

---

# Einführung in die Optimierung

---

Vorlesungsnotizen Wintersemester 2022/23,  
Prof. Dr. Yann Disser



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



---

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Einteilung von Optimierungsproblemen . . . . .	5
<b>2</b>	<b>Konvexe Mengen und Funktionen</b>	<b>9</b>
2.1	Konvexe Mengen . . . . .	9
2.2	Trennungssätze . . . . .	13
2.3	Extrempunkte, Extremrichtungen, Stützeigenschaften . . . . .	16
2.4	Konvexe Funktionen . . . . .	18
2.5	Optimalitätsresultate für konvexe Optimierung . . . . .	22
<b>3</b>	<b>Polyeder</b>	<b>25</b>
3.1	Seitenflächen, Ecken und Facetten . . . . .	28
3.2	Dimension und Darstellung von Seitenflächen . . . . .	32
<b>4</b>	<b>Grundlagen der Linearen Optimierung</b>	<b>39</b>
4.1	Duales Problem und schwache Dualität . . . . .	39
4.2	Beispiel: Batterierecycling . . . . .	43
4.3	Zulässige Richtungen und Lösungen . . . . .	46
4.4	Optimalitätsbedingungen . . . . .	48
<b>5</b>	<b>Der Simplex-Algorithmus</b>	<b>53</b>
5.1	Basen, Basislösungen und Degeneriertheit . . . . .	53
5.2	Die Grundversion des Simplex-Verfahrens . . . . .	55
5.3	Pivotregeln . . . . .	60
5.4	Laufzeit des Simplex-Verfahrens . . . . .	61
5.5	Finden einer Startbasis . . . . .	63
5.6	Das duale Simplex-Verfahren . . . . .	65
5.7	Sensitivitätsanalyse . . . . .	69
<b>6</b>	<b>Die Ellipsoidmethode</b>	<b>73</b>
6.1	Reduktion auf volldimensionale Zulässigkeitsprobleme . . . . .	74
6.2	Die Ellipsoidmethode . . . . .	81
6.3	Separieren und Optimieren . . . . .	89
<b>7</b>	<b>Grundlagen der Nichtlinearen Optimierung</b>	<b>93</b>
7.1	Optimalitätsbedingungen . . . . .	93
7.1.1	Tangential- und Linearisierungskegel . . . . .	93
7.1.2	KKT-Bedingungen . . . . .	96
7.2	Quadratische Probleme . . . . .	98
7.2.1	Gleichungsrestriktionen . . . . .	98
7.2.2	Strategie der aktiven Menge . . . . .	99



*Dieses Skript orientiert sich inhaltlich am Skript zur gleichnamigen Vorlesung, gehalten von Marc Pfetsch und Stefan Ulbrich an der TU Darmstadt.*

---

---

# 1 Einleitung

Mathematische Optimierung beschäftigt sich allgemein mit der Lösung des folgenden Problems:<sup>1</sup>

**Minimierungsproblem (Suchproblem)**

**Input:** Funktion  $f : \mathcal{D} \rightarrow \mathbb{R}$ , Menge  $\mathcal{X} \subseteq \mathcal{D}$

**Problem:** finde  $\mathbf{x}^* \in \mathcal{X}$  mit  $f(\mathbf{x}^*) = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ , sofern es existiert

bzw. des analogen Maximierungsproblems. Optimierungsprobleme sind Instanzen dieses allgemeinen Problems und werden üblicherweise geschrieben als

$$\text{(MIN)} \quad \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \quad \text{bzw.} \quad \text{(MAX)} \quad \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$$

wobei die Abkürzung „s.t.“ für „subject to“ steht, also für „unter den Nebenbedingungen“.

In dieser Vorlesung beschränken wir uns auf den Fall  $\mathcal{D} \subseteq \mathbb{R}^n$  für  $n \in \mathbb{N}$ . Für ein gegebenes Optimierungsproblem bezeichnen wir  $f : \mathcal{D} \rightarrow \mathbb{R}$  als die *Zielfunktion* und  $\mathcal{X} \subseteq \mathcal{D} \subseteq \mathbb{R}^n$  als die *zulässige Menge*, bestehend aus den *zulässigen Lösungen*  $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathcal{X}$ . Die Komponenten  $x_i$  von  $\mathbf{x}$  nennen wir die *Variablen* des Optimierungsproblems. Die Einschränkung  $\mathbf{x} \in \mathcal{X}$  nennen wir die *Nebenbedingungen*. Dabei werden die zulässige Menge  $\mathcal{X}$ , bzw. die Nebenbedingungen, oft über Gleichungen und Ungleichungen beschrieben, also mittels Funktionen  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , so dass

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}.$$

Ist  $\mathcal{X} = \mathbb{R}^n$ , sprechen wir von *unrestringierter Optimierung* bzw. *Optimierung ohne Nebenbedingungen*.

