.$$

Darin sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  mindestens einmal stetig differenzierbar.

Wir werden **notwendige Bedingungen** und **hinreichende Bedingungen** für ein lokales Minimum entwickeln.

- **Notwendige Bedingungen** sind Bedingungen, die ein lokales Minimum  $\hat{x}$  zwangsläufig erfüllt. Bei der Herleitung wird daher stets vorausgesetzt, dass bereits ein lokales Minimum  $\hat{x}$  bekannt ist. Allgemein werden wir sehen, dass die Bedingung  $\nabla f(\hat{x}) = 0$  für unrestringierte Optimierungsprobleme eine notwendige Bedingung darstellt. Sie ist aber nicht hinreichend.
- Im Gegensatz zu den notwendigen Bedingungen ist bei **hinreichenden Bedingungen** a priori nicht bekannt, ob es sich bei einem Kandidaten  $\hat{x}$  um ein lokales Minimum handelt oder nicht. Hinreichende Bedingungen sind Bedingungen, mit denen entschieden werden kann, ob es sich bei  $\hat{x}$  um ein lokales Minimum handelt oder nicht. Später wird gezeigt, dass die Bedingung „ $H_f(\hat{x})$  positiv definit“, wobei  $H_f(\hat{x})$  die Hessematrix von  $f$  in  $\hat{x}$  bezeichnet, in der unrestringierten Optimierung eine hinreichende Bedingung für ein Minimum ist. Sie ist aber nicht notwendig.
- Ideal sind Bedingungen, die sowohl notwendig als auch hinreichend sind. Derartige Bedingungen sind aber nur für spezielle Optimierungsprobleme, etwa konvexe Probleme, bekannt.

Anschließend werden numerische Verfahren diskutiert und analysiert. Die meisten der vorzustellenden Verfahren versuchen, die notwendigen Bedingungen zu erfüllen.



Zur Erinnerung:

- Der **Gradient** einer Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  an der Stelle  $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$  ist formal definiert als der 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 Spezialfall  $n = 1$  ist der Gradient gerade die erste Ableitung von  $f$ . Bekanntlich zeigt der Gradient einer Funktion  $f$  in die **Richtung des steilsten Anstiegs** von  $f$ . Außerdem steht der Gradient **senkrecht** auf den Höhenlinien von  $f$ .

- Die **Hessematrix** von  $f$  an der Stelle  $x$  ist gegeben durch

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

Häufig schreibt man auch  $H_f(x)$  anstatt  $H_f(x)$ .

Im Spezialfall  $n = 1$  ist die Hessematrix gerade die zweite Ableitung von  $f$ . Die Hessematrix beschreibt anschaulich die lokale Krümmung einer Funktion.

Die folgenden Betrachtungen basieren auf der lokalen Approximierbarkeit von  $f$  in  $\hat{x}$ . Nach dem **Satz von Taylor (Taylorentwicklung)** gilt:

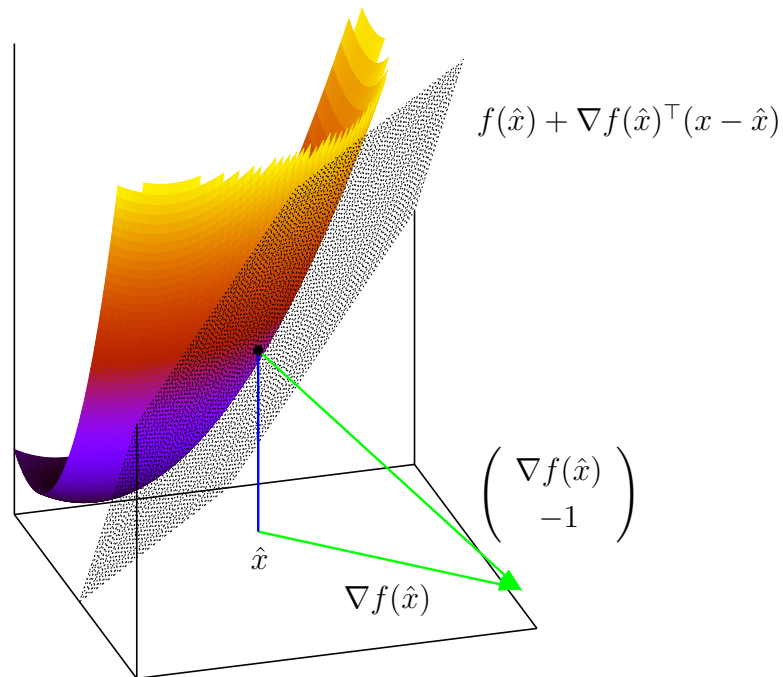
- Ist  $f$  stetig differenzierbar, so gilt

$$f(x) = f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}) + o(\|x - \hat{x}\|) \quad \text{mit} \quad \lim_{x \rightarrow \hat{x}} \frac{o(\|x - \hat{x}\|)}{\|x - \hat{x}\|} = 0.$$

Somit kann  $f$  in einer Umgebung von  $\hat{x}$  approximiert werden durch eine affin lineare Funktion:

$$f(x) \approx f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}),$$

siehe Skizze.



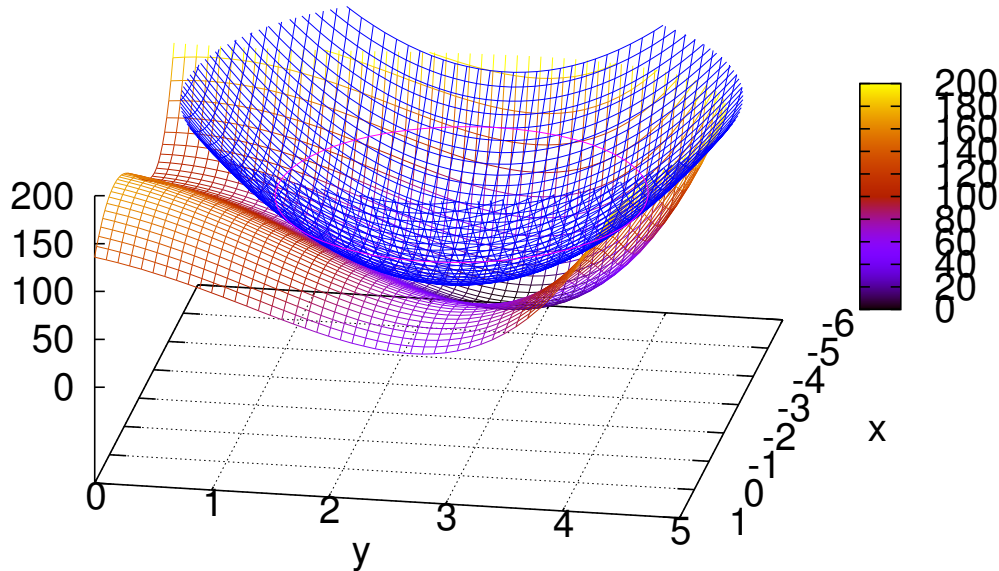
- Ist  $f$  zweimal stetig differenzierbar, so gilt

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

mit  $\lim_{x \rightarrow \hat{x}} \frac{o(\|x - \hat{x}\|^2)}{\|x - \hat{x}\|^2} = 0$ . Somit kann  $f$  in einer Umgebung von  $\hat{x}$  approximiert werden durch eine quadratische Funktion:

$$f(x) \approx f(\hat{x}) + \nabla f(\hat{x})^\top(x - \hat{x}) + \frac{1}{2}(x - \hat{x})^\top H_f(\hat{x})(x - \hat{x}). \quad (3.1)$$

## Himmelblau function and quadratic approximation at (-2,2)



### 3.1 Notwendige Bedingungen

Wir leiten Bedingungen her, die ein lokales Minimum  $\hat{x}$  erfüllen muss. Anschaulich liegt ein lokales Minimum vor, wenn es keine Richtung gibt, in die es (lokal) „bergab“ geht. Gibt es in  $\hat{x}$  hingegen eine Richtung, in die  $f$  fällt, so kann  $\hat{x}$  kein lokales Minimum sein, da ein kurzer Schritt in Richtung dieser Abstiegsrichtung genügen würde, um den Zielfunktionswert zu reduzieren. Eine formale Definition ist gegeben durch

**Definition 3.1.1** (Abstiegsrichtung)

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

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

Eine hinreichende Bedingung für eine Abstiegsrichtung liefert der folgende Hilfssatz.

**Hilfsatz 3.1.2**

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  stetig differenzierbar in  $x$ . Gilt  $\nabla f(x)^\top d < 0$ , so ist  $d$  eine Abstiegsrichtung.

tung von  $f$  in  $x$ .

**Beweis:** Die Richtungsableitung von  $f$  in  $x$  in Richtung  $d$  lautet

$$\lim_{\alpha \downarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \nabla f(x)^\top d < 0.$$

Hieraus folgt die Existenz eines  $\bar{\alpha} > 0$  mit  $f(x + \alpha d) - f(x) < 0$  für alle  $0 < \alpha \leq \bar{\alpha}$ , d.h.  $d$  ist Abstiegsrichtung. ■

### Bemerkung 3.1.3

- Die Bedingung  $\nabla f(x)^\top d < 0$  bedeutet geometrisch, dass der Winkel zwischen Gradient und Abstiegsrichtung zwischen  $90^\circ$  und  $270^\circ$  liegt. Dies lässt sich aus der für Vektoren  $a, b \in \mathbb{R}^n$  allgemein gültigen Beziehung

$$\cos \angle(a, b) = \frac{a^\top b}{\|a\| \cdot \|b\|} = \frac{\langle a, b \rangle}{\|a\| \cdot \|b\|}$$

ableiten.

- Die Bedingung in Hilfssatz 3.1.2 ist nur hinreichend für eine Abstiegsrichtung aber nicht notwendig. Betrachte z.B. ein striktes lokales Maximum  $\hat{x}$  mit  $\nabla f(\hat{x}) = 0$ . Für alle Richtungen  $d \in \mathbb{R}^n$  gilt dann  $\nabla f(\hat{x})^\top d = 0$ . Andererseits ist in einem strikten lokalen Maximum jede Richtung  $d \in \mathbb{R}^n$  eine Abstiegsrichtung. ■

Damit können wir eine erste notwendige Bedingung formulieren: Ist  $\hat{x}$  ein lokales Minimum, so gibt es keine Abstiegsrichtung von  $f$  in  $\hat{x}$ . Notwendig muss dann  $\nabla f(\hat{x})^\top d \geq 0$  gelten für alle Richtungen  $d \in \mathbb{R}^n$ , denn andernfalls wäre  $d$  Abstiegsrichtung. Die Bedingung  $\nabla f(\hat{x})^\top d \geq 0$  ist für jeden beliebigen Vektor  $d \in \mathbb{R}^n$  nur dann erfüllt, wenn bereits  $\nabla f(\hat{x}) = 0$  gilt. Damit ist bewiesen:

### Satz 3.1.4 (Notwendige Bedingung erster Ordnung)

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  stetig differenzierbar und  $\hat{x}$  ein lokales Minimum von  $f$ . Dann gilt

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

Im Spezialfall  $n = 1$  gilt  $f'(\hat{x}) = 0$ .

**Definition 3.1.5** (stationärer Punkt)

Jeder Punkt  $x$  mit  $\nabla f(x) = 0$  heißt **stationärer Punkt von  $f$** .

Das folgende Beispiel zeigt, dass die Bedingung  $\nabla f(\hat{x}) = 0$  nur notwendig, aber nicht hinreichend für ein lokales Minimum ist. Stationäre Punkte sind also nicht automatisch lokale Minima, sondern nur mögliche Kandidaten!

**Beispiel 3.1.6**

Betrachte die Funktionen  $f(x) = x^2$  und  $g(x) = -x^2$  und  $\hat{x} = 0$ .

In  $\hat{x} = 0$  gilt  $f'(\hat{x}) = 0$  und  $g'(\hat{x}) = 0$ . Allerdings besitzt  $f$  in  $\hat{x} = 0$  ein globales Minimum, während  $g$  dort ein globales Maximum besitzt. ■

Wir wollen eine weitere notwendige Bedingung herleiten. Dazu greifen wir auf die Taylorentwicklung zurück. Ist  $\hat{x}$  ein stationärer Punkt von  $f$ , so gilt  $\nabla f(\hat{x}) = 0$  und die Taylorentwicklung reduziert sich zu

$$f(x) \approx f(\hat{x}) + \frac{1}{2}(x - \hat{x})^\top H_f(\hat{x})(x - \hat{x}).$$

Damit besitzt  $f$  lokal dieselben Eigenschaften wie die quadratische Funktion

$$q(x) := f(\hat{x}) + \frac{1}{2}(x - \hat{x})^\top H_f(\hat{x})(x - \hat{x}). \quad (3.2)$$

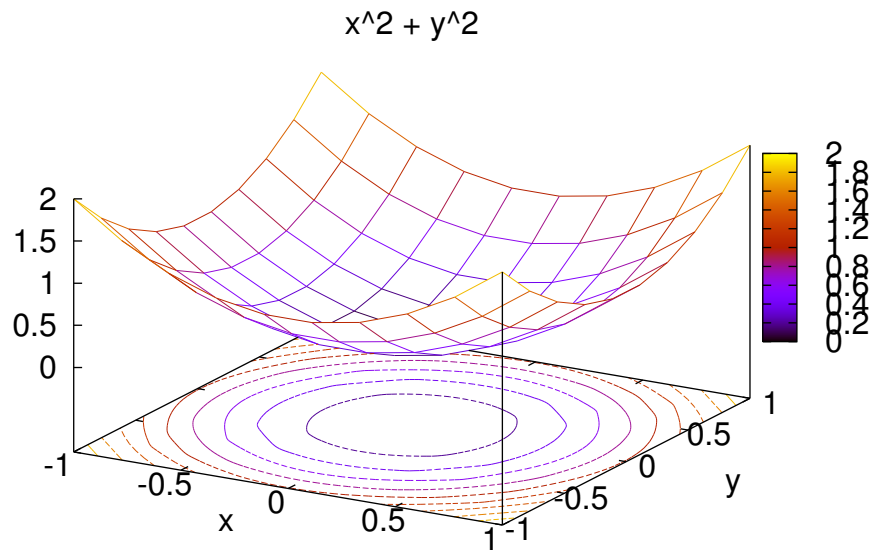
Für diese quadratische Funktion gilt

$$\begin{aligned} \nabla q(x) &= H_f(\hat{x})(x - \hat{x}), \\ H_q(x) &= H_f(\hat{x}). \end{aligned}$$

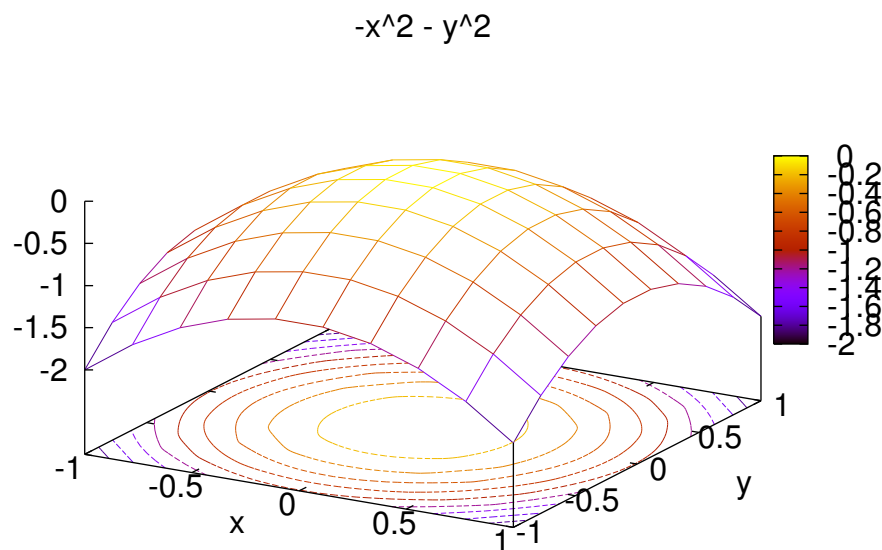
Die Hessematrix  $H_f(\hat{x})$  beschreibt anschaulich die Krümmung der Funktion  $f$  in  $\hat{x}$ .

Wir veranschaulichen mögliche Fälle im  $\mathbb{R}^2$ :

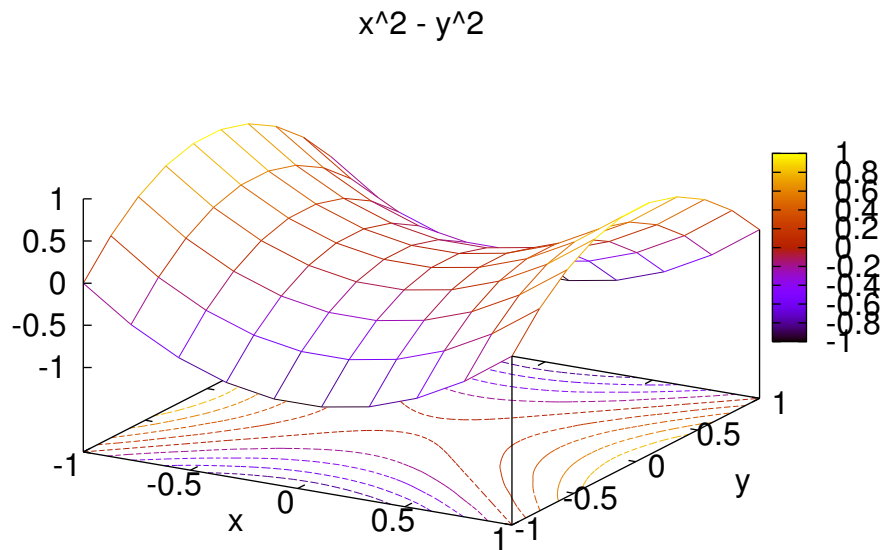
- $f(x, y) = x^2 + y^2$ ,  $H_f(\hat{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$  positiv definit:



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



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



Wir können nun eine weitere notwendige Bedingung formulieren, die besagt, dass die quadratische Form  $q$  in einem lokalen Minimum nicht negativ gekrümmt sein muss.

**Satz 3.1.7 (Notwendige Bedingung zweiter Ordnung)**

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und  $\hat{x}$  ein lokales Minimum von  $f$ . Dann ist die Hessematrix  $H_f(\hat{x})$  positiv semidefinit.

Im Spezialfall  $n = 1$  gilt  $f''(\hat{x}) \geq 0$ .

**Beweis:** Wäre  $H_f(\hat{x})$  nicht positiv semidefinit, so gäbe es ein  $d \in \mathbb{R}^n$ ,  $d \neq 0$ , mit  $d^\top H_f(\hat{x})d < 0$ . Mit dem Satz von Taylor folgt unter Ausnutzung von  $\nabla f(\hat{x}) = 0$  die Beziehung

$$f(\hat{x} + td) = f(\hat{x}) + \frac{1}{2}t^2 d^\top H_f(\hat{x} + \xi_t d)d$$

mit einem  $\xi_t \in (0, t)$ . Für hinreichend kleines  $t > 0$  folgt aus der Stetigkeit von  $H_f(\cdot)$ , dass auch  $d^\top H_f(\hat{x} + \xi_t d)d < 0$  gilt. Somit folgt  $f(\hat{x} + td) < f(\hat{x})$  für alle hinreichend kleinen Werte  $t > 0$ . Dies ist ein Widerspruch zur lokalen Minimalität von  $\hat{x}$ . ■

**Beispiel 3.1.8**

Betrachte wieder die Funktionen  $f(x) = x^2$  und  $g(x) = -x^2$ . In beiden Fällen ist  $\hat{x} = 0$  ein stationärer Punkt.

Wegen  $f''(0) = 2 > 0$  erfüllt  $f$  die notwendige Bedingung zweiter Ordnung.

Wegen  $g''(0) = -2 < 0$  erfüllt  $g$  die notwendige Bedingung zweiter Ordnung nicht, d.h.

$\hat{x} = 0$  ist kein lokales Minimum der Funktion  $g$ . ■

Das folgende Beispiel zeigt, dass die notwendige Bedingung zweiter Ordnung ebenfalls nicht hinreichend für lokale Optimalität ist.

### Beispiel 3.1.9

Die Funktion  $f(x) = x^3$  erfüllt ebenfalls  $f'(0) = 0$  und sogar  $f''(0) = 0$ , ist also positiv semidefinit. Somit sind beide notwendigen Bedingungen erfüllt. Jedoch besitzt sie weder ein lokales noch globales Minimum oder Maximum in  $x = 0$ . ■

### Beispiel 3.1.10

Betrachte  $f(x_1, x_2) := x_1^2 - x_2^4$ . Es gilt

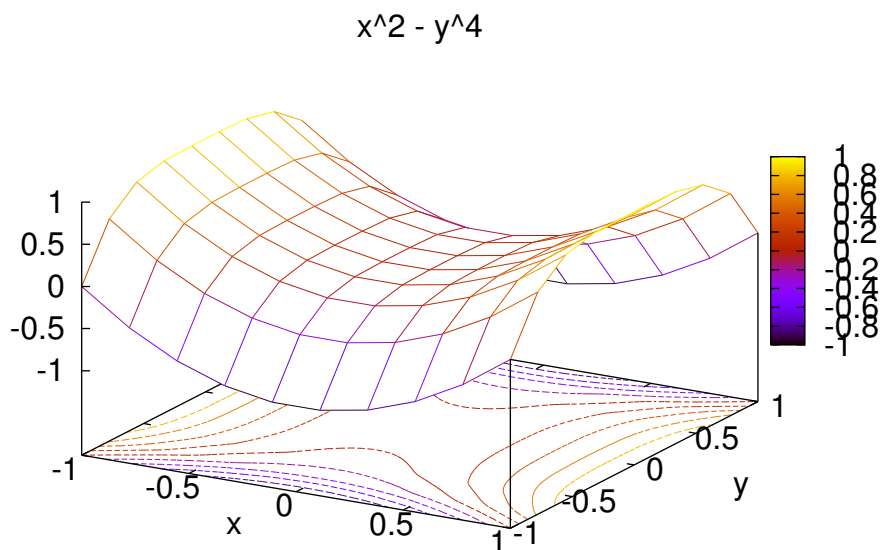
$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 \\ -4x_2^3 \end{pmatrix}.$$

Daher ist  $\hat{x} = (0, 0)^\top$  der einzige stationäre Punkt von  $f$ .

Weiter gilt

$$H_f(x_1, x_2) = \begin{pmatrix} 2 & 0 \\ 0 & -12x_2^2 \end{pmatrix} \quad \text{bzw.} \quad H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

Die Hessematrix ist positiv semidefinit, d.h.  $\hat{x}$  erfüllt die notwendige Bedingung zweiter Ordnung. Allerdings ist  $\hat{x}$  kein lokales Minimum, sondern ein Sattelpunkt.





**Fazit:**

Bestimmt man alle Punkte, die die notwendigen Bedingungen  $\nabla f(x) = 0$  und „ $H_f(x)$  positiv semidefinit“ erfüllen, so sind diese Punkte lediglich Kandidaten für ein (lokales) Minimum.

Wir werden später erkennen, dass die meisten numerischen Verfahren versuchen, stationäre Punkte zu approximieren.

**3.1.1 Hinreichende Bedingungen für unrestringierte Optimierungsprobleme**

Um entscheiden zu können, ob ein Punkt  $\hat{x}$ , der die notwendigen Bedingungen erster und zweiter Ordnung erfüllt, tatsächlich ein (lokales) Minimum ist, werden hinreichende Bedingungen benötigt.

**Satz 3.1.11**

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und  $\hat{x}$  ein stationärer Punkt mit positiv definiter Hessematrix  $H_f(\hat{x})$ . Dann ist  $\hat{x}$  ein striktes lokales Minimum von  $f$ .

Im Spezialfall  $n = 1$  muß  $f''(\hat{x}) > 0$  gelten.

**Beweis:** Taylorentwicklung von  $f$  um  $\hat{x}$  liefert

$$f(\hat{x} + d) = f(\hat{x}) + \underbrace{\nabla f(\hat{x})^\top}_{=0} d + \frac{1}{2} d^\top H_f(\hat{x} + \xi d) d$$

mit  $\xi \in (0, 1)$ . Da  $H_f(\hat{x})$  positiv definit ist, gibt es ein  $\alpha > 0$ , so dass für alle  $d \in \mathbb{R}^n$  gilt

$$d^\top H_f(\hat{x}) d \geq \alpha d^\top d.^1$$

Damit folgt mit der Cauchy-Schwarzschen Ungleichung<sup>2</sup> die Beziehung

$$\begin{aligned} f(\hat{x} + d) &= f(\hat{x}) + \frac{1}{2} d^\top H_f(\hat{x} + \xi d) d \\ &= f(\hat{x}) + \frac{1}{2} d^\top H_f(\hat{x}) d + \frac{1}{2} d^\top (H_f(\hat{x} + \xi d) - H_f(\hat{x})) d \\ &\geq f(\hat{x}) + \frac{1}{2} (\alpha - \|H_f(\hat{x} + \xi d) - H_f(\hat{x})\|) \|d\|^2. \end{aligned}$$

Aufgrund der Stetigkeit von  $H_f(\hat{x})$  gilt  $\|H_f(\hat{x} + \xi d) - H_f(\hat{x})\| \rightarrow 0$  für  $d \rightarrow 0$ . Es folgt  $f(\hat{x} + d) > f(\hat{x})$  für alle hinreichend kleinen Vektoren  $d \neq 0$  und somit ist  $\hat{x}$  striktes lokales Minimum von  $f$ . ■

<sup>1</sup>Diese Behauptung ist nur für endlichdimensionale Vektorräume richtig. Sie basiert auf der Kompaktheit der Einheitskugel in endlichdimensionalen Vektorräumen. In unendlichdimensionalen Vektorräumen gilt dies nicht mehr.

<sup>2</sup> Cauchy-Schwarzsche Ungleichung:  $|\langle a, b \rangle| = |a^\top b| \leq \|a\| \cdot \|b\|$ ,  $a, b \in \mathbb{R}^n$

**Bemerkung 3.1.12**

Die hinreichende Bedingung ist i.a. nicht notwendig. Ein Gegenbeispiel liefert die Funktion  $f(x_1, x_2) := x_1^2 + x_2^4$ . Hier ist  $\hat{x} = 0$  ein striktes globales Minimum, die Hesse-Matrix ist jedoch nicht positiv definit. ■

**Beispiel 3.1.13**

Für die Funktion

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

berechnen wir den Gradienten

$$\nabla f(x, y) = \begin{pmatrix} y^2 + 3x^2 + 2x \\ 2y(x - 1) \end{pmatrix}$$

und die Hessematrix

$$H_f(x, y) = \begin{pmatrix} 6x + 2 & 2y \\ 2y & 2(x - 1) \end{pmatrix}.$$

Aus  $\nabla f(x, y) = 0$  ergeben sich somit die stationären Punkte  $(x^0, y^0) = (0, 0)$  und  $(x^1, y^1) = (-2/3, 0)$ .

Für die zugehörigen Hesse-Matrizen erhält man

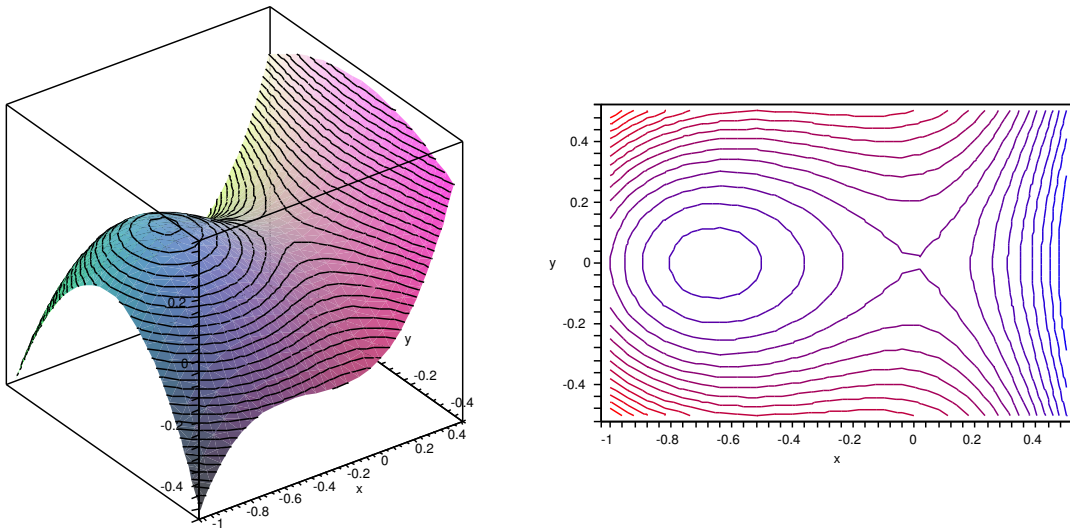
$$H_f(x^0, y^0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \text{ indefinit} \Rightarrow \begin{pmatrix} x^0 \\ y^0 \end{pmatrix} \text{ Sattelpunkt}$$

$$H_f(x^1, y^1) = \begin{pmatrix} -2 & 0 \\ 0 & -\frac{10}{3} \end{pmatrix} \text{ negativ definit} \Rightarrow \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} \text{ striktes lokales Maximum}$$

MAPLE-Befehle

```
plot3d(y^2*(x-1)+x^2*(x+1), x=-1..0.5, y=-0.5..0.5, axes=boxed,
style=PATCHCONTOUR, contours = 40, shading=XYZ);
```

```
contourplot(y^2*(x-1)+x^2*(x+1), x=-1..0.5, y=-0.5..0.5, contours=20,
coloring=[red,blue], scaling=constrained, axes=boxed);
```



### 3.2 Allgemeine Abstiegsverfahren

Wir konstruieren ein allgemeines Konzept zur Minimierung einer stetig differenzierbaren Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Viele der später diskutierten Verfahren basieren auf diesem Konzept. Der folgende Algorithmus basiert auf der Verwendung von Abstiegsrichtungen in Kombination mit einer Schrittweitensteuerung:

#### Algorithmus 3.2.1 (Abstiegsverfahren)

- (i) Bestimme einen Startpunkt  $x^{[0]} \in \mathbb{R}^n$  und setze  $i = 0$ .
- (ii) Falls ein Abbruchkriterium erfüllt ist, STOP.
- (iii) Berechne eine Abstiegsrichtung  $d^{[i]}$  und eine Schrittweite  $\alpha_i > 0$ , so dass

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

gilt und setze  $x^{[i+1]} = x^{[i]} + \alpha_i \cdot d^{[i]}$ .

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

Natürlich ist dieser Algorithmus lediglich von konzeptioneller Art, da die wesentlichen Komponenten (Bestimmung der Abstiegsrichtung, der Schrittweite und geeigneter Ab-

bruchkriterien) noch nicht näher beschrieben wurden und sehr viel Freiraum lassen. Im Vorgriff auf später seien einige Beispiele für Suchrichtungen genannt:

- **Gradientenverfahren** oder **Verfahren des steilsten Abstiegs**: Die Richtung des steilsten Abstiegs

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

ist wegen  $\nabla f(x^{[i]})^\top d^{[i]} = -\|\nabla f(x^{[i]})\|^2$  eine Abstiegsrichtung in nicht stationären Punkten. Diese naheliegende Wahl der Suchrichtung muss aber nicht die effizienteste sein.

- **Newtonverfahren**:

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

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

### Beispiel 3.2.2

Wir verwenden das (unskalierte) Gradientenverfahren zur Minimierung der Funktion

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

Die Bestimmung der Schrittweite erfolge mit exakter Liniensuche, d.h. mit

$$\alpha := \operatorname{argmin}\{f(x + \alpha d) : \alpha > 0\}.$$

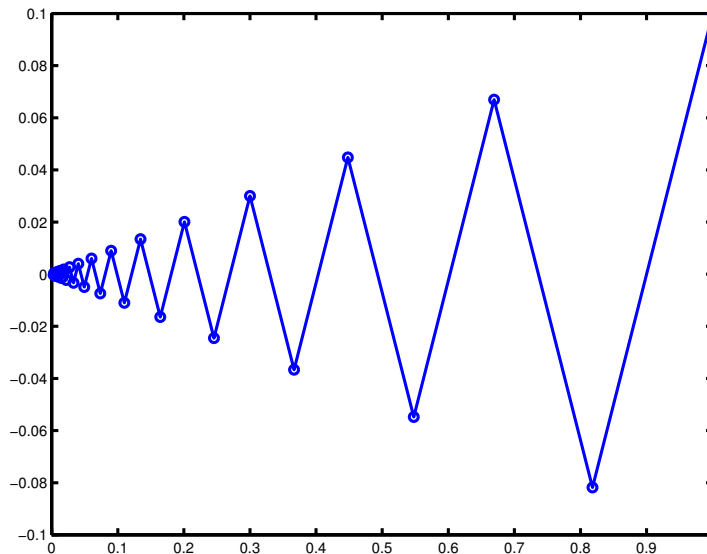
Wir erhalten

$$\begin{aligned} \nabla f(x_1, x_2) &= (2x_1, 20x_2)^\top \\ d &= -(2x_1, 20x_2)^\top \\ \varphi(\alpha) &= (x_1 + \alpha d_1)^2 + 10(x_2 + \alpha d_2)^2 \\ \varphi'(\alpha) &= 2(x_1 + \alpha d_1)d_1 + 20(x_2 + \alpha d_2)d_2 \stackrel{!}{=} 0 \end{aligned}$$

Damit:

$$\alpha := -\frac{x_1 d_1 + 10 x_2 d_2}{d_1^2 + 10 d_2^2}, \quad x_i^{\text{neu}} := x_i + \alpha d_i, \quad (i = 1, 2).$$

Mit dem Startvektor  $x^{[0]} := (1, 0.1)^\top$  benötigt das Verfahren 63 Iterationen um das Abbruchkriterium  $\|\nabla f(x)\|_2 \leq 10^{-5}$  zu erfüllen.



Würden wir dagegen die Newtonrichtung verwenden mit

$$H_f = \begin{pmatrix} 2 & 0 \\ 0 & 20 \end{pmatrix},$$

so würde sich ergeben

$$d = -H_f^{-1} \nabla f(x) = - \begin{pmatrix} 1/2 & 0 \\ 0 & 1/20 \end{pmatrix} \begin{pmatrix} 2x_1 \\ 20x_2 \end{pmatrix} = -x,$$

sowie (bei exakter Liniensuche)  $\alpha = 1$ , so dass der erste Iterationsschritt bereits die exakte Lösung  $\hat{x} = 0$  liefert. ■

### 3.3 Liniensuche, Armijo-Regel und Schrittweitenbestimmung

Im folgenden wird vorausgesetzt, dass wir im Punkt  $x^{[i]}$  bereits eine Richtung  $d^{[i]}$  mit

$$\nabla f(x^{[i]})^\top d^{[i]} < 0$$

gefunden haben. Nach Hilfssatz 3.1.2 ist  $d^{[i]}$  somit eine Abstiegsrichtung von  $f$  in  $x^{[i]}$ . Um das allgemeine Abstiegsverfahren durchführen zu können, muß also nur noch die **Schrittweite**  $\alpha_i$  bestimmt werden.

Zur Bestimmung der Schrittweite genügt es, für  $\alpha \geq 0$  die Funktion  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$\varphi(\alpha) := f(x^{[i]} + \alpha \cdot d^{[i]})$$

zu betrachten. Wegen

$$\varphi'(0) = \nabla f(x^{[i]})^\top d^{[i]} < 0$$

existieren Schrittweiten  $\alpha$  mit  $\varphi(\alpha) < \varphi(0)$ , vgl. Abbildung 3.1.

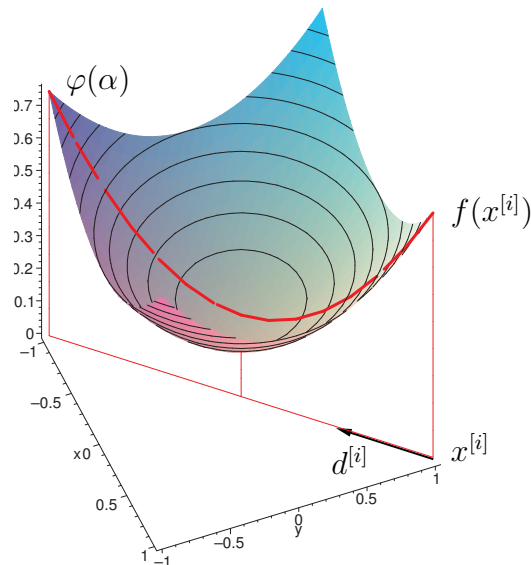


Abbildung 3.1: Eindimensionale Liniensuche ausgehend von  $x^{[i]}$  in Richtung  $d^{[i]}$ .

Die gebräuchlichste Schrittweitenstrategie ist die **Armijo-Regel**, die sich sehr leicht implementieren lässt.

### Algorithmus 3.3.1 (Armijo-Regel)

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

(ii) Falls die Bedingung

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

erfüllt ist, setze  $\alpha_i := \alpha$  und beende das Verfahren.

Andernfalls gehe zu (iii).

(iii) Setze  $\alpha := \beta \cdot \alpha$  und gehe zu (ii).

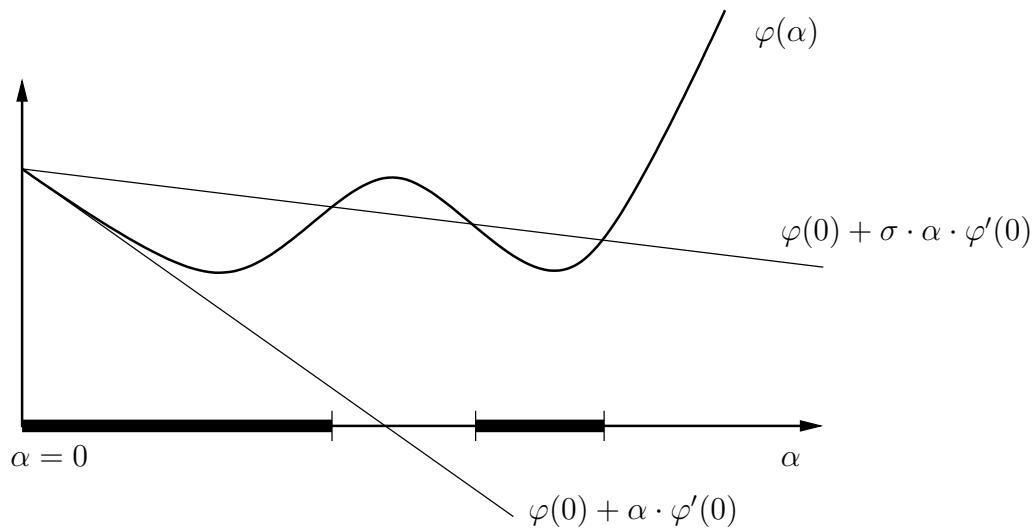


Abbildung 3.2: Schrittweiten, die der Armijoregel genügen.

Die fett eingezeichneten Intervalle in Abbildung 3.2 erfüllen die Armijo-Bedingung. Man kann zeigen, dass die Armijo-Regel wohldefiniert ist, d.h. dass der Algorithmus stets nach endlich vielen Schritten terminiert.

In der Praxis liefern die Werte  $\sigma = 0.01$  und  $\beta = 0.9$  häufig gute Ergebnisse. Genauere Untersuchungen finden sich bei Spellucci [Spe93] und Alt [Alt02].

Das Gradientenverfahren (Verfahren des steilsten Abstiegs) zur Minimierung einer stetig differenzierbaren Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  mit der Armijo-Regel zur Schrittweitenbestimmung lautet dann wie folgt:

### Algorithmus 3.3.2 (Gradientenverfahren mit Liniensuche)

- (0) Wähle eine Startschätzung  $x^{[0]} \in \mathbb{R}^n$ , einen Schrittweitenfaktor  $\beta \in (0, 1)$ , einen Steigungsfaktor  $\sigma \in (0, 1)$  und setze den Iterationsindex auf  $k = 0$ .
- (1) Falls  $\|\nabla f(x^{[k]})\| \approx 0$ , STOP.
- (2) Berechne die Suchrichtung  $d^{[k]} = -\nabla f(x^{[k]})$ .
- (3) Bestimme eine Schrittweite  $\alpha_k$  mit der Armijo-Regel.
- (4) Bestimme  $x^{[k+1]} = x^{[k]} + \alpha_k d^{[k]}$ , setze  $k := k + 1$  und gehe zu (1).

Für dieses Verfahren kann folgender Konvergenzsatz bewiesen werden:

**Satz 3.3.3**

Wird durch das Gradientenverfahren für die stetig differenzierbare Funktion  $f$ , Startwert  $x^{[0]} \in \mathbb{R}^n$  eine nicht abbrechende Folge  $\{x^{[k]}\}_{k \in \mathbb{N}}$  definiert, so gilt für jeden Häufungspunkt  $x^*$  dieser Folge  $\nabla f(x^*) = 0$ .

Das Gradientenverfahren ist einfach durchzuführen, allerdings ist die Konvergenz des Verfahrens sehr langsam (lineare Konvergenz), so dass in der Regel viele Iterationen benötigt werden. Man kann unter geeigneten Annahmen zeigen, dass für den Fortschritt in der Zielfunktion die Beziehung

$$f(x^{[k+1]}) - f(x^*) \leq \left( \frac{\kappa - 1}{\kappa + 1} \right)^2 (f(x^{[k]}) - f(x^*))$$

mit der spektralen Konditionszahl  $\kappa := \lambda_{\max}/\lambda_{\min}$  gilt. Dabei bezeichnen  $\lambda_{\max}$  bzw.  $\lambda_{\min}$  den größten bzw. kleinsten Eigenwert der Hessematrix  $H_f(x^*)$  in einem lokalen Minimum  $x^*$ . Die Konvergenz ist also umso langsamer, je größer die Kondition  $\kappa$  ist. Für  $\kappa = 100$  ergibt sich z.B. der Verkleinerungsfaktor zu ungefähr 0.96, so dass kaum eine Verbesserung in jedem Gradientenschritt eintritt.

**3.4 Newton-Verfahren**

Ein generelles Prinzip zur Bestimmung einer Abstiegsrichtung besteht darin, die Funktion  $f$  lokal in der Nähe der Iterierten  $x^{[i]}$  durch eine Approximation  $\hat{f}(y)$  zu ersetzen und diese dann ersatzweise zu minimieren, um eine geeignete Suchrichtung zu bekommen.

**Beispiel 3.4.1 (Lineare Approximation)**

Linearisierung von  $f$  in  $x$  ergibt die Funktion

$$\hat{f}(y) = f(x) + \nabla f(x)^\top (y - x).$$

Allerdings besitzt diese lineare Funktion i.a. kein Minimum auf  $\mathbb{R}^n$ . Daher schränkt man zusätzlich die Werte von  $y$  ein, indem man fordert

$$\|y - x\| \leq 1.$$

Zur Bestimmung einer Suchrichtung  $d = y - x$  betrachtet man das Optimierungsproblem

$$\text{Minimiere } \hat{f}(y) \quad \text{u. d. N.} \quad \|y - x\| \leq 1.$$

Wegen

$$\|\nabla f(x)^\top (y - x)\| \leq \|\nabla f(x)\| \cdot \|y - x\| \leq \|\nabla f(x)\|$$



löst  $\hat{y}$  in  $d = \hat{y} - x := -\nabla f(x)/\|\nabla f(x)\|$  das Optimierungsproblem. Man erhält also das Gradientenverfahren. Diese Herleitung begründet auch die Bezeichnung als Verfahren des steilsten Abstiegs. ■

Es ist nun naheliegend,  $f$  quadratisch zu approximieren, um ein weiteres Verfahren zu erhalten.

**Beispiel 3.4.2 (Quadratische Approximation)**

$f$  wird lokal durch

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

approximiert. Ist die Hessematrix  $H_f(x)$  positiv definit, so besitzt  $\hat{f}$  ein eindeutig bestimmtes Minimum  $\hat{y}$ , welches durch das lineare Gleichungssystem

$$\nabla \hat{f}(\hat{y}) = \nabla f(x) + H_f(x)(\hat{y} - x) = 0$$

bestimmt ist. Minimierung der quadratischen Approximation  $\hat{f}(y)$  führt also auf die Suchrichtung

$$d = \hat{y} - x = -H_f(x)^{-1}\nabla f(x). \quad (3.3)$$

Wegen

$$\nabla f(x)^\top d = -\nabla f(x)^\top H_f(x)^{-1}\nabla f(x) < 0$$

für  $\nabla f(x) \neq 0$  ist  $d$  eine Abstiegsrichtung von  $f$  in  $x$ . ■

Beispiel 3.4.2 beweist

**Hilfsatz 3.4.3**

Für eine zweimal stetig differenzierbare Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  führt das Prinzip der lokalen Minimierung mit quadratischem Modell auf das folgende lineare Gleichungssystem für die Suchrichtung  $d$ :

$$H_f(x)d = -\nabla f(x). \quad (3.4)$$

Ist die Hessematrix  $H_f(x)$  positiv definit, so ist  $d$  eine Abstiegsrichtung.

Die Annahme, dass die Hessematrix  $H_f(x)$  positiv definit ist – zumindest für alle vom Abstiegsverfahren erzeugten Punkte –, ist durchaus restriktiv und keinesfalls immer erfüllt. I.A. ist das Newtonverfahren daher kein Abstiegsverfahren und eine Schrittweitenstrategie kann mitunter nicht angewendet werden.

Trotzdem ist das Newtonverfahren sinnvoll, denn es kann alternativ dadurch motiviert werden, dass die notwendige Bedingung

$$\nabla f(\hat{x}) = 0 \quad (3.5)$$

für ein lokales Minimum  $\hat{x}$  von  $f$  als **Nullstellenproblem** (bzw. nichtlineares Gleichungssystem) für die Funktion  $g(x) := \nabla f(x)$  aufgefasst wird. Anwendung des aus der Numerik bekannten Newtonverfahrens liefert die Iterationsvorschrift

$$\begin{aligned} g'(x^{[i]})d^{[i]} &= -g(x^{[i]}), \\ x^{[i+1]} &= x^{[i]} + d^{[i]}, \quad i = 0, 1, 2, \dots \end{aligned}$$

$g'(\cdot)$  bezeichnet die Jacobi-Matrix von  $g$ . Die erste Gleichung lautet insbesondere

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

Sie ist identisch mit (3.4).

Wir widmen uns nun der Konvergenz des **lokalen Newtonverfahrens**, welches sich dadurch auszeichnet, dass stets die Schrittweite  $\alpha_i = 1$  gewählt wird.

#### Algorithmus 3.4.4 (Lokales Newtonverfahren)

- (i) Wähle einen Startvektor  $x^{[0]} \in \mathbb{R}^n$  und setze  $i = 0$ .
- (ii) Falls ein Abbruchkriterium erfüllt ist, STOP.
- (iii) Berechne (falls möglich) die Suchrichtung  $d^{[i]}$  als Lösung des Gleichungssystems (3.4) mit  $x = x^{[i]}$ , setze  $x^{[i+1]} = x^{[i]} + d^{[i]}$ ,  $i := i + 1$  und gehe zu (ii).

#### Definition 3.4.5 (lineare, superlineare und quadratische Konvergenz)

Sei  $\{x^{[i]}\}_{i \in \mathbb{N}}$  eine Folge.

- (a)  $\{x^{[i]}\}_{i \in \mathbb{N}}$  konvergiert **linear** gegen  $\hat{x}$ , falls es ein  $0 < c < 1$  gibt mit

$$\|x^{[i+1]} - \hat{x}\| \leq c \cdot \|x^{[i]} - \hat{x}\|$$

für alle hinreichend großen  $i$ .

(b)  $\{x^{[i]}\}_{i \in \mathbb{N}}$  konvergiert **superlinear** gegen  $\hat{x}$ , falls es eine Nullfolge  $\{c_i\}_{i \in \mathbb{N}}$ ,  $c_i \downarrow 0$  gibt mit

$$\|x^{[i+1]} - \hat{x}\| \leq c_i \cdot \|x^{[i]} - \hat{x}\| \quad \forall i \in \mathbb{N},$$

d.h.

$$\|x^{[i+1]} - \hat{x}\| = o(\|x^{[i]} - \hat{x}\|).$$

(c)  $\{x^{[i]}\}_{i \in \mathbb{N}}$  konvergiert **von der Ordnung**  $p \geq 2$  gegen  $\hat{x}$ , falls es ein  $c > 0$  gibt mit

$$\|x^{[i+1]} - \hat{x}\| \leq c \cdot \|x^{[i]} - \hat{x}\|^p \quad \forall i \in \mathbb{N},$$

d.h.

$$\|x^{[i+1]} - \hat{x}\| = \mathcal{O}(\|x^{[i]} - \hat{x}\|^p).$$

Im Fall  $p = 2$  heißt die Folge **quadratisch** konvergent.

Der folgende Satz fasst die Konvergenzeigenschaften des Newtonverfahrens zusammen.

**Satz 3.4.6 (Konvergenzsatz für das Newtonverfahren)**

$f : \mathbb{R}^n \rightarrow \mathbb{R}$  besitze ein (lokales) Minimum  $\hat{x} \in \mathbb{R}^n$ .  $f$  sei zweimal stetig differenzierbar und  $H_f(\hat{x})$  sei invertierbar.

Dann gibt es ein  $r > 0$ , so dass das lokale Newton-Verfahren für alle Startwerte  $x^{[0]} \in U_r(\hat{x})$  wohldefiniert ist, und die Folge  $\{x^{[i]}\}_{i \in \mathbb{N}}$  konvergiert superlinear gegen  $\hat{x}$  (lokal superlineare Konvergenz).

$\hat{x}$  ist das einzige lokale Minimum in  $U_r(\hat{x})$ .

Erfüllt  $H_f$  zusätzlich noch die Lipschitz-Bedingung

$$\|H_f(x) - H_f(y)\| \leq L \cdot \|x - y\| \quad \forall x, y \in \mathbb{R}^n,$$

so ist die Folge  $\{x^{[i]}\}$  sogar quadratisch konvergent (lokal quadratische Konvergenz).

**Bemerkung 3.4.7**

- Das Newtonverfahren lässt sich also entweder als **Nullstellenverfahren** zur Bestimmung eines stationären Punktes oder als **Abstiegsverfahren** (bei positiv definiten Hessematrix) interpretieren, bei dem stets mit der Schrittweite  $\alpha_i = \alpha = 1$  gearbeitet wird.
- Nach den obigen Erläuterungen liefert das Newtonverfahren – falls konvergent –

lediglich stationäre Punkte. Es könnte daher auch gegen ein Maximum oder einen Sattelpunkt konvergieren.

- Es gibt Varianten des Newtonverfahren, sogenannte **Quasi-Newton-Verfahren** oder **Variable-Metrik-Verfahren**, bei denen die Hessematrix in Schritt  $i$  durch eine geeignete Matrix  $H_i$  ersetzt wird, die in jedem Schritt aufdatiert wird.

■

### Beispiel 3.4.8 (Funktion von Himmelblau)

Wir minimieren

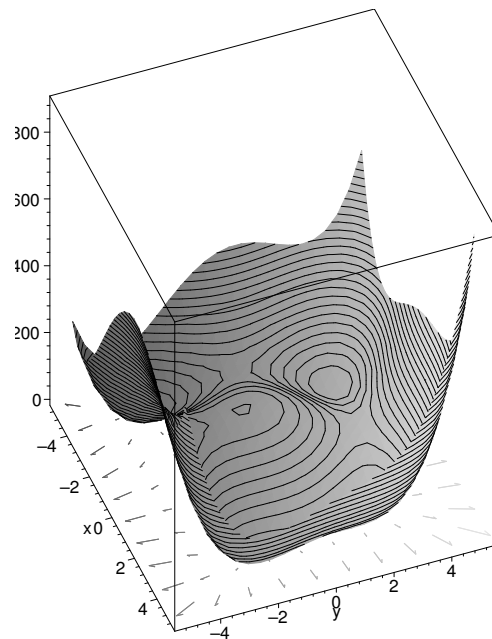
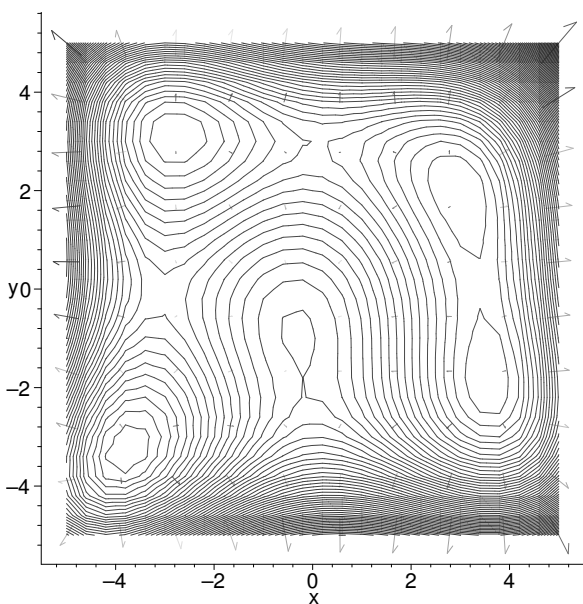
$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

mit dem Newtonverfahren. Es gilt

$$\nabla f(x, y) = \begin{pmatrix} 4x(x^2 + y - 11) + 2(x + y^2 - 7) \\ 2(x^2 + y - 11) + 4y(x + y^2 - 7) \end{pmatrix},$$

$$H_f(x, y) = \begin{pmatrix} 12x^2 + 4y - 42 & 4(x + y) \\ 4(x + y) & 4x + 12y^2 - 26 \end{pmatrix}.$$

- 4 lokale Minimalstellen (zugleich global) mit Funktionswert 0; darunter der Punkt  $(3, 2)^\top$ .
- 4 Sattelpunkte
- ein lokales Maximum in  $(-0.270845, -0.923039)^\top$



Wir diskutieren einige Varianten des Newtonverfahrens. Die Iterationsvorschrift für einen Startwert  $x^{[0]}$  und  $g(x) = \nabla f(x)$  lautet

$$\begin{aligned} A(x^{[i]})d^{[i]} &= -g(x^{[i]}), \\ x^{[i+1]} &= x^{[i]} + d^{[i]}, \quad i = 0, 1, 2, \dots \end{aligned}$$

Die Matrix  $A(\cdot) \in \mathbb{R}^{n \times n}$  wird dabei auf die folgenden Arten berechnet:

- $A(x) := H_f(x)$  (klassisches Newtonverfahren);
- finite Differenzen-Approximation der Hessematrix  $H_f(x)$ , d.h.  $A(x) = (A_{jk}(x))_{j,k=1,\dots,n}$  ist gegeben durch

$$A_{jk}(x) := \frac{g_j(x + h_i e_k) - g_j(x)}{h_i}, \quad j, k = 1, \dots, n,$$

wobei  $e_k$  den  $k$ -ten Einheitsvektor bezeichnet. Die Schrittweite  $h_i$  sei durch  $h_i = \frac{0.1}{i+1}$  gegeben, wobei  $i$  den Iterationsindex des Newtonverfahrens bezeichnet. Beachte, dass  $h_i \rightarrow 0$  für  $i \rightarrow \infty$  gilt.

- vereinfachtes Newtonverfahren mit  $A(x^{[i]}) := H_f(x^{[0]})$  für  $i = 0, 1, 2, \dots$

Als Abbruchkriterium des Newtonverfahrens verwenden wir jeweils  $\|\nabla f(x, y)\| \leq 10^{-13}$  und als Startpunkt  $(x^{[0]}, y^{[0]}) = (4, 2.5)$ .

Das klassische Newtonverfahren liefert das folgende Ergebnis:

ITER	X(1)	X(2)	GRAD	DX	P=1	P=2
0	0.4000000E+01	0.2500000E+01	0.1351240E+03	0.0000000E+00	0.0000000E+00	0.0000000E+00
1	0.3281417E+01	0.2056664E+01	0.2617493E+02	0.8443389E+00	0.2567589E+00	0.2296521E+00
2	0.3035131E+01	0.1988137E+01	0.2424694E+01	0.2556422E+00	0.1291689E+00	0.4499638E+00
3	0.3000634E+01	0.1999744E+01	0.4203618E-01	0.3639711E-01	0.1844451E-01	0.4974264E+00
4	0.3000000E+01	0.2000000E+01	0.1406811E-04	0.6837008E-03	0.3217815E-03	0.4704954E+00
5	0.3000000E+01	0.2000000E+01	0.1500787E-11	0.2200729E-06	0.1109167E-06	0.5039996E+00
6	0.3000000E+01	0.2000000E+01	0.0000000E+00	0.2440976E-13	0.0000000E+00	0.0000000E+00

Die letzten beiden Spalten testen die Folge  $(x^{[i]}, y^{[i]})^\top$ ,  $i = 0, 1, 2, \dots$  auf lineare ( $P=1$  strebt gegen Wert  $\neq 0$ ,  $P=2$  explodiert), superlineare ( $P=1$  strebt gegen 0,  $P=2$  explodiert) und quadratische Konvergenz ( $P=1$  strebt gegen 0,  $P=2$  strebt gegen Wert  $\neq 0$ ).

Die Rechnungen zeigen eine quadratische Konvergenz des klassischen Newtonverfahrens gegen das Minimum  $(3, 2)^\top$ .

Das Newtonverfahren mit finiten Differenzen liefert das folgende Ergebnis:

ITER	X(1)	X(2)	GRAD	DX	P=1	P=2
0	0.4000000E+01	0.2500000E+01	0.1351240E+03	0.0000000E+00	0.0000000E+00	0.0000000E+00
1	0.3300753E+01	0.2071139E+01	0.2853753E+02	0.8202854E+00	0.2764246E+00	0.2472417E+00
2	0.3044647E+01	0.1988635E+01	0.3188243E+01	0.2690673E+00	0.1490712E+00	0.4823498E+00
3	0.3001880E+01	0.1998866E+01	0.1165668E+00	0.4397377E-01	0.4765407E-01	0.1034366E+01
4	0.3000033E+01	0.1999965E+01	0.1813021E-02	0.2149084E-02	0.2198280E-01	0.1001284E+02
5	0.3000000E+01	0.1999999E+01	0.2684597E-04	0.4737814E-04	0.1856229E-01	0.3846122E+03
6	0.3000000E+01	0.2000000E+01	0.3878641E-06	0.8817947E-06	0.1574341E-01	0.1757351E+05
7	0.3000000E+01	0.2000000E+01	0.5091715E-08	0.1391334E-07	0.1351821E-01	0.9584736E+06
8	0.3000000E+01	0.2000000E+01	0.5957654E-10	0.1894045E-09	0.1182730E-01	0.6203365E+08
9	0.3000000E+01	0.2000000E+01	0.6269307E-12	0.2231274E-11	0.1056873E-01	0.4686628E+10
10	0.3000000E+01	0.2000000E+01	0.8758808E-14	0.2360669E-13	0.9316950E-02	0.3909375E+12

Die Rechnungen zeigen eine superlineare Konvergenz des Newtonverfahrens mit finiten Differenzen gegen das Minimum  $(3, 2)^T$ .