Die gesuchte *Optimallösung*  $\mathbf{x}^*$  ist also ein (globales) Extremum der Funktion  $f$  über der Menge  $\mathcal{X}$ . Wir unterscheiden dabei zwischen Minima und Maxima.

**Definition 1.1.** Ein Punkt  $\bar{\mathbf{x}} \in \mathcal{X}$  heißt (*globaler*) *Minimalpunkt* von  $f : \mathcal{D} \rightarrow \mathbb{R}$  über  $\mathcal{X} \subseteq \mathcal{D}$ , wenn

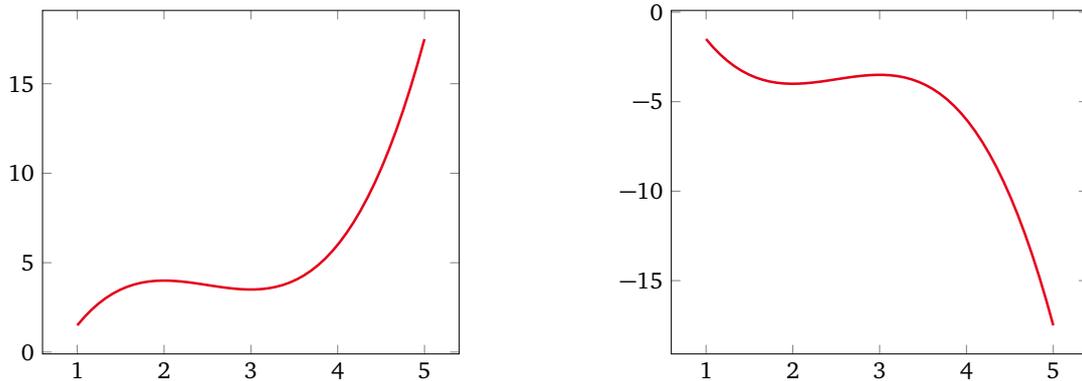
$$\bar{\mathbf{x}} \in \arg \min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} := \{\mathbf{x}' \in \mathcal{X} : f(\mathbf{x}') = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}\}$$

und (*globaler*) *Maximalpunkt* von  $f$  über  $\mathcal{X}$ , wenn

$$\bar{\mathbf{x}} \in \arg \max\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} := \{\mathbf{x}' \in \mathcal{X} : f(\mathbf{x}') = \max\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}\}.$$

---

<sup>1</sup> Wir verwenden in diesem Skript die Konvention, dass Vektoren  $\mathbf{x} \in \mathbb{R}^n$  fett gedruckt werden, im Gegensatz zu Skalaren  $x \in \mathbb{R}$  und Matrizen  $A \in \mathbb{R}^{m \times n}$ .



**Abbildung 1.1:** Die Funktionen  $f$  aus Beispiel 1.2 (links) und  $-f$  aus Beispiel 1.3 auf  $\mathcal{X} = [1, 5]$ .

**Beispiel 1.2.** Betrachte das eindimensionale Minimierungsproblem (siehe Abb. 1.1)

$$\begin{aligned} \min \quad & x^3 - 7.5x^2 + 18x - 10 \\ \text{s.t.} \quad & x - 1 \geq 0, \\ & x^2 - 5x \leq 0. \end{aligned}$$

Die Zielfunktion ist also  $f(x) = x^3 - 7.5x^2 + 18x - 10.5$  und die zulässige Menge ist  $\mathcal{X} = \{x \in \mathbb{R} : x - 1 \geq 0, x^2 - 5x \leq 0\} = [1, 5]$ . Es ist  $\arg \min\{f(x) : x \in \mathcal{X}\} = \{1\}$ , also ist die eindeutige Optimallösung der Minimalpunkt  $x^* = 1$  und der zugehörige Optimalwert ist  $f(x^*) = 1.5$ .  $\triangle$

Wegen

$$\begin{aligned} \arg \max\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} &= \arg \min\{-f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \\ \max\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} &= -\min\{-f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \end{aligned}$$

können wir natürlich jedes Maximierungsproblem als Minimierungsproblem auffassen und umgekehrt.

**Beispiel 1.3.** Wir können das Optimierungsproblem aus Beispiel 1.2 äquivalent als Maximierungsproblem schreiben als (siehe Abb. 1.1)

$$\begin{aligned} \max \quad & -x^3 + 7.5x^2 - 18x + 10 \\ \text{s.t.} \quad & x - 1 \geq 0, \\ & x^2 - 5x \leq 0. \end{aligned}$$

Das Optimum wird wieder im Maximalpunkt  $x^* = 1$  angenommen und der zugehörige Optimalwert ist  $-1.5$ .  $\triangle$

Wie man in den Beispielen am Punkt  $x = 3$  sieht, kann es im Allgemeinen auch lokale Extrema im Inneren der zulässigen Menge geben, die keine globalen Extrema sind. Dies ist eine zentrale Schwierigkeit bei der Optimierung von nichtlinearen Zielfunktionen. Wir werden uns in dieser Vorlesung auf Optimierungsprobleme beschränken bei denen nur globale Extrema auftreten können.

---

**Definition 1.4.** Sei  $f : \mathcal{D} \rightarrow \mathbb{R}$  und  $\bar{x} \in \mathcal{X}$ . Dann ist  $\bar{x}$  ein *lokaler Minimalpunkt* (bzw. *Maximalpunkt*) von  $f$  über  $\mathcal{X} \subseteq \mathcal{D}$ , wenn eine Umgebung  $\mathcal{U}(\bar{x}) \subseteq \mathbb{R}^n$  von  $\bar{x}$  existiert, so dass  $\bar{x}$  ein globaler Minimalpunkt (bzw. Maximalpunkt) von  $f$  über  $\mathcal{U}(\bar{x}) \cap \mathcal{X}$  ist.

Wir kennen aus der Analysis bereits hinreichende lokale Optimalitätsbedingungen für unrestringierte Optimierungsprobleme mit zweifach stetig differenzierbaren Zielfunktionen  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ : Es ist  $\mathbf{x}^* \in \mathbb{R}^n$  ein lokaler Minimal-/Maximalpunkt von  $f$ , wenn der Gradient  $\nabla f(\mathbf{x}^*)$  verschwindet und die Hesse-Matrix  $(\partial_i \partial_j f(\mathbf{x}^*))_{i,j \in \{1, \dots, n\}}$  positiv/negativ definit ist. Außerdem liefert der Extremwertsatz (siehe *Analysis II*) die Existenz der globalen Extrema für stetige  $f$  auf kompakten  $\mathcal{X}$ . Ferner haben wir die notwendige Bedingung für lokale Extrema von stetig differenzierbaren Zielfunktionen  $f$  auf Untermannigfaltigkeiten  $\mathcal{X}$  gesehen, dass der Tangentialanteil des Gradienten bezüglich  $\mathcal{X}$  verschwindet.

In der Optimierung beschäftigen wir uns mit der Existenz globaler Lösungen und, vor allem, mit effizienten algorithmischen (also konstruktiven) Verfahren zum Finden einer globalen Optimallösung. Dabei wird  $\mathcal{X}$  im Allgemeinen nicht beschränkt sein.

---

## 1.1 Einteilung von Optimierungsproblemen

---

Diese Vorlesung soll die Grundlagen der mathematischen Optimierung vermitteln, die insbesondere für weiterführende Vorlesungen in der diskreten und nichtlinearen Optimierung gebraucht werden. Wir geben nun einen Ausblick auf die wichtigsten Klassen von Optimierungsproblemen.

---

### Lineare Optimierungsprobleme

---

Von einem linearen Optimierungsproblem spricht man, wenn sowohl die Zielfunktion, als auch die Nebenbedingungen affin linear sind. Letzteres meint, dass die zulässige Menge ausgedrückt werden kann als

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$$

mit affin linearen Funktionen  $\mathbf{h}$  und  $\mathbf{g}$  (siehe auch Definition 4.1). Wir werden uns in dieser Vorlesung ausführlich mit der Struktur von linearen Optimierungsproblemen beschäftigen (Kapitel 4) und Lösungsmethoden für sie entwickeln (Kapitel 5 und 6). Lineare Optimierungsproblemen haben eine besondere Bedeutung, da sie eine Vielzahl von Fragestellungen abbilden können, aber dennoch in polynomieller Zeit lösbar sind (siehe Kapitel 6).

**Beispiel 1.5.** Das Problem einen maximalen  $s$ - $t$ -Fluss  $f : E \rightarrow \mathbb{R}_{\geq 0}$  in einem Netzwerk  $(G = (V, E), \mu : E \rightarrow \mathbb{R}_{\geq 0}, s \in V, t \in V)$  zu finden (siehe ADM), lässt sich wie folgt als lineares Programm formulieren:

$$\begin{aligned} \max \quad & \sum_{e \in \delta^+(s)} x_e && \text{(Ausfluss von } s) \\ \text{s.t. } \quad & x_e \leq \mu(e) && \forall e \in E, \text{ (Kapazitätsschranken)} \\ & \sum_{e \in \delta^-(v)} x_e = \sum_{e \in \delta^+(v)} x_e && \forall v \in V \setminus \{s, t\}, \text{ (Flusserhaltung)} \\ & \mathbf{x} \geq \mathbf{0}, && \text{(Positivität)} \end{aligned}$$

wobei der Vektor  $\mathbf{x} \in \mathbb{R}^E$  die Flusswerte entlang der Kanten beschreibt. △