Das vereinfachte Newtonverfahren liefert das folgende Ergebnis:

ITER	X(1)	X(2)	GRAD	DX	P=1	P=2
0	0.400000E+01	0.250000E+01	0.1351240E+03	0.000000E+00	0.000000E+00	0.000000E+00
1	0.3281417E+01	0.2056664E+01	0.2617493E+02	0.8443389E+00	0.2567589E+00	0.2296521E+00
3	0.3071606E+01	0.1980562E+01	0.5151481E+01	0.6534609E-01	0.5470076E+00	0.4032719E+01
5	0.3022629E+01	0.1985541E+01	0.1403370E+01	0.1757236E-01	0.6120523E+00	0.1394972E+02
7	0.3007754E+01	0.1993240E+01	0.4467921E+00	0.6376896E-02	0.6191372E+00	0.3726386E+02
9	0.3002760E+01	0.1997232E+01	0.1540965E+00	0.2446909E-02	0.6154210E+00	0.9688894E+02
11	0.3001001E+01	0.1998926E+01	0.5514621E-01	0.9306565E-03	0.6121533E+00	0.2552127E+03
13	0.3000366E+01	0.1999593E+01	0.2006926E-01	0.3497360E-03	0.6102365E+00	0.6801452E+03
15	0.3000135E+01	0.1999848E+01	0.7361719E-02	0.1304451E-03	0.6092050E+00	0.1825138E+04
17	0.3000050E+01	0.1999943E+01	0.2710854E-02	0.4845351E-04	0.6086619E+00	0.4915933E+04
19	0.3000018E+01	0.1999979E+01	0.1000183E-02	0.1795865E-04	0.6083767E+00	0.1326686E+05
21	0.3000007E+01	0.1999992E+01	0.3693921E-03	0.6648483E-05	0.6082266E+00	0.3584082E+05
23	0.3000003E+01	0.1999997E+01	0.1364969E-03	0.2459848E-05	0.6081473E+00	0.9687759E+05
25	0.3000001E+01	0.1999999E+01	0.5045188E-04	0.9098193E-06	0.6081053E+00	0.2619347E+06
...						
61	0.3000000E+01	0.2000000E+01	0.8472333E-12	0.1511068E-13	0.6121321E+00	0.1587489E+14
63	0.3000000E+01	0.2000000E+01	0.3142520E-12	0.5625244E-14	0.6151352E+00	0.4292624E+14
65	0.3000000E+01	0.2000000E+01	0.1217558E-12	0.2090737E-14	0.6172799E+00	0.1154325E+15
66	0.3000000E+01	0.2000000E+01	0.7240833E-13	0.1287557E-14	0.6201737E+00	0.1878784E+15

Sogar das vereinfachte Newtonverfahren konvergiert, allerdings nur linear. ■

Wie Satz 3.4.6 erkennen läßt, ist das lokale Newtonverfahren i.a. nicht global (d.h. für jeden Startpunkt  $x^{[0]}$ ) konvergent. Es gibt verschiedene Möglichkeiten, um das lokale Newton-Verfahren durch Zusatzmaßnahmen zu einem global konvergenten Verfahren zu erweitern. Um die quadratische Konvergenz des lokalen Newtonverfahrens zu sichern, hat man dabei darauf zu achten, dass das globale Newtonverfahren in der Nähe der Lösung in das lokale Verfahren (also mit Schrittweite  $\alpha_i = 1$ ) übergeht.

## Kapitel 4

### Lineare Optimierung

In diesem Kapitel werden lineare Optimierungsprobleme diskutiert. Lineare Optimierungsprobleme zeichnen sich dadurch aus, dass die Zielfunktion und die Nebenbedingungen im Standardoptimierungsproblem SOP durch lineare Funktionen gegeben sind. Zur Erinnerung: Was ist eine lineare Funktion?

#### Definition 4.0.1 (lineare Funktion)

$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  heißt **linear** genau dann, wenn

$$\begin{aligned}f(x + y) &= f(x) + f(y), \\f(cx) &= cf(x)\end{aligned}$$

für alle  $x, y \in \mathbb{R}^n$  und jedes  $c \in \mathbb{R}$  gelten.

#### Beispiel 4.0.2 (lineare vs. nichtlineare Funktion)

Die Funktionen

$$\begin{aligned}f_1(x) &= x, \\f_2(x) &= 3x_1 - 5x_2, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \\f_3(x) &= c^\top x, \quad x, c \in \mathbb{R}^n, \\f_4(x) &= Ax, \quad x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n},\end{aligned}$$

sind linear (Beweis?).

Die Funktionen

$$\begin{aligned}f_5(x) &= 1, \\f_6(x) &= x + 1, \quad x \in \mathbb{R}, \\f_7(x) &= Ax + b, \quad x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, \\f_8(x) &= x^2, \\f_9(x) &= \sin(x), \\f_{10}(x) &= a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0, \quad a_i \in \mathbb{R}, (a_n, \dots, a_2, a_0) \neq 0\end{aligned}$$

sind nicht linear (Beweis?). ■

**Tatsache:**

Jede lineare Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  kann in der Form  $f(x) = Ax$  mit einer Matrix  $A \in \mathbb{R}^{m \times n}$  dargestellt werden.

**Notation:**

Wir verwenden folgende Notation:

- $x \in \mathbb{R}^n$  ist der Spaltenvektor  $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$  mit Komponenten  $x_1, x_2, \dots, x_n \in \mathbb{R}$ .

- Mit  $x^\top$  bezeichnen wir den Zeilenvektor  $(x_1, x_2, \dots, x_n)$ . Dieser heißt transponierter Vektor von  $x$ .

- $A \in \mathbb{R}^{m \times n}$  bezeichnet die Matrix  $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$  mit Einträgen  $a_{ij}$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq n$ .

- Mit  $A^\top \in \mathbb{R}^{n \times m}$  bezeichnen wir die transponierte Matrix von  $A$ , d.h.

$$A^\top = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

Ein allgemeines lineares Optimierungsproblem, auch lineares Programm genannt, lautet wie folgt.

**Definition 4.0.3** (Lineares Programm (LP))

Minimiere

$$c^\top x = c_1 x_1 + \dots + c_n x_n$$

unter den Nebenbedingungen

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \left\{ \begin{array}{l} \leq \\ \geq \\ = \end{array} \right\} b_i, \quad i = 1, \dots, m.$$



Zusätzlich seien einige (oder alle) Komponenten  $x_i$ ,  $i \in I$ , wobei  $I \subseteq \{1, \dots, n\}$  eine Indexmenge ist, beschränkt durch Vorzeichenbedingungen

$$x_i \left\{ \begin{array}{l} \leq \\ \geq \end{array} \right\} 0, \quad i \in I.$$

Hierbei sind die Daten

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

gegeben.

#### Bemerkung 4.0.4

- Die Relationen „ $\leq$ “, „ $=$ “ und „ $\geq$ “ angewendet auf Vektoren  $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$  und  $y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$  sind komponentenweise zu verstehen, z.B.:

$$x \leq y \quad \Leftrightarrow \quad x_i \leq y_i, \quad i = 1, \dots, n.$$

- Die Variablen  $x_i$ ,  $i \notin I$ , heißen **freie Variablen**, da sie nicht durch Vorzeichenbedingungen eingeschränkt sind. ■

LP ist ein lineares Programm in sehr allgemeiner Form. Um nicht ständig zwischen freien und vorzeichenbeschränkten Variablen und Gleichungs- und Ungleichungsbeschränkungen unterscheiden zu müssen, beschränken wir uns im folgenden auf LPs in der sogenannten primalen Normalform.

#### Definition 4.0.5 (Primale Normalform)

Gegeben seien Vektoren  $c \in \mathbb{R}^n$  und  $b \in \mathbb{R}^m$ , sowie die Matrix  $A \in \mathbb{R}^{m \times n}$ .

Das lineare Optimierungsproblem in primaler Normalform lautet:

$$\text{Minimiere } c^\top x \quad \text{unter den Nebenbedingungen } Ax = b, \quad x \geq 0. \quad (4.1)$$

Der zulässige Bereich des Problems ist gegeben durch die Menge

$$M := \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}.$$

**Bemerkung 4.0.6**

Die primale Normalform ist in der Regel nur für den Fall  $m < n$  interessant. Ist  $A$  invertierbar und  $m = n$ , so ist  $x$  bereits durch  $x = A^{-1}b$  eindeutig festgelegt, so dass nichts zu optimieren ist. Im Fall  $m > n$  und  $\text{Rang}(A) = n$  besitzt das Gleichungssystem  $Ax = b$  höchstens eine Lösung  $x$ , so dass ebenfalls nichts zu optimieren ist. Es gelte im folgenden also stets  $m < n$ . ■

Die vermeintliche Einschränkung auf Probleme in primaler Normalform ist in Wirklichkeit keine Einschränkung, da allgemeine LPs stets äquivalent auf Probleme in primaler Normalform transformiert werden können. Wir verdeutlichen dies an einem Beispiel.

**Beispiel 4.0.7 (Transformation auf primale Normalform)**

Betrachte das folgende lineare Optimierungsproblem:

Maximiere

$$-x_1 - 2x_2$$

unter den Nebenbedingungen

$$\begin{aligned} x_1 &\geq -1, \\ x_1 - 2x_2 &\leq 0, \\ x_1 + x_2 &= 1, \\ x_2 &\geq 0. \end{aligned}$$

Das Problem liegt nicht in primaler Normalform vor und soll auf letztere transformiert werden. Die Variable  $x_1$  ist eine **freie Variable**. Sie wird ersetzt durch

$$x_1 = x_1^+ - x_1^- \quad \text{mit} \quad x_1^+, x_1^- \geq 0.$$

Für die Ungleichungsnebenbedingungen führen wir Schlupfvariable  $s_1, s_2 \geq 0$  ein mit

$$\begin{aligned} -(x_1^+ - x_1^-) + s_1 &= 1, \\ x_1^+ - x_1^- - 2x_2 + s_2 &= 0. \end{aligned}$$

Schließlich erhalten wir ein Minimierungsproblem in primaler Normalform durch Multi-

plikation der Zielfunktion mit  $-1$ :

$$\text{Minimiere } (1, -1, 2, 0, 0) \begin{pmatrix} x_1^+ \\ x_1^- \\ x_2 \\ s_1 \\ s_2 \end{pmatrix}$$

$$\text{u.d.N. } \begin{pmatrix} -1 & 1 & 0 & 1 & 0 \\ 1 & -1 & -2 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^+ \\ x_1^- \\ x_2 \\ s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} x_1^+ \\ x_1^- \\ x_2 \\ s_1 \\ s_2 \end{pmatrix} \geq 0.$$

■

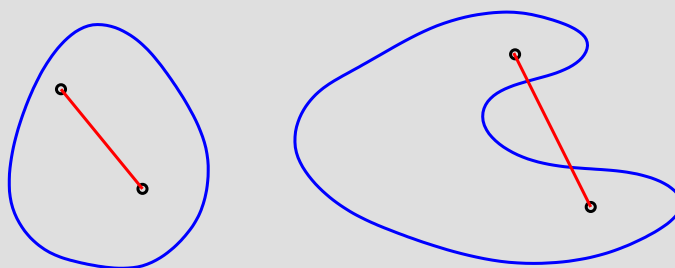
Lineare Optimierungsprobleme gehören zur Klasse der konvexen Optimierungsprobleme, die sich dadurch auszeichnet, dass die zu minimierende Zielfunktion und der zulässige Bereich konvex sind.

### Definition 4.0.8

(a) Eine Menge  $M \subseteq \mathbb{R}^n$  heißt **konvex**, falls mit zwei Punkten stets auch deren gesamte Verbindungsstrecke zu  $M$  gehört, wenn also

$$x, y \in M \Rightarrow \lambda x + (1 - \lambda)y \in M \quad \text{für alle } 0 \leq \lambda \leq 1$$

gilt.



Konvexe und nicht konvexe Menge

(b) Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt **konvex auf der konvexen Menge  $M$** , falls

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{für alle } x, y \in M, 0 \leq \lambda \leq 1$$

gilt.

Für die primale Normalform weist man leicht nach (Übung!), dass die Zielfunktion  $f(x) = c^\top x$  und der zulässige Bereich  $M = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  konvex sind. Für konvexe Optimierungsprobleme (und damit auch für lineare Programme) gilt der folgende Satz.

**Satz 4.0.9**

Seien  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $M \subseteq \mathbb{R}^n$  konvex und  $\hat{x} \in M$  lokales Minimum von  $f$  auf  $M$ . Dann ist  $\hat{x}$  auch ein globales Minimum von  $f$  auf  $M$ .

**Beweis:** Nehmen wir an,  $\hat{x}$  sei ein lokales, aber kein globales Minimum. Dann gibt es ein  $y \in M$  mit  $f(y) < f(\hat{x})$ . Für  $\lambda \in [0, 1)$  folgt

$$f(\lambda\hat{x} + (1 - \lambda)y) \leq \lambda f(\hat{x}) + (1 - \lambda)f(y) < \lambda f(\hat{x}) + (1 - \lambda)f(\hat{x}) = f(\hat{x}).$$

Damit gibt es in jeder Umgebung von  $\hat{x}$  Punkte mit kleinerem Zielfunktionswert. Widerspruch! ■

## 4.1 Ecken und zulässige Basislösungen

Wir kommen nochmal auf Beispiel 2.2.1 zurück.

**Beispiel 4.1.1** (vgl. Beispiel 2.2.1)

Der Landwirt aus Beispiel 2.2.1 muss das folgende lineare Optimierungsproblem lösen:  
Minimiere

$$f(x_1, x_2) = -100x_1 - 250x_2$$

unter den Nebenbedingungen

$$\begin{aligned} x_1 + x_2 &\leq 40, \\ 40x_1 + 120x_2 &\leq 2400, \\ 6x_1 + 12x_2 &\leq 312, \\ x_1, x_2 &\geq 0. \end{aligned}$$

Zur Untersuchung des zulässigen Bereichs definieren wir die linearen Funktionen

$$\begin{aligned} g_1(x_1, x_2) &:= x_1 + x_2, \\ g_2(x_1, x_2) &:= 40x_1 + 120x_2, \\ g_3(x_1, x_2) &:= 6x_1 + 12x_2. \end{aligned}$$

Zunächst wird die erste Restriktion  $g_1(x_1, x_2) = x_1 + x_2 \leq 40$  untersucht, vgl. Abbildung 4.1.

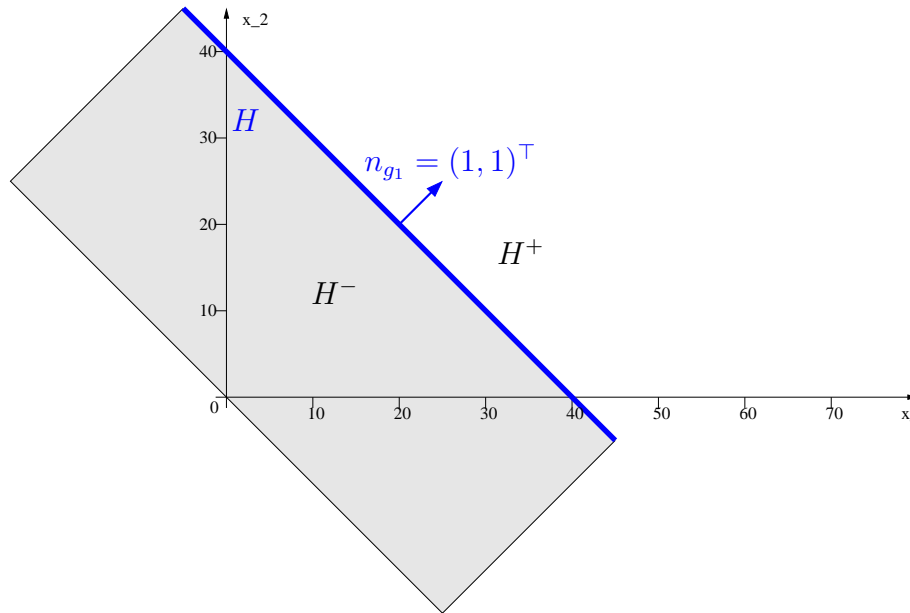


Abbildung 4.1: Geometrie der Restriktion  $g_1(x_1, x_2) = x_1 + x_2 \leq 40$ : Normalenvektor  $n_{g_1} = (1, 1)^\top$  (Länge ist skaliert), Hyperebene  $H$  und Halbräume  $H^+$  und  $H^-$ . Alle Punkte in  $H^-$  sind zulässig bzgl. dieser Restriktion.

Die Gleichung  $g_1(x_1, x_2) = x_1 + x_2 = 40$  definiert eine Gerade im  $\mathbb{R}^2$  – eine sogenannte **Hyperebene**

$$H := \{(x_1, x_2)^\top \in \mathbb{R}^2 \mid x_1 + x_2 = 40\}.$$

Der Koeffizientenvektor  $n_{g_1} = (1, 1)^\top$  ist der **Normalenvektor** der Hyperebene  $H$  und zeigt in Richtung steigender Werte der Funktion  $g_1$ . Es ist gleichzeitig auch der Gradient von  $g_1$ . Wird also ein Punkt in  $H$  in Richtung  $n_{g_1}$  verschoben, so führt dies auf einen Punkt  $(x_1, x_2)^\top$  mit  $g_1(x_1, x_2) > 40$ . Wird ein Punkt in  $H$  in die andere Richtung, also  $-n_{g_1}$ , verschoben, so führt dies auf einen Punkt  $(x_1, x_2)^\top$  mit  $g_1(x_1, x_2) < 40$ . Die Hyperebene  $H$  trennt also den  $\mathbb{R}^2$  in die zwei **Halbräume**

$$H^- := \{(x_1, x_2)^\top \in \mathbb{R}^2 \mid x_1 + x_2 \leq 40\},$$

$$H^+ := \{(x_1, x_2)^\top \in \mathbb{R}^2 \mid x_1 + x_2 > 40\}.$$

Offenbar sind die Punkte in  $H^+$  nicht zulässig, während diejenigen in  $H^-$  die Restriktion  $x_1 + x_2 \leq 40$  erfüllen.

Entsprechend können die weiteren Restriktionen inklusive  $x_1 \geq 0$  und  $x_2 \geq 0$  untersucht werden. Der zulässige Bereich ist in Abbildung 4.2 skizziert und ergibt sich als Durchschnitt der Halbräume  $H^-$  für die jeweiligen Restriktionen.

Bisher wurde die Zielfunktion nicht betrachtet. Die gestrichelte Gerade in Abbildung 4.3 entspricht der Geraden

$$f(x_1, x_2) = -100x_1 - 250x_2 = -3500,$$

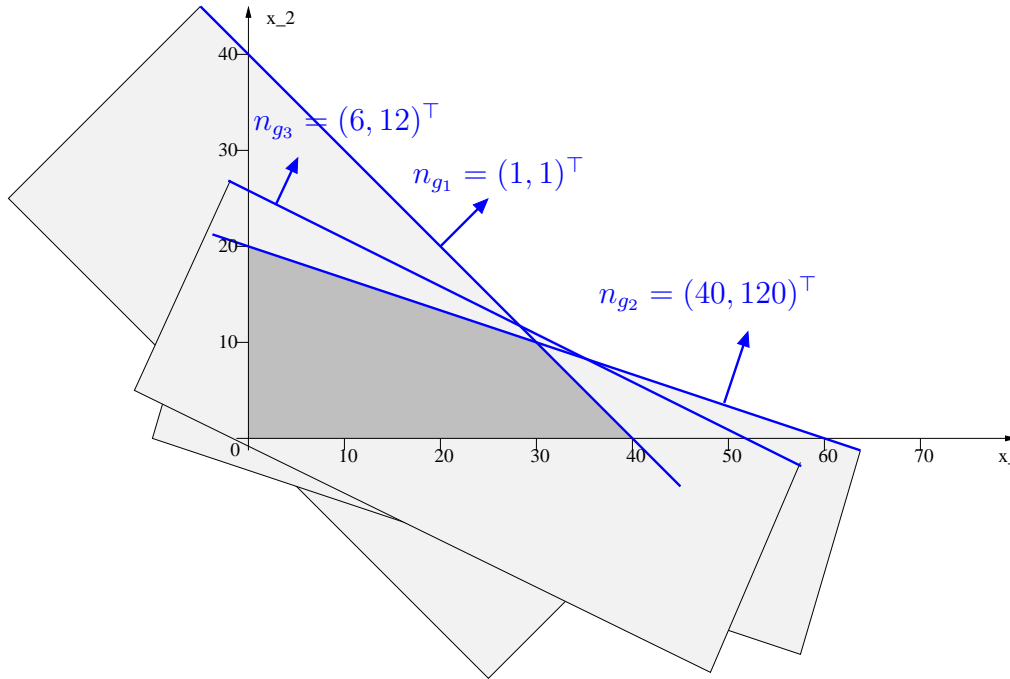


Abbildung 4.2: Zulässiger Bereich des linearen Optimierungsproblems: Durchschnitt von negativen Halbräumen für die durch  $g_1, g_2, g_3, x_1 \geq 0$  und  $x_2 \geq 0$  definierten Restriktionen (Längen der Normalenvektoren sind skaliert).

*d.h. alle Punkte auf dieser Geraden besitzen den Zielfunktionswert  $-3500$ . Die durchgezeichnete Gerade entspricht der Geraden*

$$f(x_1, x_2) = -100x_1 - 250x_2 = -5500,$$

*d.h. alle Punkte auf dieser Geraden besitzen den Zielfunktionswert  $-5500$ . Offenbar steigen die Funktionswerte von  $f$  in Richtung des Normalenvektors bzw. Gradienten  $n_f(x_1, x_2) = (-100, -250)^T$ . Da  $f$  minimiert werden soll, muss die Gerade in Richtung  $-n_f$  verschoben werden solange sie noch zulässige Punkte enthält. Die zulässigen Punkte auf der extremsten unter diesen Geraden sind dann optimal. In diesem Beispiel ergibt sich die optimale Lösung  $x_1 = 30$  und  $x_2 = 10$  mit Zielfunktionswert  $f(x_1, x_2) = -5500$ . Die optimale Lösung wird in einer **Ecke** des zulässigen Bereichs angenommen. In diesem Beispiel sind die Ecken des zulässigen Bereichs offenbar durch die Punkte  $(0, 0)$ ,  $(0, 20)$ ,  $(30, 10)$  und  $(40, 0)$  gegeben.*

■

Allgemein ist der zulässige Bereich eines linearen Optimierungsproblems mit Restriktionen  $Ax \leq b$  und  $x \geq 0$  gegeben durch den Durchschnitt der zugehörigen negativen Halbräume. Der Durchschnitt von endlich vielen abgeschlossenen Halbräumen ist eine **polyedrische Menge (Polyeder)**. Ein beschränktes Polyeder heißt **Polytop**. Abbildung 4.4 skizziert

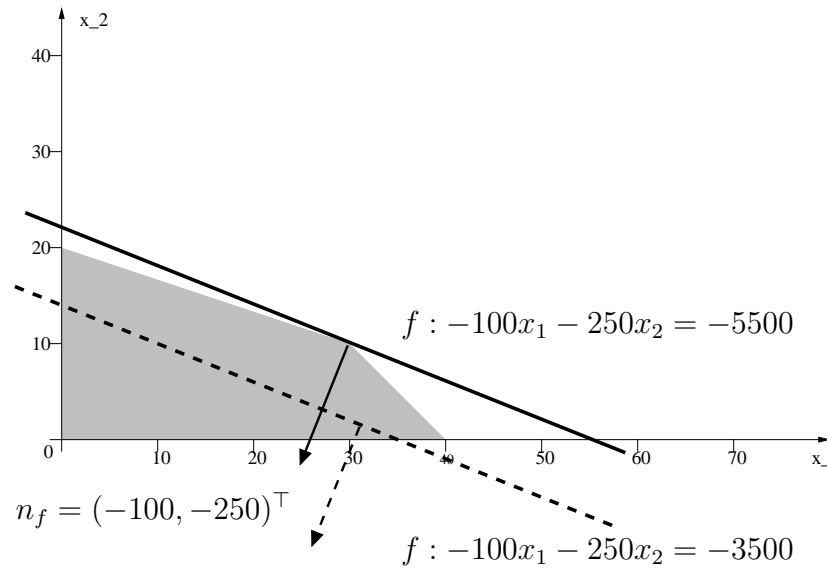


Abbildung 4.3: Grafische Lösung: Bewege Zielfunktionsgerade in Richtung  $-n_f$  solange der Durchschnitt mit dem zulässigen Bereich nicht leer ist. Die optimale Lösung ist  $(x_1, x_2)^T = (30, 10)^T$ .

verschiedene Konstellationen, die in der linearen Optimierung auftreten können.

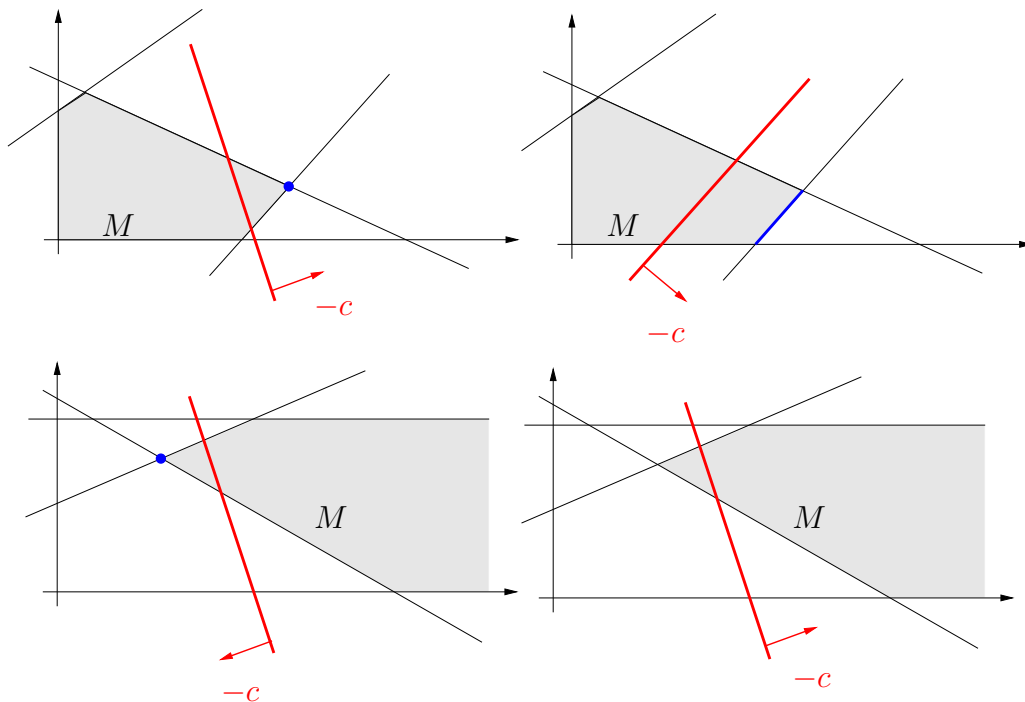


Abbildung 4.4: Verschiedene Geometrien des zulässigen Bereichs  $M$  (grau), Höhenlinie der Zielfunktion (Gerade), eindeutige optimale Lösung (links), unendlich viele optimale Lösungen (rechts oben) und keine optimale Lösung (rechts unten).

Die Beispiele erlauben folgende Beobachtungen:

- Der zulässige Bereich  $M$  eines linearen Optimierungsproblems kann **beschränkt** oder **unbeschränkt** sein.
- Falls der zulässige Bereich beschränkt ist, dann existiert stets eine optimale Lösung nach dem Satz von Weierstrass, welcher besagt, dass eine stetige Funktion auf einer kompakten Menge ihr Minimum und Maximum annimmt. Falls der zulässige Bereich unbeschränkt ist, ist die Existenz einer optimalen Lösung nicht gesichert.
- Es kann genau eine optimale Lösung oder unendlich viele optimale Lösungen geben.
- Existiert eine optimale Lösung, so befindet sich stets ein Eckpunkt unter den optimalen Lösungen.

Die letzte Vermutung (sie gilt allgemein) ist die Hauptmotivation für das berühmte Simplexverfahren. Das Simplexverfahren ist eines der populärsten Verfahren zur Lösung linearer Optimierungsprobleme und wurde 1947 von G. B. Dantzig entwickelt. Die wesentliche Idee des Simplexverfahrens ist, von einem Eckpunkt des zulässigen Bereichs zu einem benachbarten Eckpunkt des zulässigen Bereichs zu gelangen, dabei den Zielfunktionswert (möglichst) zu verringern und dies solange zu wiederholen, bis ein optimaler Eckpunkt erreicht wird. Satz 4.1.2 rechtfertigt diese Idee. Er besagt, dass es ausreicht, nur die Eckpunkte des zulässigen Bereichs eines linearen Optimierungsproblems zu untersuchen, um **eine** optimale Lösung (nicht alle optimale Lösungen!) zu bestimmen.