---

## Diskrete Optimierungsprobleme

---

Das definierende Merkmal von diskreten Optimierungsproblemen ist das Auftreten von Ganzzahligkeitsbedingungen der Form  $x_i \in \mathbb{Z}$  als zusätzliche Nebenbedingungen. Diskrete Optimierungsprobleme haben ansonsten meistens (affin) lineare Zielfunktionen und Nebenbedingungen. Die Ganzzahligkeitsbedingungen erlauben es wesentlich mehr algorithmische Fragen abzubilden als durch lineare Optimierungsprobleme. Das folgende Beispiel zeigt, dass diskrete Optimierungsprobleme im Allgemeinen NP-schwer zu lösen sind.

**Beispiel 1.6.** Eine SAT-Instanz mit Klauseln  $\{C_i\}_{i \in \{1, \dots, m\}}$  und Variablen  $\{x_j\}_{j \in \{1, \dots, n\}}$  (siehe ADM) ist erfüllbar genau dann, wenn das lineare Ungleichungssystem

$$\sum_{j: x_j \in C_i} \hat{x}_j + \sum_{j: \bar{x}_j \in C_i} (1 - \hat{x}_j) \geq 1 \quad \forall i \in \{1, \dots, m\}$$

eine Lösung  $\hat{x} \in \{0, 1\}^n$  hat. △

---

## Kombinatorische Optimierungsprobleme

---

Von kombinatorischen Optimierungsproblemen sprechen wir, wenn die zulässige Menge  $\mathcal{X}$  endlich ist. Beispiel 1.6 zeigt, dass diese Probleme im Allgemeinen NP-schwer sind.

---

## Kontinuierliche (nichtlineare) Optimierungsprobleme

---

Dies sind allgemein Optimierungsprobleme, bei denen keine Ganzzahligkeitsbedingungen auftreten. Die Zielfunktion und Nebenbedingungen sind insbesondere auch nichtlinear zugelassen.

---

## Globale Optimierungsprobleme

---

Dies sind Optimierungsprobleme, bei denen lokale Extrema auftreten können, die nicht schon global optimal sind. Hier müssen spezielle Optimierungstechniken entwickelt werden.

---

## Konvexe Optimierungsprobleme

---

Wir werden sehen (Kapitel 2), dass Konvexität von Zielfunktion und zulässiger Menge hinreichend ist, damit nur globale Extrema existieren. Konvexe Optimierungsprobleme sind Optimierungsprobleme, die dieser hinreichenden Bedingung genügen. Für diese Probleme ist es möglich, Algorithmen zu entwickeln, die auf lokaler Information basieren, wie den Gradienten der Zielfunktion  $f$  und der Funktionen  $\mathbf{g}$  und  $\mathbf{h}$  in der Beschreibung der Nebenbedingungen.

**Beispiel 1.7.** Sei der Vektor  $(r_1, \dots, r_n)^\top \in \mathbb{R}^n$  eine Zufallsvariable mit Erwartungswert  $\boldsymbol{\mu} \in \mathbb{R}^n$  und Kovarianzmatrix  $\Sigma \in \mathbb{R}^{n \times n}$ , die die Rendite von  $n$  verschiedenen Wertpapieren beschreibt. Wir wollen entscheiden, welchen Anteil  $x_i$  unseres Budgets wir jeweils in das  $i$ -te Wertpapier investieren sollen. Wollen wir eine erwartete Rendite von  $\rho \in \mathbb{R}$  bei minimaler Varianz erzielen, so führt dies auf das (wie wir sehen werden) konvexe quadratische Minimierungsproblem

$$\min \mathbf{x}^\top \Sigma \mathbf{x} \quad \text{s.t.} \quad \mathbf{1}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq \mathbf{0}, \quad \boldsymbol{\mu}^\top \mathbf{x} \geq \rho.$$

---

Wollen wir hingegen die erwartete Rendite maximieren bei Varianz höchstens  $v \in \mathbb{R}$ , so erhalten wir das konvexe Maximierungsproblem mit quadratischen Nebenbedingungen

$$\max \boldsymbol{\mu}^\top \mathbf{x} \quad \text{s.t.} \quad \mathbf{1}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq \mathbf{0}, \quad \mathbf{x}^\top \Sigma \mathbf{x} \leq v. \quad \triangle$$



---

## 2 Konvexe Mengen und Funktionen

Wir beschäftigen uns in diesem Kapitel mit dem Begriff der Konvexität von Mengen und Funktionen. Die Bedeutung dieses Begriffs für die Optimierung liegt darin, dass die Konvexität der Zielfunktion und der zulässigen Menge zusammen eine hinreichende Bedingung liefern, dass nur globale Extrema existieren.