### Satz 4.1.2 (Fundamentalsatz der linearen Programmierung)

Sei das lineare Optimierungsproblem in primaler Normalform

$$\text{Minimiere } c^\top x \quad \text{u.d.N.} \quad Ax = b, \quad x \geq 0$$

gegeben mit  $c, x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$ ,  $A \in \mathbb{R}^{m \times n}$ . Der zulässige Bereich  $M = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  sei nichtleer. Dann gelten folgende Aussagen:

- (a) Entweder ist die Zielfunktion auf  $M$  nach unten unbeschränkt, oder es existiert eine optimale Lösung und mindestens ein Eckpunkt von  $M$  ist unter den optimalen Lösungen.
- (b) Ist  $M$  beschränkt, dann existiert eine optimale Lösung und  $x \in M$  ist optimal genau dann, wenn  $x$  eine Konvexkombination von optimalen Eckpunkten ist. Hierin heißt  $x$  Konvexkombination von Punkten  $x^1, \dots, x^N$ , falls  $x$  sich darstellen lässt als

$$x = \sum_{i=1}^N \lambda_i x^i \quad \text{mit} \quad \sum_{i=1}^N \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = 1, \dots, N.$$



Zur Konstruktion des Simplexverfahrens müssen allerdings folgende Fragen beantwortet werden:

- Wie können (zulässige) Eckpunkte berechnet werden?
- Wie kann zu einem gegebenen Eckpunkt ein benachbarter Eckpunkt berechnet werden?
- Wie kann festgestellt werden, ob ein Eckpunkt bereits optimal ist?

Zur Beantwortung dieser Fragen muss zunächst der anschauliche Begriff eines Eckpunkts mathematisch präzisiert werden. Für konvexe Mengen eignet sich die folgende geometrische Definition, die besagt, dass ein Eckpunkt einer Menge  $M$  nicht als echte Konvexkombination von zwei Punkten aus  $M$  dargestellt werden kann. Die Einschränkung auf konvexe Mengen ist für lineare Optimierungsprobleme ausreichend, da man leicht zeigt, dass der zulässige Bereich eines linearen Optimierungsproblems stets konvex ist.

#### Definition 4.1.3 (Ecke, Eckpunkt)

Sei  $M$  eine konvexe Menge.  $x \in M$  heißt **Ecke (oder Eckpunkt) von  $M$** , falls aus der Darstellung  $x = \lambda x_1 + (1 - \lambda)x_2$  mit  $0 < \lambda < 1$  und  $x_1, x_2 \in M$  stets folgt, dass  $x_1 = x_2 = x$  gilt.

Die Menge  $\{\lambda x_1 + (1 - \lambda)x_2 \mid 0 < \lambda < 1\}$  ist die Verbindungsstrecke zwischen den Punkten  $x_1$  und  $x_2$ , wobei  $x_1$  und  $x_2$  nicht enthalten sind. Ist  $x$  also eine Ecke, dann kann  $x$  nicht auf einer solchen Strecke liegen, es sei denn, es gilt  $x = x_1 = x_2$ . Leider ist die Definition 4.1.3 nicht konstruktiv und erlaubt es nicht, einen Eckpunkt des zulässigen Bereichs eines linearen Optimierungsproblems tatsächlich zu berechnen. Aus diesem Grund wird eine äquivalente Charakterisierung von Eckpunkten untersucht.

#### Definition 4.1.4 (zulässige Basislösung)

Seien  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  und  $M = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  gegeben. Mit  $a^j$ ,  $j = 1, \dots, n$ , seien die Spalten von  $A$  bezeichnet.

$x \in M$  heißt **zulässige Basislösung von  $M$** , falls die Spaltenvektoren  $\{a^j \mid j \in B_+\}$  mit  $B_+ = \{j \in \{1, \dots, n\} \mid x_j > 0\}$  linear unabhängig sind.

Den Zusammenhang zwischen einer zulässigen Basislösung und einer Ecke stellt der folgende Satz her:

**Satz 4.1.5**

*$x \in M$  ist eine Ecke von  $M$  genau dann, wenn  $x$  eine zulässige Basislösung von  $M$  ist.*

**Wichtige Notation:**

Sei  $B \subseteq \{1, \dots, n\}$  eine beliebige Indexmenge.

- Sei  $x = (x_1, \dots, x_n)^\top$  ein Vektor. Dann bezeichnet  $x_B$  den Vektor mit den Komponenten  $x_j$ ,  $j \in B$ .
- Sei  $A$  eine  $m \times n$ -Matrix mit Spalten  $a^j$ ,  $j = 1, \dots, n$ . Dann bezeichnet  $A_B$  die Matrix mit den Spalten  $a^j$ ,  $j \in B$ .

Konkretes Beispiel:

$$x = \begin{pmatrix} -1 \\ 10 \\ 3 \\ -4 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}, \quad B = \{2, 4\} \Rightarrow x_B = \begin{pmatrix} 10 \\ -4 \end{pmatrix}, \quad A_B = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}.$$

Die Indexmenge  $B_+$  besteht aus höchstens  $m$  Indizes, evtl. aber aus weniger als  $m$  Indizes. Da der Rang von  $A$  gleich  $m$  ist, lassen sich die Spalten  $(a^j)_{j \in B_+}$  aber stets zu einem System von genau  $m$  linear unabhängigen Spalten  $(a^j)_{j \in B}$  von  $A$  mit Indexmenge  $B_+ \subseteq B$  erweitern.

Fassen wir diese Spalten zu einer regulären Matrix  $A_B$  und die verbleibenden Spalten zu einer Matrix  $A_N$  mit  $N = \{1, \dots, n\} \setminus B$  zusammen, so gilt für beliebiges  $x$  mit  $Ax = b$  dann

$$b = Ax = A_B x_B + A_N x_N \quad \Rightarrow \quad x_B = A_B^{-1} b - A_B^{-1} A_N x_N.$$

Eine zulässige Basislösung  $x$  erfüllt wegen  $x_N = 0$  speziell

$$x_B = A_B^{-1} b \geq 0, \quad x_N = 0.$$

Damit kann man Eckpunkte (bzw. zulässige Basislösungen) durch Lösen von linearen Gleichungssystemen berechnen.

**Algorithmus 4.1.6** (Berechnung eines Eckpunkts)

(1) Wähle  $m$  linear unabhängige Spalten  $a^j$ ,  $j \in B$ , mit  $B \subseteq \{1, \dots, n\}$  von  $A$  und setze  $N = \{1, \dots, n\} \setminus B$ .

(2) Setze  $x_N = 0$  und löse das lineare Gleichungssystem

$$A_B x_B = b.$$

(3) Falls  $x_B \geq 0$  gilt, ist  $x$  zulässige Basislösung: STOP. Falls es einen Index  $j \in B$  mit  $x_j < 0$  gibt, dann ist  $x$  nicht zulässig. Gehe zu (1) und wiederhole die Prozedur mit einer anderen Wahl von linear unabhängigen Spalten von  $A$ .

Wegen  $\text{Rang}(A) = m$  gibt es  $m$  linear unabhängige Spalten von  $A$  in Schritt (1). Ein naiver Ansatz zur Lösung des linearen Optimierungsproblems besteht in der Berechnung aller zulässigen Basislösungen. Hiervon gibt es höchstens  $\binom{n}{m}$ , was gerade die Anzahl der Möglichkeiten ist,  $m$  Spalten aus insgesamt  $n$  Spalten auszuwählen. Diese Vorgehensweise ist nicht effizient. Zudem kann durch Betrachtung der endlich vielen Ecken des zulässigen Bereichs nicht erkannt werden, ob die Zielfunktion auf  $M$  nach unten unbeschränkt ist und das Problem keine optimale Lösung besitzt.

**Beispiel 4.1.7**

Betrachte  $M = \{x \in \mathbb{R}^4 \mid Ax = b, x \geq 0\}$  mit

$$A = \begin{pmatrix} 2 & 3 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 6 \\ 2 \end{pmatrix}.$$

Es gibt  $\binom{4}{2} = 6$  mögliche Kombinationen von zwei Spalten von  $A$ :

$$\begin{aligned} B_1 = \{1, 2\} &\Rightarrow A_{B_1}^{-1}b = \begin{pmatrix} 2 \\ \frac{2}{3} \end{pmatrix}, & x^1 &= \left(2, \frac{2}{3}, 0, 0\right)^\top, \\ B_2 = \{1, 3\} &\Rightarrow A_{B_2}^{-1}b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, & x^2 &= (2, 0, 2, 0)^\top, \\ B_3 = \{1, 4\} &\Rightarrow A_{B_3}^{-1}b = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, & x^3 &= (3, 0, 0, -1)^\top, \\ B_4 = \{2, 3\} &\Rightarrow A_{B_4} \text{ ist singular}, \\ B_5 = \{2, 4\} &\Rightarrow A_{B_5}^{-1}b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, & x^5 &= (0, 2, 0, 2)^\top, \\ B_6 = \{3, 4\} &\Rightarrow A_{B_6}^{-1}b = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, & x^6 &= (0, 0, 6, 2)^\top. \end{aligned}$$

$x^3$  ist nicht zulässig. Die Eckpunkte des zulässigen Bereichs sind also gegeben durch  $x^1, x^2, x^5, x^6$ . ■

### Definition 4.1.8

- (i) Sei  $\text{Rang}(A) = m$  und  $x$  eine zulässige Basislösung. Jedes System  $\{a^j \mid j \in B\}$  von  $m$  linear unabhängigen Spalten von  $A$ , welches jene Spalten  $a^j$  mit  $x_j > 0$  enthält, heißt **Basis von  $x$** .
- (ii) Sei  $\{a^j \mid j \in B\}$  eine Basis von  $x$ . Die Indexmenge  $B$  heißt **Basisindexmenge**, die Indexmenge  $N := \{1, \dots, n\} \setminus B$  heißt **Nichtbasisindexmenge**, die Matrix  $A_B := (a^j)_{j \in B}$  heißt **Basismatrix**, die Matrix  $A_N := (a^j)_{j \in N}$  heißt **Nichtbasismatrix**, der Vektor  $x_B := (x_j)_{j \in B}$  heißt **Basisvariable** und der Vektor  $x_N := (x_j)_{j \in N}$  heißt **Nichtbasisvariable**.

## 4.2 Das primale Simplexverfahren

Wir haben in Satz 4.1.5 gesehen, dass ein Eckpunkt algebraisch über linear unabhängige Spalten von  $A$  charakterisiert werden kann. Laut Satz 4.1.2 genügt es, die Eckpunkte des zulässigen Bereichs zu berechnen, um mindestens eine optimale Lösung – falls eine solche existiert – zu erhalten. Darauf basiert das **Simplexverfahren**.

Der schematische Ablauf des Simplexverfahrens lautet wie folgt:

**(0) Phase 0:**

Transformiere das vorliegende lineare Optimierungsproblem auf primale Normalform.

**(1) Phase 1:**

Bestimme eine zulässige Basislösung (Eckpunkt)  $x$  mit Basisindexmenge  $B$ , Nichtbasisindexmenge  $N$ , Basisvariable  $x_B \geq 0$  und Nichtbasisvariable  $x_N = 0$ .

**(2) Phase 2:**

Bestimme solange benachbarte zulässige Basislösungen (Eckpunkte)  $x^+$  mit Basisindexmenge  $B^+$  und Nichtbasisindexmenge  $N^+$  bis entweder ein Optimum erreicht ist, oder bis entschieden werden kann, dass das Problem keine Lösung besitzt.

Im folgenden widmen wir uns

- der effizienten Berechnung (zulässiger) Eckpunkte,
- der Berechnung benachbarter Eckpunkte ,
- Optimalitätskriterien für Eckpunkte.

**4.2.1 Basiswechsel beim Simplexverfahren**

In diesem Abschnitt wird auf die Konstruktion einer benachbarten zulässigen Basislösung  $x^+$  in Phase 2 des Simplexverfahrens eingegangen, wenn eine zulässige Basislösung  $x$  gegeben ist. Ausgehend von Basis- bzw. Nichtbasisindexmengen  $B$  und  $N$  in Iteration  $j$  des Simplexverfahrens werden neue Indexmengen

$$\begin{aligned} B^+ &= (B \setminus \{p\}) \cup \{q\}, \\ N^+ &= (N \setminus \{q\}) \cup \{p\} \end{aligned}$$

konstruiert, wobei  $p \in B$  und  $q \in N$  geeignet gewählte Indizes sind. Da hier lediglich der Basisindex  $p$  und der Nichtbasisindex  $q$  gegeneinander ausgetauscht werden, spricht man auch von einem **Basiswechsel**. Die Indizes  $p$  und  $q$  werden dabei nicht beliebig gewählt, sondern so, dass die folgenden beiden Mindestanforderungen erfüllt sind:

**(i) Zulässigkeit:**

Mit  $x \in M$  muß auch  $x^+ \in M$  gelten, d.h

$$Ax = b, x \geq 0 \quad \Rightarrow \quad Ax^+ = b, x^+ \geq 0$$

(ii) **Abstiegseigenschaft:**

Der Zielfunktionswert fällt monoton, d.h.

$$c^\top x^+ \leq c^\top x.$$

**Voraussetzung:**

- Wir setzen im folgenden stets voraus, dass  $\text{Rang}(A) = m$  gilt. Dies ist keine wesentliche Einschränkung, sondern schließt lediglich überflüssige Nebenbedingungen aus.

Es sei eine Basis mit Basisindexmenge  $B$  gegeben, mithin ist  $A_B$  invertierbar. Dann kann die Nebenbedingung  $Ax = b$  nach den Basisvariablen  $x_B$  aufgelöst werden:

$$Ax = A_B x_B + A_N x_N = b \quad \Rightarrow \quad x_B = \underbrace{A_B^{-1} b}_{=\beta_B} - \underbrace{A_B^{-1} A_N}_{=\Gamma_B^N} x_N. \quad (4.2)$$

Einsetzen in die Zielfunktion liefert die Darstellung

$$c^\top x = c_B^\top x_B + c_N^\top x_N = \underbrace{c_B^\top \beta_B}_{=d} - \underbrace{(c_B^\top \Gamma_B^N - c_N^\top)}_{=\zeta_N^\top} x_N =: d - \zeta_N^\top x_N. \quad (4.3)$$

Hierin besitzt der Vektor  $\beta_B$  die Komponenten  $\beta_i, i \in B$ , der Vektor  $\zeta_N$  die Komponenten  $\zeta_i, i \in N$ , und die Matrix  $\Gamma_B^N$  die Komponenten  $\gamma_{ij}, i \in B, j \in N$ .

Für eine zulässige Basislösung  $x$  gilt  $x_N = 0$  in (4.2) und (4.3) und somit

$$x_B = \beta_B \geq 0, \quad x_N = 0, \quad c^\top x = d. \quad (4.4)$$

Wir wollen nun die Basis abändern, um zu einer benachbarten Basis zu gelangen. Dazu wird die Kantenhalbgerade

$$z(t) = x + ts, \quad t \geq 0, \quad (4.5)$$

mit noch zu bestimmender Suchrichtung  $s$  und Schrittweite  $t \geq 0$  betrachtet, vgl. Abbildung 4.5.

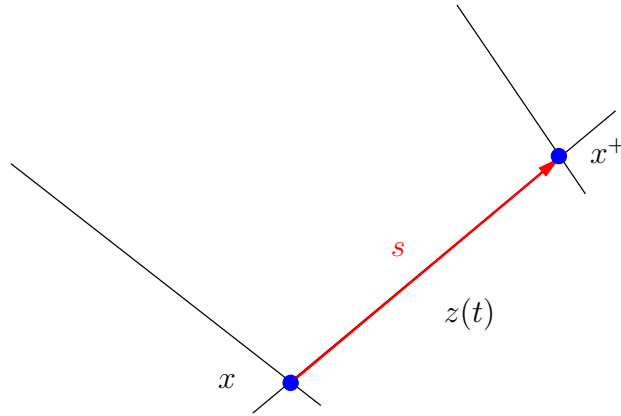


Abbildung 4.5: Idee des Basiswechsels beim Simplexverfahren: Es wird entlang einer Kante gesucht, so dass die Zielfunktion monoton fällt.

Die Suchrichtung  $s$  wird speziell so gewählt, dass lediglich eine Nichtbasisvariable  $x_q$  mit geeignetem  $q \in N$  abgeändert wird, während die übrigen Nichtbasisvariablen mit  $x_j = 0$ ,  $j \in N$ ,  $j \neq q$ , unverändert bleiben. Für einen geeigneten Index  $q \in N$  gelte also

$$s_q := 1, \quad s_j := 0 \text{ für } j \in N, j \neq q, \quad (4.6)$$

und somit

$$z_q(t) = t \geq 0, \quad z_j(t) = 0 \text{ für } j \in N, j \neq q. \quad (4.7)$$

Natürlich sollen die Werte  $z(t)$  darüber hinaus auch zulässig sein, d.h. wie in (4.2) muß gelten

$$b = Az(t) = \underbrace{Ax}_{=b} + tAs \Rightarrow 0 = As = A_B s_B + A_N s_N.$$

Auflösen ergibt

$$s_B = -\Gamma_B^N s_N, \quad (4.8)$$

so dass die Suchrichtung  $s$  mit (4.6) und (4.8) vollständig festgelegt ist. Die Zielfunktionswerte berechnen sich zu

$$\begin{aligned} c^\top z(t) &\stackrel{(4.5)}{=} c^\top x + t c^\top s \\ &\stackrel{(4.4)}{=} d + t c_B^\top s_B + t c_N^\top s_N \\ &\stackrel{(4.8)}{=} d - t (c_B^\top \Gamma_B^N - c_N^\top) s_N \\ &= d - t \zeta_N^\top s_N \\ &\stackrel{(4.6)}{=} d - t \zeta_q. \end{aligned} \quad (4.9)$$

Aus dieser Darstellung läßt sich sofort ablesen, dass der Zielfunktionswert entlang der Richtung  $s$  für  $t \geq 0$  abnimmt, wenn  $\zeta_q > 0$  gilt.

Gilt andererseits  $\zeta_j \leq 0$  für alle  $j \in N$ , so ist entlang  $s$  kein Abstieg in der Zielfunktion möglich. Es bleibt zu klären, ob es dann möglicherweise andere Punkte gibt, die nicht auf der Kantenhalbgeraden  $z$  liegen, und einen kleineren Zielfunktionswert annehmen. Dies ist nicht der Fall, denn für beliebiges  $\hat{x} \in M$  gilt  $\hat{x} \geq 0$  und  $A_B \hat{x}_B + A_N \hat{x}_N = b$  bzw.  $\hat{x}_B = \beta_B - \Gamma_B^N \hat{x}_N$ . Aus  $\zeta_j \leq 0$  und  $\hat{x}_N \geq 0$  folgt

$$c^\top \hat{x} = c_B^\top \hat{x}_B + c_N^\top \hat{x}_N = c_B^\top \beta_B - (c_B^\top \Gamma_B^N - c_N^\top) \hat{x}_N = d - \zeta_N^\top \hat{x}_N \geq d = c^\top x.$$

Damit ist die aktuelle Basislösung  $x$  bereits optimal. Wir fassen zusammen:

### Zwischenresultat zur Wahl der Pivotspalte $q$ :

Um die Abstiegseigenschaft in (ii) zu erreichen, muß ein Index  $q \in N$  mit  $\zeta_q > 0$  gewählt werden. Gilt  $\zeta_j \leq 0$  für alle  $j \in N$ , so ist die aktuelle Basislösung  $x$  optimal.

Nun wird versucht, die Zulässigkeitsbedingung in (i) zu erfüllen. Es muß gelten

$$z_B(t) = x_B + t s_B \stackrel{(4.8)}{=} \beta_B - t \Gamma_B^N s_N \geq 0$$

bzw. in Komponentenschreibweise

$$\begin{aligned} z_i(t) &= \beta_i - t \sum_{j \in N} \gamma_{ij} s_j \\ &= \beta_i - t \underbrace{\gamma_{iq}}_{=1} \underbrace{s_q}_{=1} - t \sum_{j \in N, j \neq q} \underbrace{\gamma_{ij}}_{=0} \underbrace{s_j}_{=0} \\ &\stackrel{(4.6)}{=} \beta_i - \gamma_{iq} t \geq 0, \quad i \in B. \end{aligned}$$

Die Bedingungen  $\beta_i - \gamma_{iq} t \geq 0$ ,  $i \in B$ , schränken die Schrittweite  $t \geq 0$  ein. Wir diskutieren zwei Fälle:

- (a) **Fall 1:** Es gelte  $\gamma_{iq} \leq 0$  für alle  $i \in B$ .

Wegen  $\beta_i \geq 0$  gilt dann  $z_i(t) = \beta_i - \gamma_{iq} t \geq 0$  für alle  $t \geq 0$  und alle  $i \in B$ . Mit anderen Worten:  $z(t)$  bleibt für alle  $t \geq 0$  zulässig.

Gilt nun zusätzlich  $\zeta_q > 0$ , so ist die Zielfunktion wegen (4.9) für  $t \rightarrow \infty$  nicht nach unten beschränkt. Das lineare Optimierungsproblem besitzt keine Lösung!

### Zwischenresultat: Unlösbarkeit des linearen Optimierungsproblems

Gilt für ein  $q \in N$   $\zeta_q > 0$  und  $\gamma_{iq} \leq 0$  für alle  $i \in B$ , so besitzt das lineare Optimierungsproblem keine Lösung. Die Zielfunktion ist nach unten unbeschränkt.

- (b) **Fall 2:** Es gelte  $\gamma_{iq} > 0$  für mindestens ein  $i \in B$ .

Aus  $\beta_i - \gamma_{iq} t \geq 0$  folgt dann  $t \leq \frac{\beta_i}{\gamma_{iq}}$ . Diese Forderung schränkt die Schrittweite  $t$  ein.



**Zwischenresultat zur Wahl der Pivotzeile  $p$ :**

Die Zulässigkeit in (i) wird durch die Wahl eines Index  $p \in B$  mit

$$t_{min} := \frac{\beta_p}{\gamma_{pq}} := \min \left\{ \frac{\beta_i}{\gamma_{iq}} \mid \gamma_{iq} > 0, i \in B \right\}$$

erreicht.

**4.2.2 Der Algorithmus**

Mit diesen Betrachtungen ist das Simplexverfahren definiert, wobei man sich noch überlegen muss, dass die neue Basislösung tatsächlich eine Basislösung ist. Dies kann mit dem Austauschatz von Steinitz erfolgen, worauf wir nicht weiter eingehen möchten. Wir fassen alles in einem Algorithmus zusammen:

**Algorithmus 4.2.1 (Primales Simplexverfahren)****(0) Phase 1:**

Bestimme eine zulässige Basislösung (Eckpunkt)  $x$  mit Basisindexmenge  $B$ , Nichtbasisindexmenge  $N$ , Basismatrix  $A_B$ , Nichtbasismatrix  $A_N$ , Basisvariable  $x_B \geq 0$  und Nichtbasisvariable  $x_N = 0$  der primalen Normalform 4.0.5 mit  $\text{Rang}(A) = m$ .

(1) Berechne  $\Gamma_B^N := A_B^{-1}A_N$ ,  $\beta_B := A_B^{-1}b$  und die negativen reduzierten Kosten  $\zeta_N^\top := c_B^\top \Gamma_B^N - c_N^\top$ .

**(2) Überprüfung der Optimalität:**

Falls  $\zeta_j \leq 0$  für alle  $j \in N$  gilt, so beende das Verfahren. Die aktuelle Basislösung  $x_B = \beta_B$ ,  $x_N = 0$  ist optimal. Der Zielfunktionswert beträgt  $d = c_B^\top \beta_B$ .

**(3) Überprüfung der Unbeschränktheit:**

Gibt es einen Index  $q$  mit  $\zeta_q > 0$  und  $\gamma_{iq} \leq 0$  für alle  $i \in B$ , so besitzt das lineare Optimierungsproblem keine Lösung. STOP.

**(4) Bestimmung des Pivotelements:**

Wähle einen Index  $q$  mit  $\zeta_q > 0$ .  $q$  legt die Pivotspalte fest. Wähle einen Index  $p$  mit

$$\frac{\beta_p}{\gamma_{pq}} = \min \left\{ \frac{\beta_i}{\gamma_{iq}} \mid \gamma_{iq} > 0, i \in B \right\}.$$

$p$  legt die Pivotzeile fest.

(5) **Führe Basiswechsel durch:**

Setze  $B := (B \setminus \{p\}) \cup \{q\}$  und  $N := (N \setminus \{q\}) \cup \{p\}$ .

(6) Gehe zu (1).

Das Simplexverfahren wird häufig in einer kompakten Darstellung – dem sogenannten **Simplextableau** – durchgeführt. Die Beziehungen

$$\begin{aligned} x_B &= (A_B)^{-1}b - \Gamma_B^N x_N, \\ c^\top x &= c_B^\top x_B - (c_B^\top \Gamma_B^N - c_N^\top) x_N, \end{aligned}$$

werden in einem Tableau nach folgendem Schema zusammengefasst:

	$x_N$	
$x_B$	$\Gamma_B^N = (A_B)^{-1}A_N$	$\beta_B = (A_B)^{-1}b$
	$\zeta_N^\top = c_B^\top \Gamma_B^N - c_N^\top$	$c_B^\top x_B$

Da  $x_N = 0$  in einer Basislösung gilt, kann der aktuelle Wert der Basisvariablen  $x_B$  direkt im Tableau abgelesen werden:  $x_B = (A_B)^{-1}b$ . Der zugehörige Zielfunktionswert lautet  $c^\top p = c_B^\top x_B = c_B^\top (A_B)^{-1}b$ .

Hat man sich für eine Pivotspalte  $q \in N$  entschieden, so werden die Quotienten  $\frac{x_B}{\Gamma_B^q}$  benötigt, welche aus Gründen der Übersichtlichkeit in einer weiteren Spalte des Tableaus aufgelistet werden können:

	$x_N$		
$x_B$	$\Gamma_B^N$	$\beta_B$	$\frac{x_B}{\Gamma_B^q}$
	$c_B^\top \Gamma_B^N - c_N^\top$	$c_B^\top x_B$	

Hierin bezeichnet  $\Gamma_B^q$  die Spalte von  $\Gamma_B^N$ , die zum Index  $q \in N$  gehört. Die Division  $x_B/\Gamma_B^q$  ist komponentenweise zu verstehen.

### Beispiel 4.2.2

Betrachte das lineare Optimierungsproblem

$$\text{Minimiere } c^\top x \quad \text{u.d.N.} \quad Ax = b, \quad x \geq 0$$

mit den Daten ( $x_3, x_4, x_5$  sind Schlupfvariable):

$$c = \begin{pmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 0 \end{pmatrix}, A = \begin{pmatrix} 1 & 4 & 1 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 24 \\ 21 \\ 9 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}.$$

Eine zulässige Basislösung ist gegeben durch die Basisindexmenge  $B = \{3, 4, 5\}$  mit  $N = \{1, 2\}$  und

$$A_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, A_N = \begin{pmatrix} 1 & 4 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}, c_B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, c_N = \begin{pmatrix} -2 \\ -5 \end{pmatrix}.$$

und

$$\Gamma_B^N = \begin{pmatrix} \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \\ \gamma_{51} & \gamma_{52} \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 24 \\ 21 \\ 9 \end{pmatrix}, \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Es ergibt sich das folgende Starttableau für das Simplexverfahren:

	$x_1$	$x_2$		
$x_3$	1	4	24	6
$x_4$	3	1	21	21
$x_5$	1	1	9	9
	2	5	0	

Das Tableau ist noch nicht optimal und das Problem ist nicht unbeschränkt. Wir wählen den Nichtbasisindex  $\boxed{q = 2 \in N}$  mit  $\zeta_q = 5 > 0$  als Pivotspalte (wir hätten auch  $q = 1$  wählen können) und berechnen die Quotienten

$$\frac{x_3}{\gamma_{32}} = \frac{24}{4} = 6, \quad \frac{x_4}{\gamma_{42}} = \frac{21}{1} = 21, \quad \frac{x_5}{\gamma_{52}} = \frac{9}{1} = 9.$$

Diese Werte stehen in der rechten Spalte des Starttableaus. Der erste Quotient ist minimal und definiert die Pivotzeile  $\boxed{p = 3 \in B}$ .