---

### 2.1 Konvexe Mengen

---

Wir definieren die Konvexität einer Menge über Konvexkombinationen von zwei Punkten und zeigen im Anschluss, dass wir äquivalent dazu auch Konvexkombinationen beliebig vieler Punkte betrachten können (siehe auch Abb. 2.1).

**Definition 2.1.** Für  $\mathbf{x}_1, \dots, \mathbf{x}_p \in \mathbb{R}^n$  und  $\lambda_1, \dots, \lambda_p \geq 0$  mit  $p \in \mathbb{N}$  heißt

$$\mathbf{z} = \sum_{i=1}^p \lambda_i \mathbf{x}_i$$

eine *konische Kombination* der Punkte  $\mathbf{x}_1, \dots, \mathbf{x}_p$ . Ist  $\sum_{i=1}^p \lambda_i = 1$ , so heißt  $\mathbf{z}$  eine *Konvexkombination* von  $\mathbf{x}_1, \dots, \mathbf{x}_p$ . Gilt zusätzlich  $\lambda_1, \dots, \lambda_p \in (0, 1)$ , so heißt  $\mathbf{z}$  *echte Konvexkombination*.

**Definition 2.2.** Eine Menge  $\mathcal{C} \subseteq \mathbb{R}^n$  heißt *konvex*, wenn sie alle Konvexkombinationen zweier Punkte  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$  enthält.

**Satz 2.3.** Eine Menge  $\mathcal{C} \subseteq \mathbb{R}^n$  ist konvex genau dann, wenn sie alle Konvexkombinationen von Punkten aus  $\mathcal{C}$  enthält.

*Beweis.* Ist jede Konvexkombinationen von Punkten aus  $\mathcal{C}$  wieder in  $\mathcal{C}$  enthalten, dann auch jede Konvexkombination aus zwei Punkten. Damit ist  $\mathcal{C}$  per Definition konvex.

Nehmen wir also andersherum an,  $\mathcal{C}$  sei konvex. Es genügt Konvexkombinationen  $\mathbf{z} = \sum_{i=1}^p \lambda_i \mathbf{x}_i$  mit  $\mathbf{x}_1, \dots, \mathbf{x}_p \in \mathcal{C}$ ,  $\lambda_1, \dots, \lambda_p > 0$  und  $p \in \mathbb{N}$  zu betrachten. Für  $p = 1$  ist  $\mathbf{z} = \mathbf{x}_1 \in \mathcal{C}$ . Wir beweisen nun, dass  $\mathbf{z} \in \mathcal{C}$  für  $p \geq 2$  per Induktion über  $p$ .

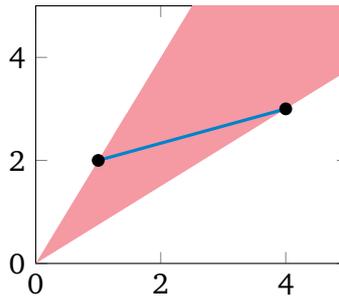
Wegen  $p > 1$  und  $\lambda_1 > 0$  gilt  $\lambda_p < 1$  und wir können  $\mathbf{z}$  schreiben als

$$\mathbf{z} = (1 - \lambda_p) \mathbf{y} + \lambda_p \mathbf{x}_p$$

mit

$$\mathbf{y} := \sum_{i=1}^{p-1} \frac{\lambda_i}{1 - \lambda_p} \mathbf{x}_i.$$

Wegen  $\frac{\lambda_i}{1 - \lambda_p} > 0$  für alle  $i \in \{1, \dots, p-1\}$  und  $\sum_{i=1}^{p-1} \frac{\lambda_i}{1 - \lambda_p} = 1$  ist  $\mathbf{y}$  eine Konvexkombination von  $p-1$  Punkten aus  $\mathcal{C}$ , also gilt per Induktion  $\mathbf{y} \in \mathcal{C}$  und  $\mathbf{z}$  ist eine Konvexkombination aus zwei Punkten in  $\mathcal{C}$ . Somit ist  $\mathbf{z} \in \mathcal{C}$  per Definition der Konvexität von  $\mathcal{C}$ .  $\square$



**Abbildung 2.1:** Zwei Punkte in  $\mathbb{R}^2$  mit ihren konischen Kombinationen (rot) und Konvexkombinationen (blau).