Basiswechsel liefert die Basisindexmenge  $B = \{2, 4, 5\}$ , die Nichtbasismenge  $N = \{1, 3\}$ , sowie die Matrizen

$$A_B = \begin{pmatrix} 4 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, A_N = \begin{pmatrix} 1 & 1 \\ 3 & 0 \\ 1 & 0 \end{pmatrix}, c_B = \begin{pmatrix} -5 \\ 0 \\ 0 \end{pmatrix}, c_N = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

und

$$\Gamma_B^N = \begin{pmatrix} \gamma_{21} & \gamma_{23} \\ \gamma_{41} & \gamma_{43} \\ \gamma_{51} & \gamma_{53} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{11}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \end{pmatrix}, \begin{pmatrix} x_2 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 6 \\ 15 \\ 3 \end{pmatrix}, \begin{pmatrix} \zeta_1 \\ \zeta_3 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} \\ -\frac{5}{4} \end{pmatrix}.$$

Damit folgt Tableau 1:

	$x_1$	$x_3$		
$x_2$	$\frac{1}{4}$	$\frac{1}{4}$	6	$\frac{24}{4}$
$x_4$	$\frac{11}{4}$	$-\frac{1}{4}$	15	$\frac{60}{11}$
$x_5$	$\frac{3}{4}$	$-\frac{1}{4}$	3	4
	$\frac{3}{4}$	$-\frac{5}{4}$	-30	

Das Tableau ist nicht optimal und das Problem ist nicht unbeschränkt. Wir wählen die Pivotspalte  $\boxed{q=1}$  und die Pivotzeile  $\boxed{p=5}$  und führen einen Basiswechsel durch. Dieser liefert die Basisindexmenge  $B = \{2, 4, 1\}$ , die Nichtbasismenge  $N = \{5, 3\}$ , sowie die Matrizen

$$A_B = \begin{pmatrix} 4 & 0 & 1 \\ 1 & 1 & 3 \\ 1 & 0 & 1 \end{pmatrix}, A_N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, c_B = \begin{pmatrix} -5 \\ 0 \\ -2 \end{pmatrix}, c_N = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

und

$$\Gamma_B^N = \begin{pmatrix} \gamma_{25} & \gamma_{23} \\ \gamma_{45} & \gamma_{43} \\ \gamma_{15} & \gamma_{13} \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} \\ -\frac{11}{3} & \frac{2}{3} \\ \frac{4}{3} & -\frac{1}{3} \end{pmatrix}, \begin{pmatrix} x_2 \\ x_4 \\ x_1 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} \zeta_5 \\ \zeta_3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Damit folgt Tableau 2:

	$x_5$	$x_3$		
$x_2$	$-\frac{1}{3}$	$\frac{1}{3}$	5	
$x_4$	$-\frac{11}{3}$	$\frac{2}{3}$	4	
$x_1$	$\frac{4}{3}$	$-\frac{1}{3}$	4	
	-1	-1	-33	

Tableau 2 ist optimal, da  $\zeta_N < 0$  gilt. Die zugehörige optimale Basislösung lautet  $x_2 = 5$ ,  $x_4 = 4$ ,  $x_1 = 4$ ,  $x_5 = 0$  und  $x_3 = 0$ . Der optimale Zielfunktionswert beträgt  $c^\top x = -33$ . ■

### Bemerkung 4.2.3

Anstatt die Inverse  $A_B^{-1}$  in Schritt (1) von Algorithmus 4.2.1 explizit zu berechnen, werden die linearen Gleichungssysteme  $A_B \Gamma_B^N = A_N$  und  $A_B x_B = b$  gelöst. Hierbei können effiziente numerische Verfahren (LR-Zerlegung, QR-Zerlegung, ggf. Verfahren zur Lösung von dünnbesetzten Gleichungssystemen) eingesetzt werden. Ebenso gibt es Update-Techniken

zur effizienten Berechnung der erwähnten Zerlegungen beim Basiswechsel. Man spricht hier auch vom **revidierten Simplexverfahren**. ■

### 4.2.3 Updateformeln für das Simplextableau

Die Neuberechnung des Tableauihalts bei einem Basiswechsel mit Hilfe der Inversen Basismatrix  $A_B^{-1}$  ist für Handrechnungen mitunter sehr mühsam. Es lassen sich jedoch effiziente Updateformeln herleiten. Hierzu nehmen wir an, dass ein Simplextableau zur Basisindexmenge  $B$  vorliegt. Wir wollen nun die Basis  $B$  abändern, um zu einer benachbarten Basis zu gelangen. Dazu seien nach den Regeln des Simplexverfahrens bereits die Pivotspalte  $q \in N$  und die Pivotzeile  $p \in B$  gewählt worden.

Betrachte dazu die Zeile  $p$  im Tableau:

$$x_p = \beta_p - \sum_{j \in N} \gamma_{pj} x_j = \beta_p - \gamma_{pq} x_q - \sum_{j \in N, j \neq q} \gamma_{pj} x_j.$$

Für das Pivotelement gilt  $\gamma_{pq} \neq 0$  und die Gleichung kann nach  $x_q$  aufgelöst werden:

$$\begin{aligned} x_q &= \frac{1}{\gamma_{pq}} \left( \beta_p - x_p - \sum_{j \in N, j \neq q} \gamma_{pj} x_j \right) = \underbrace{\frac{\beta_p}{\gamma_{pq}}}_{=\beta_q^+} - \underbrace{\frac{1}{\gamma_{pq}}}_{=\gamma_{qp}^+} x_p - \sum_{j \in N, j \neq q} \underbrace{\frac{\gamma_{pj}}{\gamma_{pq}}}_{=\gamma_{qj}^+} x_j \\ &= \beta_q^+ - \sum_{j \in N^+} \gamma_{qj}^+ x_j, \end{aligned}$$

wobei  $N^+ = (N \setminus \{q\}) \cup \{p\}$  die neue Nichtbasisindexmenge nach dem Basiswechsel bezeichnet.

Ersetzen von  $x_q$  in den restlichen Gleichungen

$$x_i = \beta_i - \sum_{j \in N} \gamma_{ij} x_j = \beta_i - \gamma_{iq} x_q - \sum_{j \in N, j \neq q} \gamma_{ij} x_j, \quad i \neq p, i \in B,$$

liefert für  $i \neq p, i \in B$ , die Formeln

$$\begin{aligned} x_i &= \underbrace{\left( \beta_i - \gamma_{iq} \frac{\beta_p}{\gamma_{pq}} \right)}_{=\beta_i^+} - \underbrace{\left( -\frac{\gamma_{iq}}{\gamma_{pq}} \right)}_{=\gamma_{ip}^+} x_p - \sum_{j \in N, j \neq q} \underbrace{\left( \gamma_{ij} - \gamma_{iq} \frac{\gamma_{pj}}{\gamma_{pq}} \right)}_{=\gamma_{ij}^+} x_j \\ &= \beta_i^+ - \sum_{j \in N^+} \gamma_{ij}^+ x_j. \end{aligned}$$

Einsetzen in die Zielfunktion liefert

$$\begin{aligned}
 c^\top x &= d - \sum_{j \in N} \zeta_j x_j = d - \zeta_q x_q - \sum_{j \in N, j \neq q} \zeta_j x_j \\
 &= \underbrace{d - \zeta_q \frac{\beta_p}{\gamma_{pq}}}_{=d^+} - \underbrace{\left( -\frac{\zeta_q}{\gamma_{pq}} \right)}_{=\zeta_p^+} x_p - \sum_{j \in N, j \neq q} \underbrace{\left( \zeta_j - \zeta_q \frac{\gamma_{pj}}{\gamma_{pq}} \right)}_{=\zeta_j^+} x_j \\
 &= d^+ - \sum_{j \in N^+} \zeta_j^+ x_j
 \end{aligned}$$

Für die neue Basisindexmenge  $B^+ = (B \setminus \{p\}) \cup \{q\}$  und die neue Nichtbasisindexmenge  $N^+ = (N \setminus \{q\}) \cup \{p\}$  ergibt sich somit die neue Basisdarstellung

$$\begin{aligned}
 x_{B^+} &= \beta^+ - \Gamma^+ x_{N^+}, \\
 c^\top x^+ &= d^+ - \zeta^{+\top} x_{N^+},
 \end{aligned}$$

welche wieder in einem Tableau dargestellt werden kann. Beachtet man die oben hergeleiteten Updateformeln, so berechnet sich der Tableaueinhalt für die neue Basis zur Indexmenge  $B^+$  aus dem alten Tableaueinhalt (mit Daten  $\gamma_{ij}$ ,  $\beta_i$ ,  $\zeta_j$ ) nach folgendem Schema:

	$x_j, j \in N \setminus \{q\}$	$x_p$	
$x_i, i \in B \setminus \{p\}$	$\gamma_{ij} - \frac{\gamma_{iq}\gamma_{pj}}{\gamma_{pq}}$	$-\frac{\gamma_{iq}}{\gamma_{pq}}$	$\beta_i - \frac{\gamma_{iq}\beta_p}{\gamma_{pq}}$
$x_q$	$\frac{\gamma_{pj}}{\gamma_{pq}}$	$\frac{1}{\gamma_{pq}}$	$\frac{\beta_p}{\gamma_{pq}}$
	$\zeta_j - \frac{\zeta_q\gamma_{pj}}{\gamma_{pq}}$	$-\frac{\zeta_q}{\gamma_{pq}}$	$d - \frac{\zeta_q\beta_p}{\gamma_{pq}}$

#### Beispiel 4.2.4 (vgl. Beispiel 4.2.2)

Betrachte das Problem in primaler Normalform

$$\text{Minimiere } c^\top x \text{ unter den Nebenbedingungen } Ax = b, x \geq 0$$

mit den Daten ( $x_3, x_4, x_5$  sind Schlupfvariable):

$$c = \begin{pmatrix} -2 \\ -5 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 4 & 1 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 24 \\ 21 \\ 9 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}.$$

Eine zulässige Basislösung ist gegeben durch die Basisindexmenge  $B = \{3, 4, 5\}$  mit  $N = \{1, 2\}$  und es ergibt sich das folgende Starttableau für das Simplexverfahren:

	$x_1$	$x_2$		
$x_3$	1	4	24	6
$x_4$	3	1	21	21
$x_5$	1	1	9	9
	2	5	0	

Das Tableau ist noch nicht optimal und das Problem ist nicht unbeschränkt. Wir wählen den Nichtbasisindex  $q = 2 \in N$  mit  $\zeta_q = 5 > 0$  als Pivotspalte (wir hätten auch  $q = 1$  wählen können) und berechnen die Quotienten

$$\frac{x_3}{\gamma_{32}} = \frac{24}{4} = 6, \quad \frac{x_4}{\gamma_{42}} = \frac{21}{1} = 21, \quad \frac{x_5}{\gamma_{52}} = \frac{9}{1} = 9.$$

Diese Werte stehen in der rechten Spalte des Starttableaus. Der erste Quotient ist minimal und definiert die Pivotzeile  $p = 3 \in B$ .

Anwendung der Updateregeln liefert das Tableau 1:

	$x_1$	$x_3$		
$x_2$	$\frac{1}{4}$	$\frac{1}{4}$	6	$\frac{24}{4}$
$x_4$	$\frac{11}{4}$	$-\frac{1}{4}$	15	$\frac{60}{11}$
$x_5$	$\frac{3}{4}$	$-\frac{1}{4}$	3	4
	$\frac{3}{4}$	$-\frac{5}{4}$	-30	

Das Tableau ist nicht optimal und das Problem ist nicht unbeschränkt. Wir wählen die Pivotspalte  $q = 1$  und die Pivotzeile  $p = 5$  und führen einen Basiswechsel durch. Dieser liefert die Basisindexmenge  $B = \{2, 4, 1\}$ , die Nichtbasismenge  $N = \{5, 3\}$  und das Tableau 2:

	$x_5$	$x_3$		
$x_2$	$-\frac{1}{3}$	$\frac{1}{3}$	5	
$x_4$	$-\frac{11}{3}$	$\frac{2}{3}$	4	
$x_1$	$\frac{4}{3}$	$-\frac{1}{3}$	4	
	-1	-1	-33	

Tableau 2 ist optimal, da  $\zeta_N < 0$  gilt. Die zugehörige optimale Basislösung lautet  $x_2 = 5$ ,  $x_4 = 4$ ,  $x_1 = 4$ ,  $x_5 = 0$  und  $x_3 = 0$ . Der optimale Zielfunktionswert beträgt  $c^\top x = -33$ . ■

#### Bemerkung 4.2.5 (Aufwand des Simplexverfahrens)

Da  $A$  nur endlich viele Spalten enthält, gibt es maximal  $\binom{n}{m} = \frac{n!}{m!(n-m)!}$  Möglichkeiten,  $m$

linear unabhängige Spalten aus insgesamt  $n$  Spalten auszuwählen. Daher ist auch die Menge der Basisindexmengen endlich. Da jede Basisindexmenge einem Eckpunkt entspricht, ist die Anzahl der Eckpunkte endlich.

Klee und Minty [KM72] haben an einem Beispiel – dem sogenannten Klee und Minty Würfel – gezeigt, dass das Simplexverfahren im schlimmsten Fall exponentielle Komplexität besitzt und somit nicht zur Klasse der polynomialen Algorithmen gehört. Dies war ein enttäuschendes Resultat aus theoretischer Sicht und trug maßgeblich zur Entwicklung von innere Punkte-Verfahren bei, für die unter geeigneten Voraussetzungen eine polynomiale Komplexität gezeigt werden kann, vgl. [Kha79, Kar84, Wri97]. Die Tatsache, dass das Simplexverfahren – selbst für große Probleme – immer noch eines der am häufigsten benutzten Verfahren im OR ist, zeigt, dass die worst-case Komplexität in der Praxis in der Regel nicht beobachtet wird. Tiefer gehende Untersuchungen von Borgwardt [Bor87] zeigen: Falls die Eingabedaten  $A$ ,  $b$  und  $c$  sinnvoll verteilt sind, ist die erwartete Komplexität des Simplexverfahrens polynomial. ■

#### Bemerkung 4.2.6 (Phase 1 des Simplexverfahrens)

Eine zulässige Startbasislösung für das Simplexverfahren kann durch Lösen des linearen Hilfsproblems

$$\text{Minimiere } e^\top s = \sum_{i=1}^m s_i \quad \text{u.d.N.} \quad Ax + Is = b, \quad x \geq 0, \quad s \geq 0, \quad (4.10)$$

erreicht werden, wobei  $e = (1, \dots, 1)^\top \in \mathbb{R}^m$  ist. Hierbei gelte o.B.d.A.  $b \geq 0$ , was durch Multiplikation entsprechender Gleichungen in  $Ax = b$  mit  $-1$  stets erreicht werden kann. Für das Hilfsproblem beschreibt  $s = b$  und  $x = 0$  eine zulässige Basislösung.

Für das Hilfsproblem (4.10) gilt:

- (i) Hat das Hilfsproblem die Simplexlösung  $s = 0$ , so ist das zugehörige  $x$  eine zulässige Basislösung für das Ausgangsproblem.
- (ii) Besitzt das Hilfsproblem eine optimale Lösung  $s \neq 0$  und somit  $e^\top s > 0$ , so ist dies gleichbedeutend damit, dass das Ausgangsproblem keinen zulässigen Punkt besitzt. ■

#### 4.2.4 Endlichkeit des Simplexverfahrens

Es gibt noch zwei Freiheitsgrade in Schritt (4) des Algorithmus 4.2.1:

- Welcher Index  $q$  mit  $\zeta_q > 0$  soll gewählt werden? Die Beispiele 4.2.2 und 4.2.4 zeigen, dass es im allgemeinen mehrere Möglichkeiten gibt.



- Welcher Index  $p$  mit

$$\frac{\beta_p}{\gamma_{pq}} = \min \left\{ \frac{\beta_i}{\gamma_{iq}} \mid \gamma_{iq} > 0, i \in B \right\}$$

soll gewählt werden? Im allgemeinen gibt es auch hier mehrere Möglichkeiten.

Eine übliche Wahl geht auf Dantzig zurück und wählt diejenige Pivotspalte  $k$  mit

$$\zeta_k = \max\{\zeta_j \mid \zeta_j > 0, j \in N\}. \quad (4.11)$$

Von dieser Wahl erhofft man sich eine möglichst starke Abnahme der Zielfunktion beim Basiswechsel. Hoffmann [Hof53] und Marshall und Suurballe [MS69] zeigten jedoch an einem Beispiel, dass diese Wahl zu Zyklen im Simplexverfahren führen kann, so dass das Verfahren nicht immer terminiert. Dieser Fall tritt allerdings nur in sogenannten entarteten Basislösungen auf.

### Beispiel 4.2.7 (Schleifenbildung beim Simplexverfahren)

Das folgende Beispiel wurde von Marshall und Suurballe [MS69] zum Nachweis von Zyklen im Simplexverfahren verwendet:

$$\text{Minimiere } c^\top x \quad \text{u.d.N.} \quad Ax = b, \quad x \geq 0_{\mathbb{R}^7}$$

mit den Daten ( $x_5, x_6, x_7$  sind Schlupfvariable)

$$c = \begin{pmatrix} -10 \\ 57 \\ 9 \\ 24 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.5 & -5.5 & -2.5 & 9 & 1 & 0 & 0 \\ 0.5 & -1.5 & -0.5 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix}.$$

Anwendung des Simplexverfahrens, wobei die Pivotspalte stets gemäß (4.11) gewählt wird, führt auf die folgenden Tableaus. Beachte, dass das letzte Tableau bis auf Permutationen der Spalten dem ersten Tableau entspricht. Folglich führt eine wiederholte Wahl derselben Pivotelemente zu einem Zyklus.

Starttableau:

	$x_1$	$x_2$	$x_3$	$x_4$		
$x_5$	0.5	-5.5	-2.5	9	0	0
$x_6$	0.5	-1.5	-0.5	1	0	0
$x_7$	1	0	0	0	1	1
	10	-57	-9	-24	0	

Pivotzeile und -spalte:  $p = 5$ ,  $q = 1$ .

Tableau 1:

	$x_5$	$x_2$	$x_3$	$x_4$		
$x_1$	2	-11	-5	18	0	-
$x_6$	-1	4	2	-8	0	0
$x_7$	-2	11	5	-18	1	$\frac{1}{11}$
	-20	53	41	-204	0	

Pivotzeile und -spalte:  $p = 6$ ,  $q = 2$ .

Tableau 2:

	$x_5$	$x_6$	$x_3$	$x_4$		
$x_1$	-0.75	2.75	0.5	-4	0	0
$x_2$	-0.25	0.25	0.5	-2	0	0
$x_7$	0.75	-2.75	-0.5	4	1	-
	-6.75	-13.25	14.5	-98	0	

Pivotzeile und -spalte:  $p = 1$ ,  $q = 3$ .

Tableau 3:

	$x_5$	$x_6$	$x_1$	$x_4$		
$x_3$	-1.5	5.5	2	-8	0	-
$x_2$	0.5	-2.5	-1	2	0	0
$x_7$	0	0	1	0	1	-
	15	-93	-29	18	0	

Pivotzeile und -spalte:  $p = 2$ ,  $q = 4$ .

Tableau 4:

	$x_5$	$x_6$	$x_1$	$x_2$		
$x_3$	0.5	-4.5	-2	4	0	0
$x_4$	0.25	-1.25	-0.5	0.5	0	0
$x_7$	0	0	1	0	1	-
	10.5	-70.5	-20	-9	0	

Pivotzeile und -spalte:  $p = 3$ ,  $q = 5$ .

Tableau 5:

	$x_3$	$x_6$	$x_1$	$x_2$		
$x_5$	2	-9	-4	8	0	-
$x_4$	-0.5	1	0.5	-1.5	0	0
$x_7$	0	0	1	0	1	-
	-21	24	22	-93	0	

Pivotzeile und -spalte:  $p = 4$ ,  $q = 6$ .

Tableau 6:

	$x_3$	$x_4$	$x_1$	$x_2$	
$x_5$	-2.5	9	0.5	-5.5	0
$x_6$	-0.5	1	0.5	-1.5	0
$x_7$	0	0	1	0	1
	-9	-24	10	-57	0

Es ist interessant zu beobachten, dass alle Tableaus dieselbe (entartete) Basislösung  $x = (0, 0, 0, 0, 0, 0, 1)^\top$  beschreiben, obwohl sich die Tableaus selbst unterscheiden. Geometrisch bleibt das Verfahren also in derselben Ecke des zulässigen Bereichs hängen. ■

Wir wollen einige Überlegungen anstellen, wann die Schleifenbildung eintreten kann. Zunächst nehmen wir an, dass bei der Durchführung des Simplexverfahrens nur Pivotzeilen  $p$  mit  $\beta_p > 0$  gewählt werden (dies war beim vorangehenden Beispiel nicht der Fall). Die Updateformeln aus Abschnitt 4.2.3 zeigen, dass die Zielfunktion in diesem Fall strikt fällt, d.h. es gilt  $d^+ < d$ . Da das Simplexverfahren per Konstruktion ausschließlich zulässige Basislösungen berechnet, von denen es nur endlich viele gibt und von denen mindestens eine optimal ist, endet das Verfahren in diesem Fall nach endlich vielen Schritten (falls überhaupt eine Lösung existiert).

Eine Schleife kann also nur dann auftreten, wenn eine Pivotzeile  $p$  mit  $\beta_p = 0$  gewählt wird, da dann  $d^+ = d$  aus den Updateformeln aus Abschnitt 4.2.3 folgt. Der Zielfunktionswert ändert sich somit nicht beim Übergang von der Basislösung  $x$  zur benachbarten Basislösung  $x^+$ . Desweiteren ändert sich gemäß Abschnitt 4.2.3 auch  $\beta^+$  nicht, d.h. es gilt  $x_{B^+} = \beta_{B^+}^+ = \beta_B = x_B$  (vgl. Beispiel 4.2.7). Das Verfahren stagniert also in der Basislösung  $x = x^+$ . Beachte, dass sich die Darstellung der Basislösung, also das Tableau, sehr wohl ändert, da ein Basiswechsel stattfindet. In diesem Fall spricht man von einer entarteten Basislösung.

#### Definition 4.2.8 (Entartung)

Eine zulässige Basislösung  $x$  mit  $x_i = 0$  für mindestens ein  $i \in B$  heißt **entartet**. Andernfalls heißt sie **nicht entartet**.

Als Fazit halten wir fest, dass eine Schleifenbildung nur auftreten kann (aber nicht muß), wenn entartete Basislösungen auftreten.

Es gibt eine einfache Regel, die die Schleifenbildung verhindert (der Beweis ist allerdings sehr kompliziert, so dass wir ihn hier nicht führen wollen).

#### Blandsche Regel

Wähle für die Pivotspalte  $q \in N$  und die Pivotzeile  $p \in B$  in Schritt (4) des Algorithmus

mus 4.2.1 stets den Kandidaten mit dem kleinsten Index. Dann treten keine Schleifen auf und das Simplexverfahren endet nach endlich vielen Schritten.

### Bemerkung 4.2.9

*In der Praxis tritt die Schleifenbildung in der Regel nicht auf, so dass dort häufig auf die Blandsche Regel verzichtet wird.*

## 4.3 Dualität, Sensitivität und Parametrische Optimierung

Häufig hängen Optimierungsprobleme von Parametern ab, die selbst **keine** Optimierungsvariablen sind. Es stellt sich dann die Frage, wie die Lösung des Optimierungsproblems von den Parametern abhängt bzw. wie sensitiv diese Abhängigkeit ist? Wir werden sehen, dass die Lösung unter geeigneten Voraussetzungen (u.a. hinreichende Bedingung zweiter Ordnung, LICQ) sogar stetig differenzierbar von den Parametern abhängt. Zunächst beschränken wir uns aber auf lineare Optimierungsprobleme, um das sogenannte duale Optimierungsproblem und dessen Bedeutung für die Sensitivitätsanalyse zu untersuchen.

### 4.3.1 Dualität

Dem Standardproblem

$$\text{Minimiere } c^\top x \quad \text{u.d.N.} \quad Ax = b, x \geq 0 \quad (4.12)$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $c \in \mathbb{R}^n$ , welches im folgenden stets als **primales lineares Optimierungsproblem** oder **Primalproblem** bezeichnet wird, kann ein sogenanntes **duales lineares Optimierungsproblem** oder **Dualproblem** zugeordnet werden.

### Beispiel 4.3.1

*Eine Firma produziert drei Produkte  $P_1$ ,  $P_2$  und  $P_3$ . Der Gewinn pro Einheit der Produkte betrage 10, 5 bzw. 5.5 Geldeinheiten. Die Produktion benötigt Rohmaterialien  $B_1$ ,  $B_2$ ,  $B_3$  und  $B_4$ , von denen 1500, 200, 1200 und 900 Einheiten verfügbar sind. Zur Herstellung einer Einheit der jeweiligen Produkte werden die in der folgenden Tabelle angegebenen Mengen an Rohmaterialien benötigt.*

	$P_1$	$P_2$	$P_3$
$B_1$	30	10	50
$B_2$	5	0	3
$B_3$	20	10	50
$B_4$	10	20	30

*Es bezeichne  $x_i$ ,  $i = 1, 2, 3$ , die produzierten Mengen der Produkte  $P_i$ ,  $i = 1, 2, 3$ . Die Firma möchte ihren Gewinn maximieren und muss daher das folgende primale Problem lösen:*

Maximiere

$$10x_1 + 5x_2 + 5.5x_3$$

unter den Nebenbedingungen

$$30x_1 + 10x_2 + 50x_3 \leq 1500,$$

$$5x_1 + 3x_3 \leq 200,$$

$$20x_1 + 10x_2 + 50x_3 \leq 1200,$$

$$10x_1 + 20x_2 + 30x_3 \leq 900,$$

$$x_1, x_2, x_3 \geq 0.$$

*Angenommen, eine zweite Firma bietet der ersten Firma an, alle Rohmaterialien aufzukaufen. Die zweite Firma bietet den Preis  $y_i \geq 0$  pro Einheit des Rohmaterials  $B_i$ ,  $i = 1, \dots, 4$ . Natürlich möchte die zweite Firma ihre Kosten*

$$1500y_1 + 200y_2 + 1200y_3 + 900y_4$$

*minimieren. Desweiteren wird die erste Firma das Angebot nur dann akzeptieren, wenn die Preise pro Einheit der Produkte  $P_j$ ,  $j = 1, 2, 3$ , größer oder gleich dem (nicht realisierten) Gewinn  $c_j$ ,  $j = 1, 2, 3$ , aus dem Verkauf der Produkte sind, d.h. es muss gelten*

$$30y_1 + 5y_2 + 20y_3 + 10y_4 \geq 10,$$

$$10y_1 + 10y_3 + 20y_4 \geq 5,$$

$$50y_1 + 3y_2 + 50y_3 + 30y_4 \geq 5.5.$$

*Insgesamt muss die zweite Firma also das folgende sogenannte duale Problem lösen:*

Minimiere

$$1500y_1 + 200y_2 + 1200y_3 + 900y_4$$

unter den Nebenbedingungen

$$30y_1 + 5y_2 + 20y_3 + 10y_4 \geq 10,$$

$$10y_1 + 10y_3 + 20y_4 \geq 5,$$

$$50y_1 + 3y_2 + 50y_3 + 30y_4 \geq 5.5,$$

$$y_1, y_2, y_3, y_4 \geq 0.$$

■

Duale Optimierungsprobleme sind aus folgenden Gründen wichtig:

- Duale Programme liefern untere Schranken für den optimalen Zielfunktionswert des primalen Problems.
- Gelegentlich ist das duale Programm einfacher zu lösen als das primale Problem.
- Das duale Optimierungsproblem erlaubt, Änderungen des Minimalwerts des Primalproblems bei Störungen der Problemdata abzuschätzen.

Diesen letzten Aspekt, die so genannte Sensitivität des Ausgangsproblems gegenüber Störungen, werden wir später im Beispiel 4.3.7 auch geometrisch illustrieren.

### Problem 4.3.2 (Duales lineares Optimierungsproblem)

Das dem primalen Optimierungsproblem (4.12) zugeordnete **duale Problem** lautet:

$$\text{Maximiere } b^\top y \quad \text{unter der Nebenbedingung } A^\top y \leq c. \quad (4.13)$$

### Bemerkung 4.3.3

Das duale Problem des dualen Problems liefert wieder das primale Problem (Beweis durch Ausrechnen). ■

Eine erste Eigenschaft des dualen Problems liefert

### Satz 4.3.4 (Schwacher Dualitätssatz)

Für alle primal zulässigen Punkte  $x$  (mit  $Ax = b$ ,  $x \geq 0$ ) und alle dual zulässigen Punkte  $y$  (mit  $A^\top y \leq c$ ) gilt

$$b^\top y \leq c^\top x.$$

**Beweis:** Die Behauptung folgt wegen  $b = Ax$  und  $x \geq 0$  aus

$$b^\top y = (Ax)^\top y = x^\top A^\top y \leq x^\top c = c^\top x.$$

Der schwache Dualitätssatz liefert eine Motivation für das duale Problem. Dual zulässige Werte liefern untere Schranken für den optimalen Zielfunktionswert des primalen Problems. Umgekehrt liefern primal zulässige Werte obere Schranken für den optimalen Zielfunktionswert des dualen Problems. Diese Eigenschaft kann z.B. beim Branch & Bound-Verfahren für ganzzahlige Probleme zur Bestimmung unterer Schranken ausgenutzt werden.

Darüber hinaus gilt:

**Satz 4.3.5**

Seien  $x$  primal und  $y$  dual zulässig.

(i) Gilt  $b^\top y = c^\top x$ , so ist  $x$  optimal für das primale Problem in (4.12) und  $y$  ist optimal für das duale Problem 4.3.2.

(ii)  $b^\top y = c^\top x$  gilt genau dann, wenn die **Komplementaritätsbedingungen** gelten (vgl. Kapitel 1):

$$x^\top (A^\top y - c) = 0.$$

**Beweis:** Die erste Behauptung in (i) folgt sofort aus dem schwachen Dualitätssatz. Die zweite Behauptung in (ii) folgt aus der Darstellung

$$c^\top x - b^\top y = c^\top x - y^\top b = c^\top x - y^\top Ax = (c - A^\top y)^\top x.$$

Gilt  $c^\top x = b^\top y$ , so muß jeder Summand wegen  $x \geq 0$  und  $c - A^\top y \geq 0$  einzeln verschwinden, damit die Summe Null ergibt. Gelten umgekehrt die Komplementaritätsbedingungen, so verschwindet jeder Summand in der rechten Summe und es folgt  $c^\top x = b^\top y$ . ■

Es stellt sich die Frage, ob  $c^\top x = b^\top y$  tatsächlich erreicht werden kann. Sei also  $x$  eine primal optimale Basislösung mit Basisindexmenge  $B$ , die mit dem Simplexverfahren berechnet wurde, d.h. es gilt  $x_B \geq 0$  und  $x_N = 0$ . Insbesondere gilt dann das Optimalitätskriterium  $\zeta_N \leq 0$  mit  $\zeta_N^\top = c_B^\top A_B^{-1} A_N - c_N^\top$ . Es sei  $y$  definiert durch

$$y^\top := c_B^\top A_B^{-1}.$$

Wir zeigen, dass das so definierte  $y$  eine Lösung des dualen Problems ist. Zunächst gilt

$$c^\top x = c_B^\top x_B = c_B^\top A_B^{-1} b = y^\top b = b^\top y.$$

Desweiteren ist  $y$  auch dual zulässig, denn es gelten  $A_B^\top y = c_B$  und

$$A_N^\top y - c_N = A_N^\top (A_B^\top)^{-1} c_B - c_N = \zeta_N \leq 0.$$

Letzteres ist gerade das primale Optimalitätskriterium. Beides zusammen liefert  $A^\top y \leq c$ . Damit ist folgendes gezeigt:

Besitzt das primale Problem eine Optimallösung, so besitzt auch das duale Problem eine Optimallösung, wobei eine spezielle duale Optimallösung für die Basisindexmenge  $B$  durch  $y^\top = c_B^\top A_B^{-1}$  gegeben ist (es kann weitere duale Lösungen geben).

Das folgende Resultat zeigt, dass auch die Umkehrung richtig ist.

**Satz 4.3.6 (Starker Dualitätssatz)**

Das primale Problem in (4.12) besitzt genau dann eine optimale Lösung, wenn das duale Problem 4.3.2 eine optimale Lösung besitzt. Darüber hinaus stimmen dann die Zielfunktionswerte überein, d.h. es gilt  $c^\top x = b^\top y$ . ■

**4.3.2 Sensitivität und Schattenpreise**

Lösungen des dualen Problems besitzen eine für die Wirtschaftswissenschaften wichtige Interpretation. Sie geben – unter gewissen Voraussetzungen – die Sensitivität des primalen Zielfunktionswertes  $c^\top x$  bzgl. Störungen im Vektor  $b$  an und sind als **Schattenpreise** bekannt. Hintergrund:  $b$  bezeichnet häufig Kapazitätsschranken, etwa die Größe eines Lagers. Für einen Unternehmer ist nun interessant, wie sich eine Vergrößerung oder Verkleinerung des Lagers auf die Kosten auswirkt.

**Beispiel 4.3.7**

Ein Landwirt möchte 40 Hektar mit Zuckerrüben und Weizen bepflanzen. Er hat hierfür 2400 Euro und 312 Arbeitstage zur Verfügung. Für jeden Hektar belaufen sich seine Anpflanzungskosten auf 40 Euro für Zuckerrüben und auf 120 Euro für Weizen. Für Zuckerrüben benötigt er 6 Arbeitstage pro Hektar und für Weizen 12 Arbeitstage pro Hektar. Der Gewinn beläuft sich auf 100 Euro pro Hektar für Zuckerrüben und auf 250 Euro pro Hektar für Weizen.

Der Landwirt möchte 2400 Euro investieren, aber zusätzlich unvorhergesehene Ausgaben in seiner Kalkulation berücksichtigen. Er nimmt daher an, dass  $2400 + \delta$  Euro zur Verfügung stehen, wobei  $\delta \in \mathbb{R}$  eine durch unvorhergesehene Ausgaben verursachte Störung bezeichnet. Natürlich möchte der Landwirt seinen Gewinn maximieren:

Minimiere

$$f(x_1, x_2) = -100x_1 - 250x_2$$

unter den Nebenbedingungen

$$\begin{aligned} x_1 + x_2 &\leq 40, \\ 40x_1 + 120x_2 &\leq 2400 + \delta, \\ 6x_1 + 12x_2 &\leq 312, \\ x_1, x_2 &\geq 0. \end{aligned}$$

Das ungestörte Problem mit  $\delta = 0$  können wir grafisch lösen:



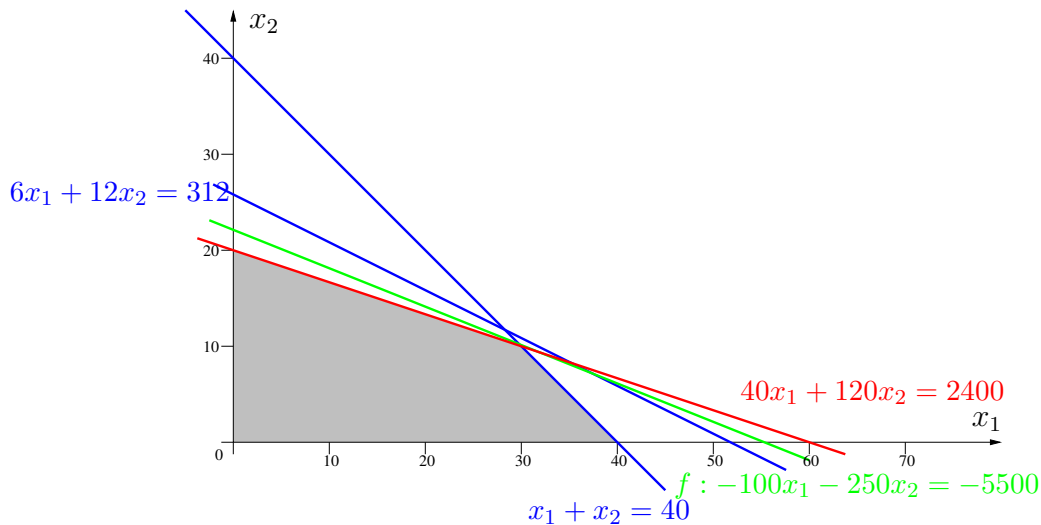


Abbildung 4.6: Lösung des ungestörten Problems mit  $\delta = 0$ . Die optimale Lösung lautet  $(x_1, x_2)^\top = (30, 10)^\top$  mit Zielfunktionswert  $-5500$ .

Die optimale Lösung ist eine Ecke der zulässigen Menge und erfüllt die ersten beiden Nebenbedingungen ohne Schlupf, alle übrigen strikt.

Was geschieht, wenn Störungen  $\delta \neq 0$  auftreten?

$\delta$  beeinflusst nur die zweite Nebenbedingung, wobei deren Steigung nicht verändert wird. Eine Änderung in  $\delta$  führt also zu einer Parallelverschiebung der Geraden  $40x_1 + 120x_2 = 2400$ , Abbildung 4.7 zeigt die Situation für  $\delta = -600$ .

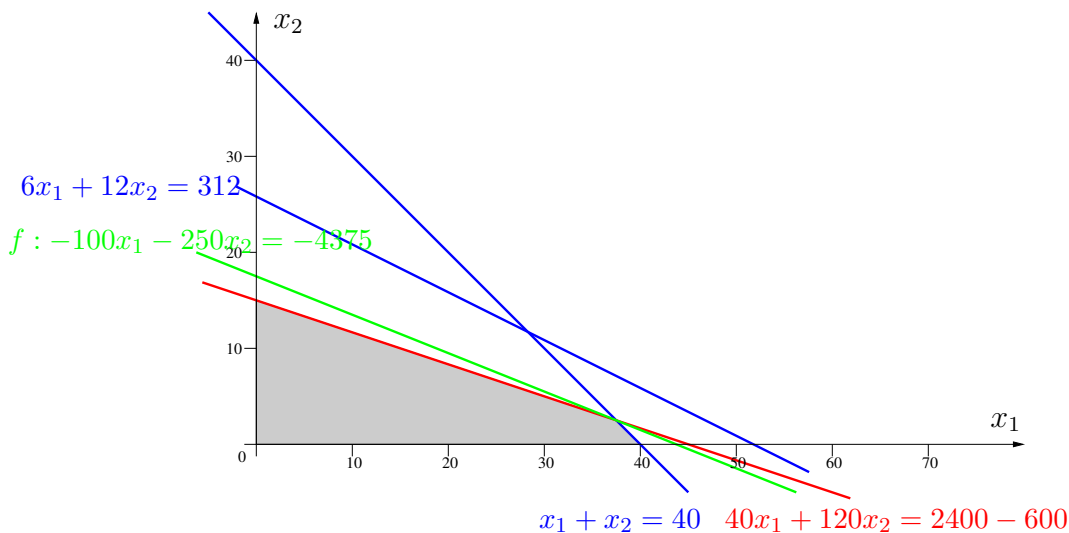


Abbildung 4.7: Lösung des Problems für  $\delta = -600$ . Die optimale Lösung lautet  $(x_1, x_2)^\top = (37.5, 2.5)^\top$  mit Zielfunktionswert  $-4375$ .

Offenbar haben sich der zulässige Bereich und die optimale Lösung verändert, aber die optimale Lösung erfüllt immer noch die ersten beiden Nebenbedingungen ohne Schlupf, die übrigen Nebenbedingungen strikt.

Was geschieht, wenn wir  $\delta$  weiter reduzieren? Abbildung 4.8 zeigt die Situation für  $\delta = -1200$ .

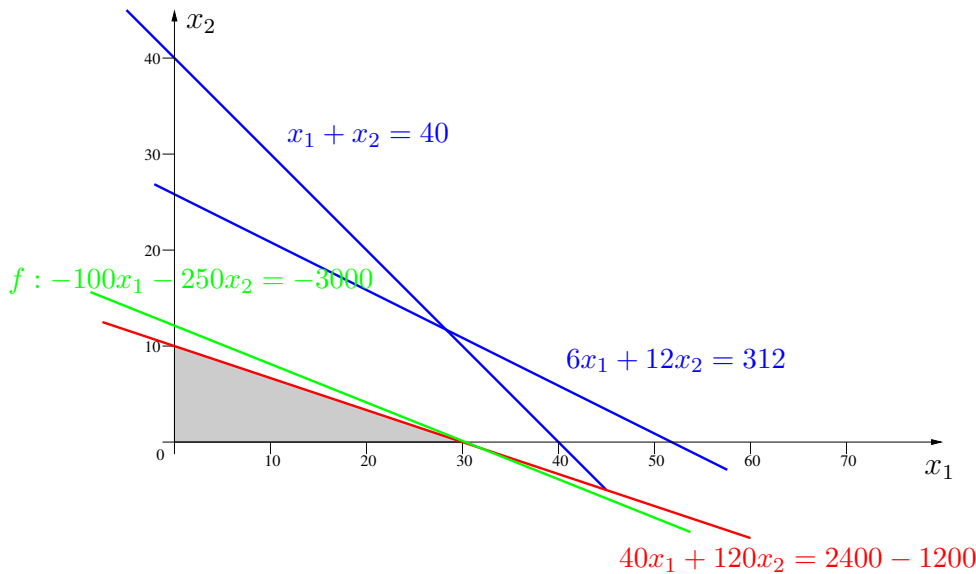


Abbildung 4.8: Lösung des Problems für  $\delta = -1200$ . Die optimale Lösung lautet  $(x_1, x_2)^\top = (30, 0)^\top$  mit Zielfunktionswert  $-3000$ .

Jetzt erfüllt die Optimallösung die zweite Nebenbedingung und die zweite Vorzeichenbedingung ohne Schlupf, alle übrigen Nebenbedingungen strikt. Die Struktur der Optimallösung hat sich also geändert, was sich auch im qualitativen Verhalten der Minimalwerte niederschlägt.

Eine grafische Lösung des Problems für alle Werte von  $\delta$  liefert die Minimallösungen

$$\begin{pmatrix} x_1(\delta) \\ x_2(\delta) \end{pmatrix} = \begin{cases} \begin{pmatrix} 0 \\ 26 \end{pmatrix}, & \text{falls } 720 \leq \delta, \\ \begin{pmatrix} 36 - \frac{1}{20}\delta \\ 8 + \frac{1}{40}\delta \end{pmatrix}, & \text{falls } 160 \leq \delta < 720, \\ \begin{pmatrix} 30 - \frac{1}{80}\delta \\ 10 + \frac{1}{80}\delta \end{pmatrix}, & \text{falls } -800 \leq \delta < 160, \\ \begin{pmatrix} 60 + \frac{1}{40}\delta \\ 0 \end{pmatrix}, & \text{falls } -2400 \leq \delta < -800, \\ \text{keine Lösung}, & \text{falls } \delta < -2400, \end{cases}$$

sowie die sogenannte Minimalwertfunktion

$$w(\delta) = c^\top x(\delta) = \begin{cases} -6500, & \text{falls } 720 \leq \delta, \\ -5600 - 1.25\delta, & \text{falls } 160 \leq \delta < 720, \\ -5500 - 1.875\delta, & \text{falls } -800 \leq \delta < 160, \\ -6000 - 2.5\delta, & \text{falls } -2400 \leq \delta < -800, \\ \infty, & \text{falls } \delta < -2400, \end{cases}$$

welche die optimalen Zielfunktionswerte angibt und offenbar stetig, stückweise linear und konvex ist auf dem Intervall  $[-2400, \infty)$ .

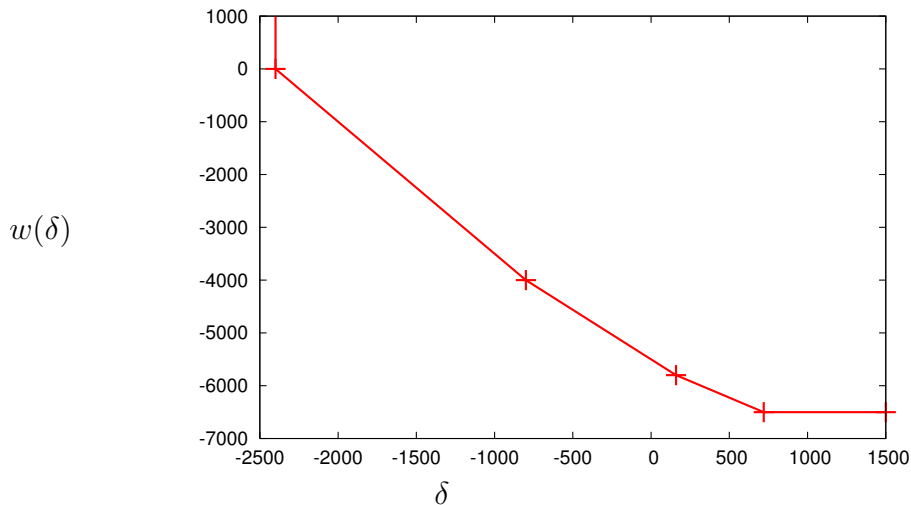


Abbildung 4.9: Minimalwertfunktion

Wir beschränken uns im folgenden auf Störungen des Vektors  $b$  im Standardproblem (4.12). Störungen in  $c$  oder  $A$  können ebenfalls betrachtet werden, sind aber komplizierter. ■

#### **Problem 4.3.8** (Gestörtes primales lineares Optimierungsproblem)

Für Störungen  $\Delta b \in \mathbb{R}^m$  lautet das gestörte Standardproblem:

$$\text{Minimiere } c^\top x \quad \text{unter den Nebenbedingungen } Ax = b + \Delta b, \quad x \geq 0. \quad (4.14)$$

Offenbar entsteht für  $\Delta b = 0$  gerade das Standardproblem (4.12), welches auch als **ungestörtes Problem** oder **Nominalproblem** bezeichnet wird. Jeder Störung  $\Delta b$  kann der zugehörige optimale Zielfunktionswert des Problems 4.3.8 zugeordnet werden (falls er existiert). Die Funktion

$$w(\Delta b) := \inf\{c^\top x \mid Ax = b + \Delta b, \quad x \geq 0\}$$

heißt **optimale Wertefunktion** des gestörten Primalproblems 4.3.8. Entsprechend kann jeder Störung die Lösung des gestörten Problems zugeordnet werden:

$$\Delta b \mapsto x(\Delta b).$$

Wir wollen nun untersuchen, wie sich die optimale Lösung des gestörten Problems 4.3.8 in der Nähe der ungestörten Lösung für  $\Delta b = 0$  verhält. Wir betrachten hier nur den Fall einer nicht entarteten Lösung mit Basislösung  $x_B > 0$ . Eine Erweiterung auf entartete Lösungen ist möglich, aber komplizierter.

Das ungestörte Problem mit  $\Delta b = 0$  besitze die optimale, nicht entartete Basislösung  $x_B(0) = A_B^{-1}b > 0$ ,  $x_N(0) = 0$  mit Basisindexmenge  $B$  und Nichtbasisindexmenge  $N$ .

Nun betrachten wir kleine Störungen  $\Delta b \neq 0$  und definieren

$$x_B(\Delta b) := A_B^{-1}(b + \Delta b), \quad x_N(\Delta b) := 0. \quad (4.15)$$

Die Werte  $x_B(\Delta b)$  und  $x_N(\Delta b)$  sind zulässig, solange

$$x_B(\Delta b) = A_B^{-1}(b + \Delta b) = x_B(0) + A_B^{-1}\Delta b \geq 0$$

gilt. Wegen  $x_B(0) = A_B^{-1}b > 0$  (keine Entartung!) kann dies für hinreichend kleine Störungen  $\Delta b$  stets erreicht werden.

Interessanterweise haben Störungen  $\Delta b$  in  $b$  keine Auswirkungen auf die negativen reduzierten Kosten

$$\zeta_N^\top = c_B^\top A_B^{-1} A_N - c_N^\top$$

im Simplexverfahren. Da  $x_B(0)$  optimale Basislösung war, gilt gemäß Simplexverfahren  $\zeta_N \leq 0$ . Diese Optimalitätsbedingung bleibt auch für  $x_B(\Delta b)$  und  $x_N(\Delta b) = 0$  in (4.15) erfüllt, so dass (4.15) optimal für das gestörte Problem ist.

Für den optimalen Zielfunktionswert des gestörten Problems ergibt sich

$$w(\Delta b) = c_B^\top x_B(\Delta b) = c_B^\top x_B(0) + c_B^\top A_B^{-1} \Delta b = w(0) + y^\top \Delta b.$$

Also:

#### **Satz 4.3.9 (Sensitivitätssatz (linearer Fall))**

Sei  $\text{Rang}(A) = m$  und das ungestörte Problem besitze eine nicht entartete optimale Basislösung mit Basisindexmenge  $B$  und Nichtbasisindexmenge  $N$ . Dann ist

$$x_B(\Delta b) = A_B^{-1}(b + \Delta b), \quad x_N(\Delta b) = 0,$$

für alle Störungen  $\Delta b = (\Delta b_1, \dots, \Delta b_m)^\top$  hinreichend nahe bei 0 optimale zulässige Basislösung für das gestörte Problem und es gilt

$$\Delta w := w(\Delta b) - w(0) = y^\top \Delta b = \sum_{i=1}^m y_i \Delta b_i,$$

wobei die duale Lösung  $y^\top = c_B^\top A_B^{-1}$  die **Sensitivität der Zielfunktion bzgl. Störungen in  $b$**  angibt. ■

Aus dieser Darstellung läßt sich folgendes ablesen: Ist  $y_i$  betragsmässig sehr gross, so haben kleine Störungen  $\Delta b_i$  einen verhältnismässig **großen Einfluss** auf den optimalen Zielfunktionswert des gestörten Problems. Ist  $y_i$  betragsmässig sehr klein, so haben kleine Störungen  $\Delta b_i$  einen verhältnismässig **kleinen Einfluss** auf den optimalen Zielfunktionswert des gestörten Problems. Im Extremfall  $y_i = 0$  haben Störungen  $\Delta b_i$  **keinen** Einfluss auf den optimalen Zielfunktionswert des gestörten Problems.

Durch differenzieren nach  $\Delta b$  ergibt sich die

Schattenpreisformel:

$$w'(\Delta b) = y.$$

**Beispiel 4.3.10** (vgl. Beispiel 4.3.7)

Wir betrachten wieder Beispiel 4.3.7:

Minimiere  $f(x_1, x_2) = -100x_1 - 250x_2$  unter den Nebenbedingungen

$$x_1 + x_2 \leq 40, \quad 40x_1 + 120x_2 \leq 2400 + \delta, \quad 6x_1 + 12x_2 \leq 312, \quad x_1, x_2 \geq 0.$$

Das Simplexverfahren angewendet auf das ungestörte Problem mit  $\delta = 0$  liefert:

Starttableau:

	$x_1$	$x_2$		
$x_3$	1	1	40	40
$x_4$	40	120	2400	60
$x_5$	6	12	312	52
	100	250	0	

Pivotzeile- und spalte:  $p = 3$ ,  $q = 1$

Tableau 1:

	$x_3$	$x_2$		
$x_1$	1	1	40	40
$x_4$	-40	80	800	10
$x_5$	-6	6	72	12
	-100	150	-4000	

Pivotzeile- und spalte:  $p = 4$ ,  $q = 2$

Tableau 2:

	$x_3$	$x_4$		
$x_1$	1.5	-0.0125	30	
$x_2$	-0.5	0.0125	10	
$x_5$	-3	-0.075	12	
	-25	-1.875	-5500	

Tableau 2 ist optimal und die nicht entartete optimale Lösung lautet  $x = (30, 10, 0, 0, 12)^\top$ . Die zugehörige duale Lösung ist

$$y^\top = c_B^\top A_B^{-1} = (-25, -1.875, 0),$$

wobei  $B = \{1, 2, 5\}$  und

$$c_B = \begin{pmatrix} -100 \\ -250 \\ 0 \end{pmatrix}, \quad A_B = \begin{pmatrix} 1 & 1 & 0 \\ 40 & 120 & 0 \\ 6 & 12 & 1 \end{pmatrix}, \quad A_B^{-1} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{80} & 0 \\ -\frac{1}{2} & \frac{1}{80} & 0 \\ -3 & -\frac{3}{40} & 1 \end{pmatrix}.$$

Der Sensitivitätssatz besagt:

$$(i) \quad \underbrace{\begin{pmatrix} x_1(\delta) \\ x_2(\delta) \\ x_5(\delta) \end{pmatrix}}_{=x_B(\delta)} = \underbrace{\begin{pmatrix} x_1(0) \\ x_2(0) \\ x_5(0) \end{pmatrix}}_{=x_B(0)} + A_B^{-1} \begin{pmatrix} 0 \\ \delta \\ 0 \end{pmatrix} = \begin{pmatrix} 30 \\ 10 \\ 12 \end{pmatrix} + \delta \begin{pmatrix} -\frac{1}{80} \\ \frac{1}{80} \\ -\frac{3}{40} \end{pmatrix}$$

ist optimal für  $\delta$  hinreichend nahe bei 0. Wie groß kann die Störung  $\delta$  sein? Nun,  $x_B(\delta)$  muss zulässig sein, d.h.

$$30 - \frac{1}{80}\delta \geq 0, \quad 10 + \frac{1}{80}\delta \geq 0, \quad 12 - \frac{3}{40}\delta \geq 0.$$

Diese drei Ungleichungen sind für  $\delta \in [-800, 160]$  erfüllt. Daher ist  $x_B(\delta)$  für alle  $\delta \in [-800, 160]$  optimal.

(ii)

$$\Delta w = w(\delta) - w(0) = y^\top \begin{pmatrix} 0 \\ \delta \\ 0 \end{pmatrix} = -1.875\delta.$$

Die Änderung des negativ genommenen Profits bzgl.  $\delta$  beträgt  $-1.875\delta$ .

Vergleiche dies mit der grafischen Untersuchung in Beispiel 4.3.7! ■

#### **Bemerkung 4.3.11 (Achtung!)**

Die Annahme, dass die ungestörte Lösung nicht entartet ist, ist wesentlich und kann nicht ohne Weiteres weggelassen werden. Die obige Argumentation funktioniert nur, wenn die Indextmengen  $B$  und  $N$  sich nicht ändern. Dies kann nur für Störungen  $\Delta b$  hinreichend nahe bei 0 garantiert werden. Für große Störungen oder im entarteten Fall ändern sich die Indextmengen  $B$  und  $N$  üblicherweise und die optimale Wertefunktion ist nicht differenzierbar, vgl. Beispiel 4.3.7. ■

## Kapitel 5

### Restringierte Optimierung

In den folgenden Abschnitten werden wir notwendige Bedingungen für restringierte Optimierungsprobleme kennen lernen. Darauf aufbauend untersuchen wir Lösungsverfahren, darunter das Lagrange-Newton-Verfahren und Penalty-Verfahren. Wir greifen dabei auf die Bücher [BS79], [GMW81], [FM90], [Spe93], [Man94], [GK99], [Alt02], [GK02] zurück.

#### **Problem 5.0.12** (Standard-Optimierungsproblem)

Finde  $x \in \mathbb{R}^n$ , so dass  $f(x)$  minimal wird unter den Nebenbedingungen

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

In Kurzform schreiben wir auch:

$$\text{Minimiere } f(x) \quad \text{u.d.N.} \quad g(x) \leq 0, \quad h(x) = 0.$$

Zur numerischen Lösung des Standard-Optimierungsproblems gibt es im Wesentlichen die folgenden Herangehensweisen:

- (a) **Penalty- und Multiplikator-Verfahren:** Diese Verfahren basieren auf der Ankopplung der Nebenbedingungen an die Zielfunktion mithilfe eines gewichteten Strafterms (Penalty-Term), der unzulässige Punkte bestraft. Dadurch werden die Nebenbedingungen eliminiert und man kann Verfahren der unrestringierten Optimierung (Gradientenverfahren, Newtonverfahren, Quasi-Newton-Verfahren) anwenden. Zielfunktion und Strafterm müssen jedoch geeignet gewichtet werden, damit man letztendlich eine zulässige Lösung bekommt.
- (b) **Sequentielle quadratische Programmierung (SQP):** SQP-Verfahren basieren auf der lokalen Approximation des Standard-Optimierungsproblems durch ein linear-quadratisches Optimierungsproblem, dessen Lösung die Suchrichtung in einem iterativen Verfahren liefert. SQP-Verfahren erweitern das Lagrange-Newton-Verfahren auf Probleme mit Ungleichungsrestriktionen.

- (c) **Innere-Punkte-Verfahren (IP):** Innere-Punkte-Verfahren verwenden sogenannte Barriere-Funktionen, um Ungleichungsnebenbedingungen zu eliminieren und diese ähnlich wie bei Penalty-Verfahren mithilfe eines gewichteten Strafterms an die Zielfunktion zu koppeln. Im Gegensatz zu Penalty- und Multiplikator-Verfahren wird dabei nicht nur das Verlassen des zulässigen Bereichs bestraft, sondern es wird bereits die Annäherung an den Rand des zulässigen Bereichs bestraft (den Rand des zulässigen Bereichs kann man sich als unüberwindliche Barriere vorstellen).
- (d) **Verfahren für Komplementaritätsprobleme:** Diese Verfahren, zu denen semiglatte Newtonverfahren oder Variationsmethoden gehören, versuchen, die KKT-Bedingungen direkt zu lösen. Dazu werden die Komplementaritätsprobleme entweder als Variationsungleichung umgeschrieben und mit geeigneten Verfahren gelöst, oder sie werden mithilfe von speziellen Funktionen als Gleichung reformuliert. Letzteres führt auf ein nichtdifferenzierbares Gleichungssystem, auf das Varianten des Newtonverfahrens angewendet werden können.

Jede dieser Verfahrensklassen verwendet zusätzlich Strategien, um Konvergenz von beliebigen Startschätzungen zu erreichen (Globalisierungsstrategien). Die üblichen Strategien sind

- eindimensionale Liniensuche (z.B. Armijo-Verfahren)
- Trust-Region-Verfahren (Approximation auf einem Vertrauensbereich)
- Filterverfahren (versuchen, nicht-dominierte Iterierte zu erzeugen, wobei Zielfunktionswert und Verletzung der Restriktionen als zwei Kriterien mitgeführt werden)

## 5.1 Notwendige Bedingungen für Standard-Optimierungsprobleme mit Gleichungsrestriktionen

Wir betrachten zunächst das Standardoptimierungsproblem ohne die Ungleichungsrestriktionen  $g_i(x) \leq 0$ ,  $i = 1, \dots, m$ , also

**Problem 5.1.1** (Nichtlineares Optimierungsproblem mit Gleichungsrestriktionen)

Finde  $x \in \mathbb{R}^n$ , so dass  $f(x)$  minimal wird unter den Nebenbedingungen

$$h_j(x) = 0, \quad j = 1, \dots, p.$$



Da wir uns bereits mit notwendigen Bedingungen für unrestringierte Optimierungsprobleme beschäftigt haben, möchten wir das restringierte Optimierungsproblem auf ein äquivalentes unrestringiertes Problem zurückführen, um dann die bekannten notwendigen Bedingungen anzuwenden. Dazu eliminieren wir die Gleichungsrestriktionen, indem wir nach einigen Komponenten des Vektors  $x$  auflösen. Diese Vorgehensweise haben wir bereits in einer Übungsaufgabe (Konservendose) durchgeführt und verallgemeinern sie nun.

Zunächst lösen wir die Gleichungsrestriktionen  $h(x) = 0$  nach  $p$  Komponenten des Vektors  $x = (x_1, \dots, x_n)^\top$  auf. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass wir nach den letzten  $p$  Komponenten von  $x$  auflösen können und partitionieren dazu den Vektor  $x$  in  $x = (y, z)^\top \in \mathbb{R}^n$  mit  $y \in \mathbb{R}^{n-p}$  und  $z \in \mathbb{R}^p$  und erhalten

$$z = z(y) \quad \text{mit} \quad h(y, z(y)) = 0.$$

Formal ist ein solches Auflösen nur dann möglich, wenn eine zusätzliche Regularitätsbedingung gilt: Die Jacobimatrix

$$\frac{\partial h}{\partial z}(y, z)$$

sei **invertierbar**. Der **Satz über implizite Funktionen**, den wir hier nur erwähnen, garantiert dann, dass man zumindest lokal eindeutig auflösen kann.