Eine besonders wichtige Familie von konvexen Mengen für die Beschreibung von Nebenbedingungen sind Hyperebenen und die von ihnen erzeugten Halbräume. Bei der folgenden Definition gilt es zu beachten, dass der Normalvektor einer Hyperebene nicht eindeutig bestimmt ist, sondern beliebig skaliert werden kann.

**Definition 2.4.** Eine *Hyperebene* ist eine Menge der Form

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\},$$

wobei  $\mathbf{a} \in \mathbb{R}^n \setminus \{0\}$  *Normalvektor* zu  $\mathcal{H}$  heißt und  $\beta \in \mathbb{R}$ . Die Hyperebene  $\mathcal{H}$  erzeugt die *abgeschlossenen Halbräume*

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\} \quad \text{und} \quad \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \geq \beta\},$$

sowie die zugehörigen *offenen Halbräume*

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} < \beta\} \quad \text{und} \quad \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} > \beta\}.$$

**Proposition 2.5** (Übung). Seien  $\mathcal{A}_1, \dots, \mathcal{A}_k \subseteq \mathbb{R}^n$  konvex. Folgende Mengen sind konvex:

- (i) jede Schnittmenge konvexer Mengen;
- (ii) jede Hyperebene und jeder von ihr erzeugte offene oder abgeschlossene Halbraum;
- (iii) das kartesische Produkt  $\mathcal{A}_1 \times \dots \times \mathcal{A}_k = \{(\mathbf{a}_1, \dots, \mathbf{a}_k) : \mathbf{a}_i \in \mathcal{A}_i \forall i \in \{1, \dots, k\}\}$ ;
- (iv) die Minkowski-Summe  $\sum_{i=1}^k \alpha_i \mathcal{A}_i = \{\sum_{i=1}^k \alpha_i \mathbf{a}_i : \mathbf{a}_i \in \mathcal{A}_i \forall i \in \{1, \dots, k\}\}$  für  $\alpha \in \mathbb{R}^k$ ;
- (v) jede Menge der Form  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \triangleleft \mathbf{b}\}$ , mit  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  und  $\triangleleft \in \{<, \leq, =, \geq, >\}$ ;
- (vi) jede offene (und auch jede abgeschlossene) Kugel

$$\mathcal{B}_\alpha(\mathbf{z}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{z}\| < \alpha\}$$

mit Mittelpunkt  $\mathbf{z} \in \mathbb{R}^n$  und Radius  $\alpha > 0$ ;

- (vii) die Mengen  $\emptyset$  und  $\mathbb{R}^n$ , sowie jede Menge  $\{\mathbf{x}\}$  mit  $\mathbf{x} \in \mathbb{R}^n$ .

Jede Menge induziert eine eindeutige konvexe Menge, nämlich ihre kleinste konvexe Obermenge.

**Definition 2.6.** Die *konvexe Hülle* einer Menge  $\mathcal{M} \subseteq \mathbb{R}^n$  ist die kleinste konvexe Menge, die  $\mathcal{M}$  enthält, also

$$\text{conv}(\mathcal{M}) := \bigcap \{ \mathcal{C} \subseteq \mathbb{R}^n : \mathcal{M} \subseteq \mathcal{C}, \mathcal{C} \text{ konvex} \}.$$

Alternativ lässt sich die konvexe Hülle als der „konvexe Abschluss“, also als Abschluss im Bezug auf Bildung von Konvexkombinationen, einer Menge charakterisieren. Die konvexe Hülle ist allerdings natürlich nicht unbedingt (folgen-)abgeschlossen (z.B. ist  $\text{conv}((0, 1)) = (0, 1)$  nicht abgeschlossen).

**Satz 2.7** (Übung). Für alle  $\mathcal{M} \subseteq \mathbb{R}^n$  ist  $\text{conv}(\mathcal{M})$  die Menge aller Konvexkombinationen von Punkten aus  $\mathcal{M}$ , also  $\text{conv}(\mathcal{M}) = \{ A\boldsymbol{\lambda} : A \in \mathbb{R}^{n \times k}, \{ \mathbf{A}_1, \dots, \mathbf{A}_k \} \subseteq \mathcal{M}, k \in \mathbb{N}, \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{1}^\top \boldsymbol{\lambda} = 1 \}$ .

Wir zeigen, dass man sich in Satz 2.7 auf Konvexkombinationen aus  $n + 1$  Punkten beschränken kann.

**Satz 2.8** (Carathéodory). Für  $\mathcal{M} \subseteq \mathbb{R}^n$  ist  $\text{conv}(\mathcal{M})$  die Menge aller Konvexkombinationen aus  $n + 1$  Punkten in  $\mathcal{M}$ .