Ersetzen der Komponente  $z$  in der Zielfunktion  $f(x) = f(y, z)$  liefert die reduzierte Zielfunktion

$$F(y) := f(y, z(y)),$$

die nur noch von  $y$  abhängt. Die notwendige Bedingung erster Ordnung in einem lokalen Minimum  $\hat{x} = (\hat{y}, \hat{z})^\top$  führt dann mit Hilfe der Kettenregel und dem Satz über implizite Funktionen auf die Beziehung

$$\begin{aligned} 0 &= \nabla F(\hat{y})^\top \\ &= \frac{\partial f}{\partial y}(\hat{y}, \hat{z}) + \frac{\partial f}{\partial z}(\hat{y}, \hat{z}) \cdot \frac{\partial z}{\partial y}(\hat{y}) \\ &= \frac{\partial f}{\partial y}(\hat{y}, \hat{z}) - \frac{\partial f}{\partial z}(\hat{y}, \hat{z}) \cdot \left( \frac{\partial h}{\partial z}(\hat{y}, \hat{z}) \right)^{-1} \cdot \frac{\partial h}{\partial y}(\hat{y}, \hat{z}). \end{aligned} \quad (5.1)$$

Definiert man den **Lagrange-Multiplikator**  $\mu_0$  durch

$$\mu_0^\top := -\frac{\partial f}{\partial z}(\hat{y}, \hat{z}) \cdot \left( \frac{\partial h}{\partial z}(\hat{y}, \hat{z}) \right)^{-1} \quad (5.2)$$

und die **Lagrange-Funktion**  $L$  durch

$$L(x, \mu) := f(x) + \mu^\top h(x) \quad \text{mit} \quad x = \begin{pmatrix} y \\ z \end{pmatrix},$$

so folgt aus (5.1) und (5.2) die Beziehung

$$\nabla_x L(\hat{x}, \mu_0) = 0,$$

wobei  $\nabla_x$  den Gradienten bezüglich des Vektors  $x$  bezeichnet. Beachte, dass  $\mu_0$  unter den Voraussetzungen des Satzes über implizite Funktionen existiert. Damit ist folgender Satz bewiesen:

**Satz 5.1.2 (Lagrange'sche Multiplikatorenregel)**

Seien  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $h = (h_1, \dots, h_p)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^p$  stetig differenzierbar.  $f$  besitze in  $\hat{x} \in \mathbb{R}^n$  ein lokales Minimum unter den Nebenbedingungen  $h(x) = 0$ . Es gelte

$$\text{Rang}(h'(\hat{x})) = p,$$

wobei  $h'(\hat{x})$  die Jacobimatrix von  $h$  in  $\hat{x}$  bezeichnet.

Dann gibt es einen Lagrange-Multiplikator  $\mu_0 = (\mu_{0,1}, \dots, \mu_{0,p})^\top \in \mathbb{R}^p$  mit

$$0 = \nabla_x L(\hat{x}, \mu_0) = \nabla f(\hat{x}) + J_h(\hat{x})^\top \mu_0 = \nabla f(\hat{x}) + \sum_{j=1}^p \mu_{0,j} \nabla h_j(\hat{x}).$$

**Beispiel 5.1.3**

Gegeben sei das nichtlineare Optimierungsproblem

$$\text{Minimiere } f(x_1, x_2) = x_1 + x_2 \quad \text{unter } h(x_1, x_2) = x_1^2 - x_2 = 0.$$

Die Lagrangefunktion lautet

$$L(x_1, x_2, \mu) = x_1 + x_2 + \mu (x_1^2 - x_2).$$

Es gilt

$$\nabla f(x_1, x_2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla h(x_1, x_2) = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}, \quad \nabla_x L(x_1, x_2, \mu) = \begin{pmatrix} 1 + 2\mu x_1 \\ 1 - \mu \end{pmatrix}.$$

Da  $\nabla h(x_1, x_2) \neq 0$  für alle  $x_1, x_2$  gilt, ist die Rangbedingung der Lagrange'schen Multiplikatorenregel erfüllt und in einem lokalen Minimum  $\hat{x} = (\hat{x}_1, \hat{x}_2)^\top$  gibt es einen Multiplikator  $\mu_0$  mit

$$\begin{pmatrix} 1 + 2\mu_0 \hat{x}_1 \\ 1 - \mu_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dies ist ein nichtlineares Gleichungssystem für  $\hat{x}_1$  und  $\mu_0$ , welches die eindeutige Lösung  $\hat{x}_1 = -\frac{1}{2}$ ,  $\mu_0 = 1$  besitzt.

## 5.1. NOTWENDIGE BEDINGUNGEN FÜR STANDARD-OPTIMIERUNGSPROBLEME MIT GLEICHUNGSRESTRIKTIONEN 87

Einsetzen von  $\hat{x}_1$  in die Gleichungsrestriktion liefert  $0 = \frac{1}{4} - \hat{x}_2$  und somit  $\hat{x}_2 = \frac{1}{4}$ . ■

Das nichtlineare Gleichungssystem

$$\nabla_x L(x, \mu) = 0, \quad h(x) = 0$$

für  $x$  und  $\mu$  kann numerisch mit dem Newtonverfahren gelöst werden:

### Algorithmus 5.1.4 (Lagrange-Newton-Verfahren)

(i) Wähle Startschätzungen  $x^{[0]} \in \mathbb{R}^n$  und  $\mu^{[0]} \in \mathbb{R}^p$ ,  $\varepsilon > 0$  und setze  $k = 0$ .

(ii) Falls  $\|F(x^{[k]}, \mu^{[k]})\| \leq \varepsilon$ , STOP.

(iii) Löse das lineare Gleichungssystem

$$\begin{pmatrix} \nabla_{xx}^2 L(x^{[k]}, \mu^{[k]}) & h'(x^{[k]})^\top \\ h'(x^{[k]}) & 0 \end{pmatrix} \begin{pmatrix} d \\ v \end{pmatrix} = - \begin{pmatrix} \nabla_x L(x^{[k]}, \mu^{[k]}) \\ h(x^{[k]}) \end{pmatrix}$$

und setze

$$x^{[k+1]} = x^{[k]} + d, \quad \mu^{[k+1]} = \mu^{[k]} + v.$$

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

### Beispiel 5.1.5

$$\begin{aligned} \text{Minimiere} \quad & 2x_1^4 + x_2^4 + 4x_1^2 - x_1x_2 + 6x_2^2 \\ \text{unter} \quad & 2x_1 - x_2 = -4, \quad x_1, x_2 \in \mathbb{R}. \end{aligned}$$

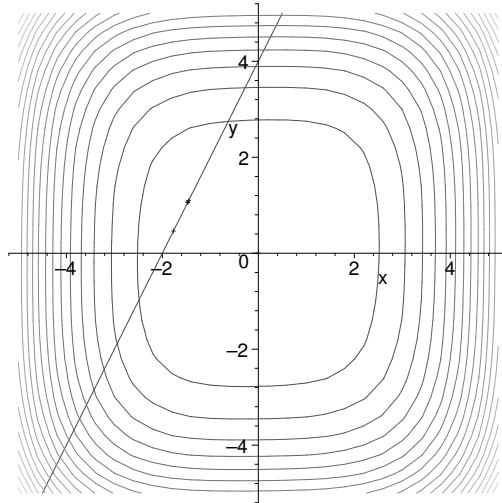
Der Gradient von  $f$  ist

$$\nabla f(x) = \begin{pmatrix} 8x_1^3 + 8x_1 - x_2 \\ 4x_2^3 - x_1 + 12x_2 \end{pmatrix}.$$

Die Hessematrix von  $f$  lautet

$$H_f(x) = \begin{pmatrix} 24x_1^2 + 8 & -1 \\ -1 & 12x_2^2 + 12 \end{pmatrix}.$$

$H_f$  ist symmetrisch und diagonaldominant und damit positiv definit. Der Rang von  $A$  ist 1. Damit ist das Newtonverfahren wohldefiniert.



Ausgabe des Lagrange-Newton-Verfahrens:

```

ITERATION 0
ZIELFUNKTION = 0.0000000000000000E+00
BESCHRAENKUNG = 0.4000000000000000E+01
NORM KKT = 0.0000000000000000E+00
X= 0.0000000000000000E+00 0.0000000000000000E+00
LAMBDA= 0.0000000000000000E+00

ITERATION 1
ZIELFUNKTION = 0.3425678372606001E+02
BESCHRAENKUNG = 0.0000000000000000E+00
NORM KKT = 0.4430579634007108E+02
X= -0.1769230769230769E+01 0.4615384615384616E+00
LAMBDA= 0.7307692307692307E+01

ITERATION 2
ZIELFUNKTION = 0.2756373273045131E+02
BESCHRAENKUNG = 0.0000000000000000E+00
NORM KKT = 0.5155005942054521E+01
X= -0.1452391998833495E+01 0.1095216002333011E+01
LAMBDA= 0.1660806234685793E+02

ITERATION 3
ZIELFUNKTION = 0.2754452288430214E+02
BESCHRAENKUNG = 0.0000000000000000E+00
NORM KKT = 0.1497798610217076E-01
X= -0.1467843643905216E+01 0.1064312712189567E+01
LAMBDA= 0.1904961295898514E+02

ITERATION 4
ZIELFUNKTION = 0.2754452202163215E+02
BESCHRAENKUNG = 0.0000000000000000E+00
NORM KKT = 0.6772437517980240E-06
X= -0.1467948113419141E+01 0.1064103773161718E+01
LAMBDA= 0.1905680332151563E+02

ITERATION 5
ZIELFUNKTION = 0.2754452202163215E+02
BESCHRAENKUNG = 0.0000000000000000E+00
NORM KKT = 0.3552713678800501E-14
X= -0.1467948118040920E+01 0.1064103763918160E+01
LAMBDA= 0.1905680364713604E+02

```

■

## 5.2 Notwendige Bedingungen für Standard-Optimierungsprobleme

Mithilfe des Satzes über implizite Funktionen konnten wir also notwendige Optimalitätsbedingungen für gleichungsrestringierte Optimierungsprobleme herleiten. Treten zusätzlich auch Ungleichungsrestriktionen auf, so kann die obige Technik des lokalen Auflörens nicht mehr ohne Weiteres angewendet werden, weil man die aktiven Beschränkungen nicht

a priori kennt.

### Definition 5.2.1

Sei  $x$  zulässig für das Standard-Optimierungsproblem. Die Beschränkung  $g_i(x) \leq 0$  heißt **aktiv in  $x$** , wenn  $g_i(x) = 0$  gilt. Sie heißt **inaktiv in  $x \in \mathbb{R}^n$** , wenn  $g_i(x) < 0$  gilt. Die Menge

$$A(x) := \{i \mid g_i(x) = 0, 1 \leq i \leq m\}$$

heißt **Indexmenge der (in  $x$ ) aktiven Ungleichungsrestriktionen**.

Man benötigt andere Beweistechniken. Anstatt hier einen formalen Beweis der nachfolgenden KKT-Bedingungen für Standard-Optimierungsprobleme mit Ungleichungsrestriktionen zu führen, begnügen wir uns mit einer geometrischen Motivation der KKT-Bedingungen<sup>1</sup> Betrachte dazu Abbildung 5.1.

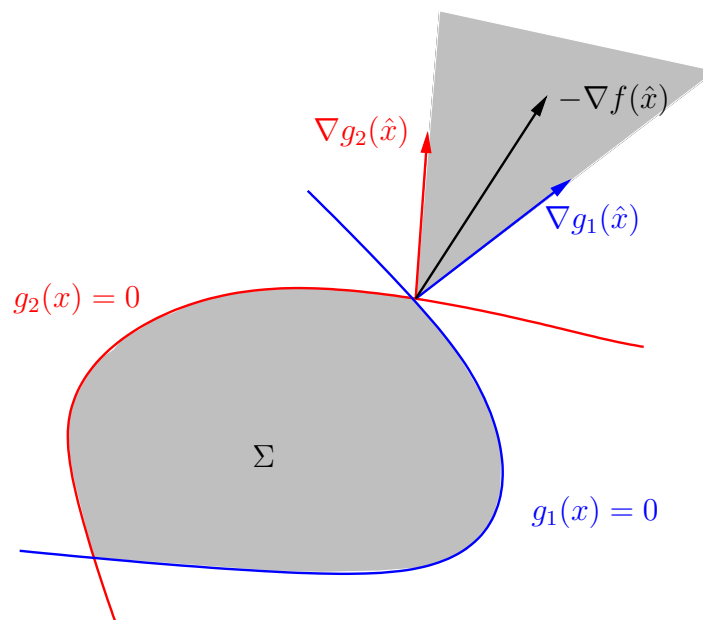


Abbildung 5.1: Illustration der KKT-Bedingungen: Falls in einem lokalen Minimum  $\hat{x}$  die Gradienten der aktiven (Un-)Gleichungsrestriktionen linear unabhängig sind, dann kann der negative Gradient der Zielfunktion als nicht-negative Linearkombination der Gradienten der aktiven Beschränkungen dargestellt werden.

Damit  $f$  in Abbildung 5.1 ein lokales Minimum im Schnittpunkt  $\hat{x}$  der beiden Ungleichungsrestriktionen  $g_1(x) \leq 0$  und  $g_2(x) \leq 0$  besitzt, muss der negative Gradient  $-\nabla f(\hat{x})$

<sup>1</sup>Ein formaler Beweis kann mithilfe des sogenannten Tangentialkegels und des linearisierenden Kegels geführt werden, wenn noch eine Constraint Qualification erfüllt ist.

in dem durch die (linear unabhängigen) Gradienten  $\nabla g_1(\hat{x})$  und  $\nabla g_2(\hat{x})$  aufgespannten Kegel (grau schattiert) liegen, d.h. es muss Multiplikatoren  $\lambda_1 \geq 0$  und  $\lambda_2 \geq 0$  geben, so dass

$$-\nabla f(\hat{x}) = \lambda_1 \nabla g_1(\hat{x}) + \lambda_2 \nabla g_2(\hat{x})$$

gilt. Wäre diese Bedingung nicht erfüllt, so gäbe es eine zulässige Abstiegsrichtung  $d$  mit  $\nabla f(\hat{x})^\top d < 0$  (dies kann man sich an der Skizze klarmachen!). Der negative Gradient von  $f$  kann in einem Optimum also als **Linearkombination der Gradienten der aktiven (Ungleichungs-)Nebenbedingungen** dargestellt werden! Formuliert man obige Gleichung um und verwendet die Lagrangefunktion für diesen Spezialfall mit 2 Ungleichungen, so erhält man

$$0 = \nabla f(\hat{x}) + \lambda_1 \nabla g_1(\hat{x}) + \lambda_2 \nabla g_2(\hat{x}) = \nabla_x L(\hat{x}, \lambda)$$

mit  $\lambda_1, \lambda_2 \geq 0$  für  $g_1(\hat{x}) = g_2(\hat{x}) = 0$ .

Dies gilt auch in allgemeiner Form für Standard-Optimierungsprobleme:

### Satz 5.2.2 (Notwendige Bedingungen erster Ordnung, KKT-Bedingungen)

Voraussetzungen:

- $\hat{x}$  ist ein lokales Minimum des Standard-Optimierungsproblems.
- Die Funktionen  $f, g_i, i = 1, \dots, m$ , und  $h_j, j = 1, \dots, p$ , sind stetig differenzierbar.
- Es gilt die Linear Independence Constraint Qualification (LICQ) in  $\hat{x}$ , d.h. die Vektoren

$$\nabla g_i(\hat{x}), i \in \mathcal{A}(\hat{x}) \quad \text{und} \quad \nabla h_j(\hat{x}), j = 1, \dots, p,$$

sind linear unabhängig.

Dann existieren eindeutig bestimmte Multiplikatoren  $\lambda = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$  und  $\mu = (\mu_1, \dots, \mu_p)^\top \in \mathbb{R}^p$ , so dass die folgenden Bedingungen gelten:

(a) **Stationarität der Lagrange-Funktion:**

$$\nabla_x L(\hat{x}, \lambda, \mu) = 0 \tag{5.3}$$

bzw.

$$\nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) + \sum_{j=1}^p \mu_j \nabla h_j(\hat{x}) = 0. \tag{5.4}$$

(c) **Komplementaritätsbedingungen:** Für  $i = 1, \dots, m$  gilt:

$$\lambda_i \geq 0 \quad \text{und} \quad \lambda_i g_i(\hat{x}) = 0. \tag{5.5}$$

(d) **Zulässigkeit:**

$$g(\hat{x}) \leq 0 \quad \text{und} \quad h(\hat{x}) = 0. \quad (5.6)$$

Jeden Punkt  $(\hat{x}, \lambda, \mu)$ , der die Bedingungen (5.3)-(5.6) erfüllt, nennen wir **KKT-Punkt** oder **stationären Punkt**. Beachte, dass KKT-Punkte lediglich Kandidaten für optimale Lösungen liefern.

**Bemerkung 5.2.3**

Die LICQ ist nur eine von vielen sogenannten Constraint Qualifications (oder Regularitätsbedingungen) an den zulässigen Bereich. Die üblichen alternativen Constraint Qualifications (z.B. von Mangasarian-Fromowitz, Robinson, Abadie, Guignard) garantieren zwar auch die Gültigkeit der KKT-Bedingungen, allerdings liefern sie allesamt nicht die Eindeutigkeit der Multiplikatoren  $\lambda$  und  $\mu$ . Zu beachten ist jedoch, dass die KKT-Bedingungen i.A. nicht mehr gelten, wenn überhaupt keine Constraint Qualification erfüllt ist (in diesem Fall gelten nur noch die sogenannten Fritz John Bedingungen). ■

**Beispiel 5.2.4**

Gegeben sei das nichtlineare Optimierungsproblem

$$\min_{x \in \mathbb{R}^2} f(x) = x_1 \quad \text{u.d.N.} \quad g(x) = (x_1 - 4)^2 + x_2^2 - 16 \leq 0.$$

Die Lagrangefunktion lautet

$$L(x_1, x_2, \lambda) = x_1 + \lambda ((x_1 - 4)^2 + x_2^2 - 16).$$

Es gilt

$$\nabla f(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \nabla g(x_1, x_2) = \begin{pmatrix} 2(x_1 - 4) \\ 2x_2 \end{pmatrix}, \quad \nabla_x L(x_1, x_2, \lambda) = \begin{pmatrix} 1 + 2\lambda(x_1 - 4) \\ 2\lambda x_2 \end{pmatrix}.$$

Beachte, dass die LICQ erfüllt ist, da  $\nabla g \neq 0$  für alle  $(x_1, x_2) \neq (4, 0)$  gilt, wobei die Beschränkung im Punkt  $(4, 0)$  wegen  $g(4, 0) = -16 < 0$  nicht aktiv ist.

Die KKT-Bedingungen lauten wie folgt:

Sei  $(x_1, x_2)^\top$  ein lokales Minimum des nichtlinearen Optimierungsproblems. Dann gibt es einen Multiplikator  $\lambda$  mit

$$\begin{pmatrix} 1 + 2\lambda(x_1 - 4) \\ 2\lambda x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5.7)$$

und den Komplementaritätsbedingungen

$$\lambda((x_1 - 4)^2 + x_2^2 - 16) = 0, \quad \lambda \geq 0, \quad (x_1 - 4)^2 + x_2^2 - 16 \leq 0.$$

Wir versuchen, einen KKT-Punkt zu bestimmen. Dazu unterscheiden wir zwei Fälle:

(i) Sei  $\lambda > 0$ . Dann muss  $x_2 = 0$  gelten. Aus den Komplementaritätsbedingungen folgt

$$0 = (x_1 - 4)^2 + x_2^2 - 16 = (x_1 - 4)^2 - 16$$

und aus der ersten Gleichung in (5.7) folgt

$$0 = 1 + 2\lambda(x_1 - 4).$$

Dies ist ein nichtlineares Gleichungssystem für  $x_1$  und  $\lambda$ .

Aus der ersten Gleichung folgt sofort entweder  $x_1 = 0$  oder  $x_1 = 8$ .

Aus der zweiten folgt dann  $\lambda = 1/8 > 0$  (falls  $x_1 = 0$ ) bzw.  $\lambda = -1/8 < 0$  (falls  $x_1 = 8$ ).

Der Punkt  $x_1 = 8, x_2 = 0, \lambda = -1/8$  erfüllt die Komplementaritätsbedingung nicht, ist daher kein KKT-Punkt.

Der Punkt  $x_1 = 0, x_2 = 0, \lambda = 1/8$  erfüllt die Komplementaritätsbedingung und die Stationaritätsbedingung, ist daher ein KKT-Punkt (sogar der einzige wegen (ii)).

(ii) Der Fall  $\lambda = 0$  kann nicht eintreten, da die erste Gleichung in (5.7) dann auf den Widerspruch  $1 = 0$  führt.

■

### 5.3 Penalty-Verfahren

Penaltyverfahren sind beliebte Verfahren, die auf der Ankopplung der Nebenbedingungen an die Zielfunktion mithilfe eines gewichteten Strafterms (Penalty-Term) basieren. Der Strafterm bestraft unzulässige Punkte. Der Vorteil der Verfahren ist, dass durch die Ankopplung der Nebenbedingungen an die Zielfunktion Nebenbedingungen eliminiert werden, so dass Verfahren der unrestringierten Optimierung angewendet werden können. Wir illustrieren die Idee der Penalty-Verfahren an einem Beispiel.

#### Beispiel 5.3.1

Gegeben sei das folgende Optimierungsproblem:

$$\text{Minimiere} \quad f(x_1, x_2) := x_1 + x_2 \quad \text{u.d.N.} \quad h(x_1, x_2) := x_1^2 - x_2 = 0.$$

Wir möchten nun die Nebenbedingung eliminieren. Wir könnten dies erreichen durch Auflösen nach  $x_2$  und Einsetzen in die Zielfunktion. Hier verfolgen wir aber eine andere Strategie und koppeln einen Strafterm, der Punkte mit  $x_1^2 - x_2 \neq 0$  bestraft, an die Zielfunktion. Ein solcher Strafterm kann z.B. mithilfe der Funktion

$$r(x_1, x_2) := (x_1^2 - x_2)^2 = h(x_1, x_2)^2$$



realisiert werden, denn es gilt  $r(x_1, x_2) = 0$  genau dann, wenn  $h(x_1, x_2) = x_1^2 - x_2 = 0$  gilt. Beachte, dass  $r$  differenzierbar ist. Man hätte auch  $|h(x_1, x_2)|$  verwenden können, aber diese Funktion ist nicht differenzierbar.

Anstelle von  $f$  minimiert man dann die sogenannte Penalty-Funktion

$$P(x_1, x_2; \eta) := f(x_1, x_2) + \frac{\eta}{2}r(x_1, x_2) = x_1 + x_2 + \frac{\eta}{2}(x_1^2 - x_2)^2,$$

wobei  $\eta > 0$  ein Gewichtungsfaktor ist.

Zur Minimierung von  $P$  können Verfahren der unrestringierten Optimierung verwendet werden. Allerdings hängt die Lösung vom Gewichtungsfaktor  $\eta$  ab und es stellt sich die Frage, unter welchen Bedingungen diese Lösung gegen die Lösung des restringierten Ausgangsproblems konvergieren.

Dazu untersuchen wir die Minima von  $P$  in Abhängigkeit von  $\eta$ . Notwendig gilt

$$0 = \nabla_x P(x_1, x_2; \eta) = \begin{pmatrix} 1 + 2\eta x_1(x_1^2 - x_2) \\ 1 - \eta(x_1^2 - x_2) \end{pmatrix}.$$

Daraus ergeben sich die stationären Punkte

$$\begin{pmatrix} x_1(\eta) \\ x_2(\eta) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{4} - \frac{1}{\eta} \end{pmatrix}.$$

Wie hängen diese Lösungen mit der Lösung des Ausgangsproblems zusammen? Dazu bestimmen wir die stationären Punkte der Lagrange-Funktion

$$L(x_1, x_2, \mu) = x_1 + x_2 + \mu(x_1^2 - x_2)$$

und erhalten

$$0 = \nabla_x L(\hat{x}_1, \hat{x}_2, \mu) = \begin{pmatrix} 1 + 2\mu\hat{x}_1 \\ 1 - \mu \end{pmatrix} \iff \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{4} \end{pmatrix}, \mu = 1.$$

Man erkennt, dass die Lösungen des Penalty-Problems für  $\eta \rightarrow \infty$  gegen die Lösung des restringierten Problems konvergieren:

$$\lim_{\eta \rightarrow \infty} \begin{pmatrix} x_1(\eta) \\ x_2(\eta) \end{pmatrix} = \lim_{\eta \rightarrow \infty} \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{4} - \frac{1}{\eta} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}.$$

Man muss die Gewichtungparameter also so wählen, dass sie gegen  $\infty$  gehen. ■

Die Beobachtungen in Beispiel 5.3.1 gelten auch allgemein unter relativ schwachen Voraussetzungen. Dazu betrachten wir zunächst das folgende gleichungsrestringierte Optimierungsproblem mit stetigen (!) Funktionen  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $h_j : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $j = 1, \dots, p$ .

**Problem 5.3.2 (Restringiertes Optimierungsproblem)**

$$\text{Minimiere } f(x) \quad \text{u.d.N. } x \in \Sigma$$

mit

$$\Sigma = \{x \in \mathbb{R}^n \mid h_j(x) = 0, j = 1, \dots, p\}.$$

Die Idee des Penalty-Verfahrens besteht darin, die Lösung  $\hat{x}$  des Ausgangsproblems iterativ durch die Lösungen von unrestringierten Hilfsproblemen zu approximieren. Diese Hilfsprobleme bestehen in der Minimierung der **Penalty-Funktion**

$$P(x; \eta) := f(x) + \frac{\eta}{2} \sum_{j=1}^p (h_j(x))^2$$

für geeignete Werte von  $\eta > 0$ . Durch die Ankopplung der Nebenbedingungen wird ein Verlassen des zulässigen Bereichs  $\Sigma$  „bestraft“. Die Konstante  $\eta$  stellt einen Gewichtungsfaktor dar, mit dessen Hilfe die „Stärke der Bestrafung“ gesteuert werden kann. Das Penalty-Verfahren ist durch folgenden Algorithmus gegeben.

**Algorithmus 5.3.3 (Penalty-Verfahren)**

(i) Wähle  $\eta_0 > 0$  und setze  $k = 0$ .

(ii) Bestimme  $x^{[k]}$  als Lösung von

$$\text{Minimiere } P(x; \eta_k) \quad \text{u.d.N. } x \in \mathbb{R}^n .$$

(iii) Ist  $h(x^{[k]}) \approx 0$ , STOP.

(iv) Bestimme  $\eta_{k+1} > \eta_k$ , setze  $k := k + 1$  und gehe zu (ii).

Da  $P$  i.A. nicht differenzierbar ist, werden für das Hilfsproblem in Schritt (ii) Verfahren der unrestringierten, nichtdifferenzierbaren Optimierung benötigt. Es stellt sich natürlich die Frage, ob das Verfahren tatsächlich gegen eine Lösung des Ausgangsproblems konvergiert.

**Satz 5.3.4 (Konvergenzsatz für das Penalty-Verfahren)**

Seien  $f$  und  $h_j$ ,  $j = 1, \dots, p$  stetig und  $\{\eta_k\}$  streng monoton wachsend mit  $\eta_k \rightarrow \infty$ . Die zulässige Menge  $\Sigma := \{x \in \mathbb{R}^n \mid h_j(x) = 0, j = 1, \dots, p\}$  sei nichtleer, und  $\{x^{[k]}\}$  sei eine durch Algorithmus 5.3.3 erzeugte Folge (die Existenz der Folge sei vorausgesetzt). Dann gelten die folgenden Aussagen:

- (a) Die Folge der Zielfunktionswerte der Penalty-Funktion  $\{P(x^{[k]}; \eta_k)\}_{k \in \mathbb{N}}$  ist monoton wachsend.
- (b) Die Folge der Verletzung der Nebenbedingungen  $\{\|h(x^{[k]})\|\}_{k \in \mathbb{N}}$  ist monoton fallend.
- (c) Die Folge der Zielfunktionswerte  $\{f(x^{[k]})\}_{k \in \mathbb{N}}$  ist monoton wachsend.
- (d) Es gilt  $\lim_{k \rightarrow \infty} h(x^{[k]}) = 0$ .
- (e) Jeder Häufungspunkt der Folge  $\{x^{[k]}\}_{k \in \mathbb{N}}$  ist eine Lösung des Ausgangsproblems.

**Beweis:** (vgl. [GK02], S. 208)

**Bemerkung 5.3.5**

Da nur die Stetigkeit der auftretenden Funktionen benötigt wird, ist das Verfahren auch auf Problemstellungen mit Ungleichungsnebenbedingungen

$$g_i(x) \leq 0, \quad i = 1, \dots, m,$$

anwendbar. Denn diese Nebenbedingungen können äquivalent als stetige Nebenbedingungen

$$\max\{0, g_i(x)\} = 0, \quad i = 1, \dots, m,$$

geschrieben werden. Die Penaltyfunktion lautet dann

$$P(x; \eta) = f(x) + \frac{\eta}{2} \sum_{j=1}^p (h_j(x))^2 + \frac{\eta}{2} \sum_{i=1}^m (\max\{0, g_i(x)\})^2.$$

■

Ein wesentlicher Nachteil des Penalty-Verfahrens ist die Tatsache, dass die Gewichtungsfaktoren  $\eta_k$  gegen  $\infty$  streben müssen, um Konvergenz zu erhalten. Dies führt dazu, dass die Teilprobleme in (ii) des Algorithmus für großes  $\eta_k$  sehr schlecht konditioniert sind<sup>2</sup> und numerisch nur sehr schwer zu lösen sind.

<sup>2</sup>Einige Eigenwerte der Hessematrix  $\nabla_{xx}^2 P(x^{[k]}; \eta_k)$  streben gegen  $\infty$  und somit strebt die Spektralkonditionszahl der Hessematrix gegen unendlich.

## 5.4 Multiplikator-Penalty-Verfahren

Multiplikator-Penalty-Verfahren ähneln den Penalty-Verfahren. Allerdings arbeiten sie mit einer exakten und differenzierbaren Penalty-Funktion – der erweiterten Lagrange-funktion.

Wir betrachten wieder das gleichungsrestringierte Problem 5.3.2, d.h.

$$\text{Minimiere} \quad f(x) \quad \text{u.d.N.} \quad h(x) = 0. \quad (5.8)$$

Darin seien die Funktionen  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $h = (h_1, \dots, h_p)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^p$  zweimal stetig differenzierbar.