*Beweis.* Sei  $\bar{\mathbf{x}} \in \text{conv}(\mathcal{M})$ . Nach Satz 2.7 existieren dann  $\mathbf{x}_1, \dots, \mathbf{x}_p \in \mathcal{M}$  und  $\boldsymbol{\lambda} \in \mathbb{R}^p, \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{1}^\top \boldsymbol{\lambda} = 1$  mit

$$\bar{\mathbf{x}} = \sum_{i=1}^p \lambda_i \mathbf{x}_i.$$

Es ist nun zu zeigen, dass wir eine solche Darstellung von  $\bar{\mathbf{x}}$  mit  $p \leq n + 1$  finden können. Dazu zeigen wir, dass im Fall  $p > n + 1$  der Vektor  $\bar{\mathbf{x}}$  die Konvexkombination von echt weniger als  $p$  Punkten aus  $\{ \mathbf{x}_1, \dots, \mathbf{x}_p \}$  ist.

Wir definieren die Vektoren  $\mathbf{y}_i = \mathbf{x}_i - \mathbf{x}_p$  für  $i \in \{ 1, \dots, p - 1 \}$ . Es ist  $p - 1 \geq n + 1$ , also sind diese Vektoren linear abhängig, das heißt, es existiert  $\boldsymbol{\alpha} \in \mathbb{R}^{p-1} \setminus \{ \mathbf{0} \}$ , so dass

$$\mathbf{0} = \sum_{i=1}^{p-1} \alpha_i \mathbf{y}_i = \sum_{i=1}^{p-1} \alpha_i \mathbf{x}_i - \mathbf{x}_p \sum_{i=1}^{p-1} \alpha_i.$$

Wir dürfen  $\alpha_p := - \sum_{i=1}^{p-1} \alpha_i \leq 0$  annehmen, sonst können wir  $\boldsymbol{\alpha}$  negieren. Damit haben wir also

$$\sum_{i=1}^p \alpha_i \mathbf{x}_i = \mathbf{0}, \tag{2.1}$$

$$\sum_{i=1}^p \alpha_i = 0. \tag{2.2}$$

Wir können also beliebige Vielfache von  $\boldsymbol{\alpha}$  zu  $\boldsymbol{\lambda}$  addieren, und dabei eine Linearkombination mit Faktoren  $\boldsymbol{\gamma}$  erhalten, die immer noch  $\bar{\mathbf{x}}$  ergibt und deren Faktoren weiterhin zu  $\mathbf{1}^\top \boldsymbol{\gamma} = 1$

summieren. Es bleibt nur noch das Vielfache genau so zu wählen, dass  $\boldsymbol{\gamma} \geq \mathbf{0}$  und dass ein Faktor  $\gamma_{i_0}$  verschwindet.

Wegen  $\boldsymbol{\alpha} \neq \mathbf{0}$  und (2.2) existiert  $i \in \{1, \dots, p\}$  mit  $\alpha_i > 0$ . Wegen  $\alpha_p \leq 0$ , können wir

$$i_0 \in \arg \min \left\{ \frac{\lambda_i}{\alpha_i} : i \in \{1, \dots, p-1\}, \alpha_i > 0 \right\}$$

wählen.

Wir setzen jetzt  $\boldsymbol{\gamma} := \boldsymbol{\lambda} - \frac{\lambda_{i_0}}{\alpha_{i_0}} \boldsymbol{\alpha} \in \mathbb{R}^p$ . Per Definition von  $i_0$  und wegen  $\boldsymbol{\lambda} \geq \mathbf{0}$  gilt  $\gamma_{i_0} = 0$  und  $\gamma_i \geq 0$  für alle  $i \in \{1, \dots, p\}$ , unabhängig vom Vorzeichen von  $\alpha_i$ . Außerdem gilt

$$\mathbf{1}^\top \boldsymbol{\gamma} = \sum_{i=1}^p \left( \lambda_i - \frac{\lambda_{i_0}}{\alpha_{i_0}} \alpha_i \right) \stackrel{(2.2)}{=} \mathbf{1}^\top \boldsymbol{\lambda} = 1.$$

Also lässt sich  $\bar{\mathbf{x}}$  als Konvexkombination ohne Vektor  $\mathbf{x}_{i_0}$  schreiben als

$$\bar{\mathbf{x}} = \sum_{i=1}^p \lambda_i \mathbf{x}_i \stackrel{(2.1)}{=} \sum_{i=1}^p \left( \lambda_i - \frac{\lambda_{i_0}}{\alpha_{i_0}} \alpha_i \right) \mathbf{x}_i = \sum_{i \in \{1, \dots, p\} \setminus \{i_0\}} \gamma_i \mathbf{x}_i. \quad \square$$

Insbesondere haben wir folgende Aussage gezeigt.