Sei  $\hat{x}$  lokales Minimum des Problems. Dann ist  $\hat{x}$  für  $\eta > 0$  auch ein lokales Minimum von

$$\text{Minimiere} \quad f(x) + \frac{\eta}{2} \|h(x)\|^2 \quad \text{u.d.N.} \quad h(x) = 0.$$

Die Lagrangefunktion für dieses Problem lautet

$$L_a(x, \mu; \eta) := f(x) + \frac{\eta}{2} \|h(x)\|^2 + \mu^\top h(x)$$

und heißt **erweiterte Lagrangefunktion (augmented Lagrangian)** oder **Multiplikator-Penalty-Funktion**.

Man versucht nun, das Ausgangsproblem (5.8) indirekt zu lösen, indem die erweiterte Lagrangefunktion minimiert wird:

$$\text{Minimiere} \quad L_a(x, \hat{\mu}; \eta) \quad \text{u.d.N.} \quad x \in \mathbb{R}^n.$$

Der Penalty-Parameter  $\eta$  muß jetzt, anders als bei den Penalty-Verfahren, nicht mehr gegen  $\infty$  streben. Darüber hinaus ist  $L_a$  differenzierbar, so dass bekannte Verfahren aus der unrestringierten Optimierung eingesetzt werden können.

**Problem:** Der optimale Lagrangemultiplikator  $\hat{\mu}$  ist unbekannt.

Wir versuchen nun,  $\hat{\mu}$  geeignet zu approximieren. Sei  $\eta$  hinreichend groß und  $x^{[k+1]}$  stationärer Punkt des Problems

$$\text{Minimiere} \quad L_a(x, \mu^{[k]}; \eta) \quad \text{u.d.N.} \quad x \in \mathbb{R}^n.$$

Dann gilt notwendig

$$0 = \nabla_x L_a(x^{[k+1]}, \mu^{[k]}; \eta) = \nabla f(x^{[k+1]}) + \sum_{j=1}^p \left( \mu_j^{[k]} + \eta h_j(x^{[k+1]}) \right) \nabla h_j(x^{[k+1]}).$$

Andererseits gilt in einem KKT-Punkt  $(\hat{x}, \hat{\mu})$  von (5.8) notwendig

$$0 = \nabla_x L(\hat{x}, \hat{\mu}) = \nabla f(\hat{x}) + \sum_{j=1}^p \hat{\mu}_j \nabla h_j(\hat{x}).$$

Ein Vergleich beider Ausdrücke liefert die naheliegende Aufdatierungsvorschrift

$$\mu^{[k+1]} := \mu^{[k]} + \eta h(x^{[k+1]}).$$

Insgesamt entsteht das Multiplier-Penalty-Verfahren:

#### Algorithmus 5.4.1 (Multiplikator-Penalty-Verfahren)

(i) Wähle  $x^{[0]} \in \mathbb{R}^n$ ,  $\mu^{[0]} \in \mathbb{R}^p$ ,  $\eta_0 > 0$ ,  $\sigma \in (0, 1)$  und setze  $k = 0$ .

(ii) Ist  $(x^{[k]}, \mu^{[k]})$  KKT-Punkt von (5.8), STOP.

(iii) Bestimme  $x^{[k+1]}$  als Lösung von

$$\text{Minimiere } L_a(x, \mu^{[k]}; \eta_k) \quad \text{u.d.N.} \quad x \in \mathbb{R}^n.$$

(iv) Setze  $\mu^{[k+1]} := \mu^{[k]} + \eta_k h(x^{[k+1]})$ .

(v) Ist  $\|h(x^{[k+1]})\| \geq \sigma \|h(x^{[k]})\|$ , so setze  $\eta_{k+1} := 10\eta_k$ . Andernfalls setze  $\eta_{k+1} := \eta_k$ .

(vi) Setze  $k := k + 1$  und gehe zu (ii).

#### Beispiel 5.4.2

Gegeben sei das nichtlineare Optimierungsproblem

$$\text{Minimiere } x^2 + y^2 \quad \text{u.d.N.} \quad x - y - 1 = 0.$$

Die erweiterte Lagrange-Funktion lautet

$$L_a(x, y, \mu, \eta) = x^2 + y^2 + \frac{\eta}{2}(x - y - 1)^2 + \mu(x - y - 1).$$

Anwendung der Multiplier-Penalty-Methode mit Startwerten

$$(x^{[0]}, y^{[0]}) = (0, 0), \quad \mu^{[0]} = 0, \quad \eta_0 = 1, \quad \sigma = 0.1$$

liefert das folgende Resultat:

K	X	Y	LAMBDA	ETA	F(X, Y)	H(X, Y)
0	0.00000E+00	0.00000E+00	0.00000E+00	0.10000E+01	0.00000E+00	0.10000E+01
1	0.25000E+00	-0.25000E+00	-0.50000E+00	0.10000E+02	0.12500E+00	0.50000E+00
2	0.47727E+00	-0.47727E+00	-0.95455E+00	0.10000E+02	0.45558E+00	0.45455E-01
3	0.49793E+00	-0.49793E+00	-0.99587E+00	0.10000E+02	0.49588E+00	0.41322E-02
4	0.49981E+00	-0.49981E+00	-0.99962E+00	0.10000E+02	0.49962E+00	0.37566E-03
5	0.49998E+00	-0.49998E+00	-0.99997E+00	0.10000E+02	0.49997E+00	0.34151E-04
6	0.50000E+00	-0.50000E+00	-0.10000E+01	0.10000E+02	0.50000E+00	0.31046E-05
7	0.50000E+00	-0.50000E+00	-0.10000E+01	0.10000E+02	0.50000E+00	0.28224E-06
8	0.50000E+00	-0.50000E+00	-0.10000E+01	0.10000E+02	0.50000E+00	0.25658E-07
9	0.50000E+00	-0.50000E+00	-0.10000E+01	0.10000E+02	0.50000E+00	0.23325E-08

■

# Kapitel A

## Zusatzmaterial

### A.1 Abbruchkriterien

Wir klären die Frage nach geeigneten Abbruchkriterien für das allgemeine Abstiegsverfahren. Dazu gibt der Benutzer eine relative Genauigkeitsschranke  $\varepsilon \approx 10^{-r}$  vor, wobei  $r$  die Anzahl der gewünschten gültigen Dezimalstellen der numerischen Lösung sei. Da numerische Rechnungen in der Regel rundungsfehlerbehaftet sind, ist es sinnlos,  $\varepsilon$  kleiner als die relative Maschinengenauigkeit  $\varepsilon_{mach}$  zu wählen. Für doppelt genaue Gleitpunktzahlen gilt  $\varepsilon_{mach} = 2^{-52} \approx 2.22 \cdot 10^{-16}$ .

Da die Bedingung  $\nabla f(\hat{x}) = 0$  notwendig für ein Minimum ist, andererseits in der numerischen Praxis aber so gut wie niemals von den Iterierten  $x^{[i]}$  erfüllt wird, ist

$$\|\nabla f(x^{[i]})\| \leq \varepsilon$$

eine sinnvolle Abbruchbedingung. Allerdings ist diese Bedingung nicht invariant bezüglich der Skalierung von  $f$ . Denn durch Multiplikation von  $f$  mit einer hinreichend kleinen positiven Zahl ist diese Bedingung nahezu immer erfüllbar. Andererseits ist es nicht sinnvoll, das Abstiegsverfahren fortzuführen, da auf Grund von Rundungsfehlern kein besseres Ergebnis zu erwarten ist. Daher sollte dieses Abbruchkriterium mit einer Warnung versehen werden. Ebenso ist es sinnvoll, eine maximale Iterationszahl  $i_{max}$  vorzuschreiben und das Abstiegsverfahren abzurechnen, sobald

$$i \geq i_{max}$$

gilt. Dies verhindert zu lange Iterationen und sollte ebenfalls mit einer Warnung versehen werden.

Gill et al. [GMW81] schlagen folgende Abbruchkriterien vor, die nicht nur die absoluten Größen  $\nabla f(x^{[i]})$  bzw.  $\|x^{[i-1]} - x^{[i]}\|$  bzw.  $f(x^{[i-1]}) - f(x^{[i]})$  überprüfen, sondern diese durch zusätzliche Faktoren  $(1 + \dots)$  in Relation setzen zur Größe der Funktionswerte  $|f(x^{[i]})|$  bzw. der Iterierten  $\|x^{[i]}\|$ :

- $f(x^{[i-1]}) - f(x^{[i]}) \leq \varepsilon \cdot (1 + |f(x^{[i]})|)$
- $\|x^{[i-1]} - x^{[i]}\| \leq \sqrt{\varepsilon} \cdot (1 + \|x^{[i]}\|)$
- $\|\nabla f(x^{[i]})\| \leq \sqrt[3]{\varepsilon} \cdot (1 + |f(x^{[i]})|)$

Für sehr kleine Werte  $|f(x^{[i]})|$  bzw.  $\|x^{[i]}\|$  gehen diese relativen Abfragen in absolute Kriterien über. Sind diese Kriterien erfüllt, so kann das Verfahren mit Erfolg beendet werden. Die Wurzel  $\sqrt{\varepsilon}$  in der zweiten Bedingung erklärt sich durch Taylorentwicklung in der Nähe des Lösungspunktes (dort gilt  $\nabla f(x^{[i]}) \approx 0$ ):

$$f(x^{[i-1]}) \approx f(x^{[i]}) + \mathcal{O}(\|x^{[i-1]} - x^{[i]}\|^2).$$

Die dritte Wurzel in der dritten Bedingung stellt eine Abschwächung der theoretisch begründbaren zweiten Wurzel dar. Die Verwendung der zweiten Wurzel stellt sich in der Praxis als zu restriktiv heraus.

## A.2 Berechnung von Ableitungen

In der Regel sind die in Anwendungen auftretenden Funktionen zu komplex, um deren Ableitungen analytisch angeben zu können. Die folgenden Varianten können häufig alternativ genutzt werden:

- Programme wie MAPLE oder MATLAB (Symbolic Toolbox) erlauben es, Ableitungen **symbolisch** zu berechnen.

In MAPLE stehen hierfür der Befehl `diff` bzw. der Operator `D` zur Verfügung. Zur Berechnung des Gradienten bzw. der Hessematrix dienen die Befehle `grad` bzw. `hessian`.

Beispiele:

```
f := (x^2+y-11)^2 + (x+y^2-7)^2;
diff(f,x);    % partielle Ableitung nach x
diff(f,y);    % partielle Ableitung nach y
```

```
g := (x,y) -> (x^2+y-11)^2 + (x+y^2-7)^2;
D[1](g);      % partielle Ableitung nach x
D[2](g);      % partielle Ableitung nach y
```

```
with(linalg):
grad(f,vector([x,y]));
grad(g(x,y),vector([x,y]));
```

```
hessian(f,vector([x,y]));
hessian(g(x,y),vector([x,y]));
```



- Die Auswertung einer von  $x$  abhängigen Funktion liege in Form eines Fortran- oder C-Programms vor. Beim **automatischen Differenzieren** wird das komplette Programm (d.h. die Anweisungsfolge des Programms) algorithmisch differenziert. Man erhält wieder ein Fortran- oder C-Programm, welches dann z.B. den Gradienten von  $f$  abhängig von den Eingabedaten  $x$  liefert. Im wesentlichen basiert das automatische Differenzieren auf der Anwendung von Produkt- und Kettenregel, gekoppelt mit einem Parser, der Fortran- oder C-Programme interpretieren kann.

Weitere Hinweise und Software finden sich unter [www.autodiff.org](http://www.autodiff.org).

- Bei der **numerischen Differentiation** werden geeignete finite-Differenzen-Schemata zur approximativen Berechnung von Ableitungen verwendet, etwa

$$\frac{\partial f}{\partial x_i}(x) \approx \frac{f(x + he_i) - f(x)}{h},$$

wobei  $e_i$  den  $i$ -ten Einheitsvektor bezeichnet. Das folgende Fortran-Programm approximiert die Jacobimatrix einer Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ :

```

CALL F( X,C )
DO 10, J=1, N
  TEMP=X(J)
  H=MAX(EPS, EPS*DABS(TEMP))
  DEL=1.0D0/H
  X(J)=X(J)+H
  CALL F( X,W )
  X(J)=TEMP
  DO 20, I=1, M
    JAC(I, J)=(W(I)-C(I))*DEL
20  CONTINUE
10  CONTINUE

```

Es sei jedoch bemerkt, daß die Berechnung zweiter Ableitungen mittels finiter Differenzen schlecht konditioniert ist.

### A.3 Globalisierung des Newton-Verfahrens

Wie Satz 3.4.6 erkennen läßt, ist das lokale Newtonverfahren i.a. nicht global (d.h. für jeden Startpunkt  $x^{[0]}$ ) konvergent. Es gibt verschiedene Möglichkeiten, um das lokale Newton-Verfahren durch Zusatzmaßnahmen zu einem global konvergenten Verfahren zu erweitern. Um die quadratische Konvergenz des lokalen Newtonverfahrens zu sichern, hat

man dabei darauf zu achten, daß das globale Newtonverfahren in der Nähe der Lösung in das lokale Verfahren (also mit Schrittweite  $\alpha_i = 1$ ) übergeht.

Geiger und Kanzow [GK99] schlagen folgende globale Variante vor:

**Algorithmus 1.3.1** (globales Newtonverfahren)

(i) Wähle einen Startvektor  $x^{[0]} \in \mathbb{R}^n$ , Konstanten  $\sigma \in (0, 1/2)$ ,  $\beta \in (0, 1)$ ,  $\varrho > 0$ ,  $p > 2$  und setze  $i = 0$ .

(ii) Falls ein Abbruchkriterium erfüllt ist, STOP.

(iii) Berechne (falls möglich) die Suchrichtung  $d^{[i]}$  als Lösung von (3.4) mit  $x = x^{[i]}$ . Besitzt das Gleichungssystem keine Lösung oder ist die Bedingung

$$\nabla f(x^{[i]})^\top d^{[i]} \leq -\varrho \|d^{[i]}\|^p$$

nicht erfüllt, so setze  $d^{[i]} = -\nabla f(x^{[i]})$ .

(iv) Bestimme eine Schrittweite  $\alpha_i > 0$  mit dem Armijo-Verfahren.

(v) Setze  $x^{[i+1]} = x^{[i]} + \alpha_i d^{[i]}$ ,  $i := i + 1$  und gehe zu (ii).

■

Unter der Voraussetzung, daß  $f$  zweimal stetig differenzierbar ist, zeigen Geiger und Kanzow [GK99] auf S. 87, daß jeder Häufungspunkt der Folge  $\{x^{[i]}\}$  ein stationärer Punkt von  $f$  ist. Darüber hinaus läßt sich der folgende Konvergenzsatz beweisen, den wir hier ohne Beweis aus Geiger und Kanzow [GK99], S. 92 zitieren.

**Satz 1.3.2** (globale Konvergenz)

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und  $\{x^{[i]}\}$  eine durch das globale Newtonverfahren erzeugte Folge. Für einen Häufungspunkt  $\hat{x}$  der Folge mit positiv definiten Hessematrix  $H_f(\hat{x})$  gelten:

(a) Die gesamte Folge  $\{x^{[i]}\}$  konvergiert gegen  $\hat{x}$  und  $\hat{x}$  ist striktes lokales Minimum von  $f$ .

(b) Für hinreichend großes  $i \in \mathbb{N}$  ist  $d^{[i]}$  stets Newton-Richtung und die Schrittweite erfüllt  $\alpha_i = 1$ .

(c)  $\{x^{[i]}\}$  konvergiert superlinear gegen  $\hat{x}$ .

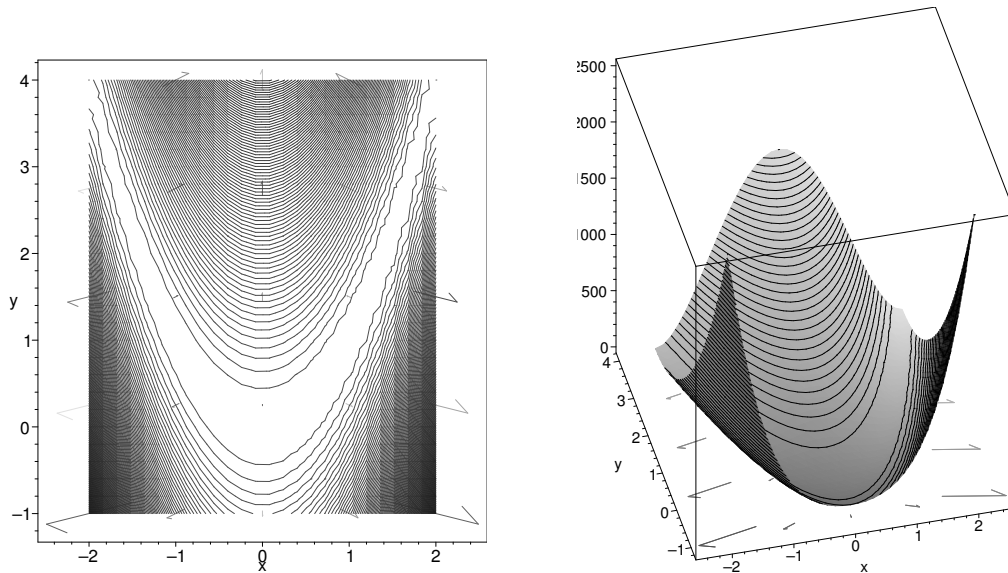
(d) Ist  $H_f$  sogar (lokal) Lipschitz-stetig, so konvergiert  $\{x^{[i]}\}$  quadratisch gegen  $\hat{x}$ .

**Beispiel 1.3.3** (Funktion von Rosenbrock)

Wir minimieren

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$$

mit dem globalen Newtonverfahren.  $f$  hat ein globales Minimum in  $(1, 1)^\top$  mit Funktionswert 0.



Die Hessematrix und deren Eigenwerte im Lösungspunkt lauten

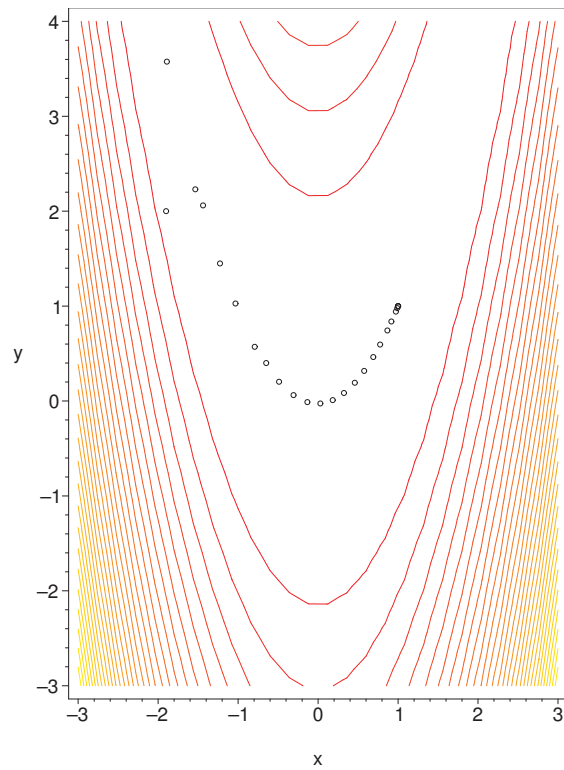
$$H_f(1, 1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}, \quad \lambda_1 \approx 1001.6, \lambda_2 \approx 0.3994.$$

Darüber hinaus ist die Funktion (als Polynom) beliebig oft differenzierbar. Alle Voraussetzungen des Konvergenzsatzes für das globale Newtonverfahren sind erfüllt. Das globale Newtonverfahren mit den Parametern  $\sigma = 0.01$ ,  $\beta = 0.5$ ,  $\varrho = 0.01$ ,  $p = 3$ ,  $\varepsilon = 10^{-13}$  und Abbruchkriterium  $\|\nabla f(x)\| \leq \varepsilon$  liefert für den Startpunkt  $x^{[0]} = (-1.9, 2)^\top$  das folgende Ergebnis:

ITER	X(1)	X(2)	GRAD	DX	P=1	P=2	ALPHA	F
0	-0.1900000E+01	0.2000000E+01	0.1270869E+04	0.0000000E+00	0.0000000E+00	0.0000000E+00	0.0000000E+00	0.2676200E+03
1	-0.1891022E+01	0.3575882E+01	0.5843040E+01	0.1575908E+01	0.1262269E+01	0.4114880E+00	0.1000000E+01	0.8358007E+01
2	-0.1535378E+01	0.2230831E+01	0.8657596E+02	0.1113020E+02	0.7278599E+00	0.1879754E+00	0.1250000E+00	0.8029711E+01
3	-0.1439014E+01	0.2061477E+01	0.1039035E+02	0.1948509E+02	0.9438102E+00	0.3348806E+00	0.1000000E+01	0.5957414E+01
4	-0.1225603E+01	0.1449594E+01	0.3196743E+02	0.2592127E+01	0.8535986E+00	0.3209034E+00	0.2500000E+00	0.5229027E+01
5	-0.1032102E+01	0.1027792E+01	0.2090905E+02	0.4640683E+00	0.8950620E+00	0.3942032E+00	0.1000000E+01	0.4268634E+01
6	-0.7927080E+00	0.5710765E+00	0.2459178E+02	0.5156536E+00	0.9070084E+00	0.4462983E+00	0.1000000E+01	0.3542240E+01
7	-0.6488530E+00	0.4003159E+00	0.9606086E+01	0.2232789E+00	0.9518327E+00	0.5163726E+00	0.1000000E+01	0.2761541E+01
8	-0.4884230E+00	0.2024721E+00	0.1235395E+02	0.5094308E+00	0.9624431E+00	0.5485510E+00	0.5000000E+00	0.2345615E+01
9	-0.3072831E+00	0.6161129E-01	0.9340990E+01	0.2294633E+00	0.9529718E+00	0.5643479E+00	0.1000000E+01	0.1816650E+01
10	-0.1344153E+00	-0.1181580E-01	0.7123217E+01	0.1878159E+00	0.9446169E+00	0.5870059E+00	0.1000000E+01	0.1376199E+01
11	0.2818625E-01	-0.2564481E-01	0.5537978E+01	0.1631886E+00	0.9295046E+00	0.6114805E+00	0.1000000E+01	0.1014325E+01
12	0.1827404E+00	0.9507067E-02	0.4778695E+01	0.1585012E+00	0.9088423E+00	0.6432327E+00	0.1000000E+01	0.7249721E+00
13	0.3241985E+00	0.8509427E-01	0.4190771E+01	0.1603865E+00	0.8857638E+00	0.6897774E+00	0.1000000E+01	0.4967493E+00
14	0.4593025E+00	0.1927057E+00	0.4299926E+01	0.1727232E+00	0.8542343E+00	0.7510176E+00	0.1000000E+01	0.3256713E+00
15	0.5755661E+00	0.3177591E+00	0.3525688E+01	0.1707500E+00	0.8269453E+00	0.8510849E+00	0.1000000E+01	0.1984157E+00
16	0.6901713E+00	0.4632021E+00	0.3992294E+01	0.1851702E+00	0.7713783E+00	0.9600343E+00	0.1000000E+01	0.1132449E+00
17	0.7755972E+00	0.5942535E+00	0.2329187E+01	0.1564355E+00	0.7480969E+00	0.1207007E+01	0.1000000E+01	0.5568209E-01
18	0.8688357E+00	0.7430797E+00	0.3104276E+01	0.1745672E+00	0.6241119E+00	0.1346035E+01	0.1000000E+01	0.2466239E-01
19	0.9168056E+00	0.8380355E+00	0.9004797E+00	0.1073013E+00	0.6292140E+00	0.2174353E+01	0.1000000E+01	0.7544806E-02

20	0.9722908E+00	0.9422709E+00	0.1297329E+01	0.1180832E+00	0.3516811E+00	0.1931445E+01	0.1000000E+01	0.1715585E-02
21	0.9894405E+00	0.9786985E+00	0.1119780E+00	0.4026267E-01	0.3712850E+00	0.5798176E+01	0.1000000E+01	0.1201525E-03
22	0.9994134E+00	0.9987276E+00	0.4341178E-01	0.2237466E-01	0.5893078E-01	0.2478671E+01	0.1000000E+01	0.1333310E-05
23	0.9999886E+00	0.9999768E+00	0.1278973E-03	0.1375217E-02	0.1847117E-01	0.1318345E+02	0.1000000E+01	0.1418487E-09
24	0.1000000E+01	0.1000000E+01	0.5718384E-07	0.2587793E-04	0.6996609E-04	0.2703508E+01	0.1000000E+01	0.2286153E-17
25	0.1000000E+01	0.1000000E+01	0.0000000E+00	0.1810704E-08	0.0000000E+00	0.0000000E+00	0.1000000E+01	0.0000000E+00

Man sieht sehr schön die quadratische Konvergenz in den letzten 4 Iterationen. Davor liegt nur lineare Konvergenz vor. Die folgende Abbildung zeigt die Lage der Iterationspunkte  $x^{[i]}$ .



■

# Kapitel B

## Ausblick

Was wir nicht diskutiert haben:

- Nichtdifferenzierbare Optimierungsprobleme (Strukturoptimierung, Eigenwertoptimierung)
- Optimierungsprobleme in unendlichdimensionalen Vektorräumen (z.B. optimale Steuerprozesse)
- Kombinatorische und gemischt-ganzzahlige Optimierung (z.B. An- und Ausschaltvorgänge, Gangschaltung, ganzzahlige Ressourcen)
- Vektor-Optimierung (mehrere Optimierungskriterien, die sich gegenseitig widersprechen)
- Globale Optimierung (Berechnung globaler Minima)
- Stochastische Optimierung (Genetische und evolutionäre Algorithmen)

Software:

- SCILAB: A Free Scientific Software Package; <http://www.scilab.org>
- GAMS: Guide to Available Mathematical Software; <http://gams.nist.gov/>
- NETLIB: collection of mathematical software, papers, and databases; <http://www.netlib.org/>
- Decision Tree for Optimization Software; <http://plato.la.asu.edu/guide.html>
- NEOS GUIDE: [www-fp.mcs.anl.gov/otc/Guide](http://www-fp.mcs.anl.gov/otc/Guide)
- CUTER: Große Sammlung von Testproblemen; <http://www.cuter.rl.ac.uk/>
- GALIB: set of C++ genetic algorithm objects; <http://lancet.mit.edu/ga/>
- Testprobleme: <http://www.princeton.edu/~rvdb/ampl/nlmodels/>
- Spezielle Optimierungsverfahren:
  - WORHP, dünn besetzte Probleme, C. Büskens, M. Gerds, <http://www.worhp.de>

- IPOPT, Innere-Punkt-Verfahren, A. Wächter, <https://projects.coin-or.org/Ipopt>
- KNITRO, <http://www.ziena.com/knitro.htm>
- FILTERSQP, SQP-Verfahren, R. Fletcher, S. Leyffer
- IPFILTER, Innere-Punkte-Verfahren, <http://www.mat.uc.pt/ipfilter/>
- NPSOL (NAG-Routine E04UCF), voll besetzte Probleme, P. Gill et al., University of California, San Diego
- SNOPT (NAG-Routine ), dünn besetzte Probleme, P. Gill et al., University of California, San Diego
- NLPQP, voll besetzte Probleme, K. Schittkowski, Universität Bayreuth

## Literaturverzeichnis

- [Alt02] Alt, W. *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*. Vieweg, Braunschweig/Wiesbaden, 2002.
- [Bor87] Borgwardt, K. H. *The simplex method – A probabilistic analysis*. Springer, Berlin-Heidelberg-New York, 1987.
- [BS79] Bazaraa, M. S. and Shetty, C. M. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 1979.
- [BSS93] Bazaraa, M. S., Sherali, H. D. and Shetty, C. M. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2nd edition, 1993.
- [FM90] Fiacco, A. V. and McCormick, G. P. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, volume 4 of *Classics In Applied Mathematics*. SIAM, Philadelphia, 1990.
- [GK99] Geiger, C. and Kanzow, C. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, Berlin-Heidelberg-New York, 1999.
- [GK02] Geiger, C. and Kanzow, C. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, Berlin-Heidelberg-New York, 2002.
- [GMW81] Gill, P. E., Murray, W. and Wright, M. H. *Practical Optimization*. Academic Press, London, 1981.
- [Hof53] Hoffman, A. J. *Cycling in the simplex algorithm*. National Bureau of Standards Report, 2974, 1953.
- [Kar84] Karmarkar, N. *A new polynomial-time algorithm for linear programming*. AT&T Bell Laboratories, Murray Hill, New Jersey 07974, 1984.
- [Kha79] Khachiyan, L. G. *A polynomial algorithm in linear programming*. Doklady Akademiia Nauk SSSR, 244; 1093–1096, 1979.
- [KM72] Klee, V. and Minty, G. J. *How good is the simplex algorithm?*. In *Inequalities III* (O. Shisha, editor), pp. 159–175. Academic Press, New York, 1972.
- [Man94] Mangasarian, O. L. *Nonlinear Programming*, volume 10 of *Classics In Applied Mathematics*. SIAM, Philadelphia, 1994.

- [MS69] Marshall, K. T. and Suurballe, J. W. *A note on cycling in the simplex method*. Naval Res. Logist. Quart., 16; 121–137, 1969.
- [NW99] Nocedal, J. and Wright, S. J. *Numerical optimization*. Springer Series in Operations Research, New York, 1999.
- [Spe93] Spellucci, P. *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser, Basel, 1993.
- [Wri97] Wright, S. E. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.