**Korollar 2.9.** Jede konische Kombination  $\sum_{i=1}^k \alpha_i \mathbf{z}_i$ ,  $\boldsymbol{\alpha} \geq \mathbf{0}$  lässt sich als konische Kombination von linear unabhängigen Vektoren aus  $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  schreiben.

*Beweis.* Es genügt zu zeigen, dass wir einen Vektor aus  $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  aus der konischen Kombination eliminieren können, falls diese Vektoren linear abhängig sind. Wir dürfen  $\boldsymbol{\alpha} > \mathbf{0}$  annehmen. Mit  $\lambda_i = \frac{1}{k+1}$  und  $\mathbf{x}_i = (k+1)\alpha_i \mathbf{z}_i$  für  $i \in \{1, \dots, k\}$ , sowie  $\lambda_{k+1} = \frac{1}{k+1}$ ,  $\mathbf{x}_{k+1} = \mathbf{0}$  und  $p = k+1$  können wir

$$\sum_{i=1}^k \alpha_i \mathbf{z}_i = \sum_{i=1}^p \lambda_i \mathbf{x}_i,$$

mit  $\mathbf{1}^\top \boldsymbol{\lambda} = 1$  und  $\boldsymbol{\lambda} \geq \mathbf{0}$  als Konvexkombination schreiben. Der Beweis von Satz 2.8 benötigt nur, dass die Vektoren  $\mathbf{y}_i = \mathbf{x}_i - \mathbf{x}_p = (k+1)\alpha_i \mathbf{z}_i$  für  $i \in \{1, \dots, p-1\}$  linear abhängig sind. Für linear abhängige  $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  ist das (wegen  $\boldsymbol{\alpha} > \mathbf{0}$ ) der Fall und wir können mit dem gleichen Verfahren einen Vektor  $\mathbf{x}_{i_0}$  mit  $i_0 \neq p$  aus der Konvexkombination  $\sum_{i=1}^p \lambda_i \mathbf{x}_i$  eliminieren. Wegen  $\mathbf{x}_p = \mathbf{0}$  ist damit

$$\sum_{i=1}^k \alpha_i \mathbf{z}_i = \sum_{i \in \{1, \dots, p\}} \lambda_i \mathbf{x}_i = \sum_{i \in \{1, \dots, p-1\} \setminus \{i_0\}} \gamma_i \mathbf{x}_i = \sum_{i \in \{1, \dots, k\} \setminus \{i_0\}} \gamma_i (k+1) \alpha_i \mathbf{z}_i$$

eine konische Kombination und wir haben einen Vektor aus  $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  eliminiert. □

Wir erwähnen eine weitere nützliche Eigenschaft.

**Proposition 2.10** (Übung). Die konvexe Hülle einer kompakten Menge ist kompakt.

---

## 2.2 Trennungssätze

---

Um die Unzulässigkeit eines Punktes  $\mathbf{x} \in \mathbb{R}^n$ , beziehungsweise einer Menge  $\mathcal{M} \subseteq \mathbb{R}^n$  von Punkten, bezüglich einer zulässigen Menge  $\mathcal{C}$  zu belegen, müssen wir  $\mathbf{x}$  und  $\mathcal{C}$  trennen. Wir zeigen im Folgenden, dass eine solche Trennung für konvexe Mengen immer bereits durch eine Hyperebene möglich ist. Es sei darauf hingewiesen, dass die allgemeinere Trennung konvexer Mengen in normierten Vektorräumen durch lineare Funktionale in der Vorlesung *Funktionalanalysis* behandelt wird.

**Definition 2.11.** Eine Hyperebene  $\mathcal{H}$  trennt  $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{R}^n$  (strikt), wenn  $\mathcal{M}_1, \mathcal{M}_2$  in unterschiedlichen abgeschlossenen (bzw. offenen) Halbräumen liegen, die von  $\mathcal{H}$  erzeugt werden.

Wir beginnen mit der strikten Trennung eines Punkts von einer abgeschlossenen, konvexen Menge.

**Proposition 2.12.** Sei  $\mathcal{C} \subseteq \mathbb{R}^n$  abgeschlossen und konvex mit  $\mathbf{0} \notin \mathcal{C} \neq \emptyset$ . Dann existiert eine Hyperebene, die  $\mathcal{C}$  und  $\{\mathbf{0}\}$  strikt trennt.

*Beweis.* Sei  $\alpha > 0$  so gewählt, dass  $\overline{\mathcal{B}}_\alpha \cap \mathcal{C} \neq \emptyset$  für  $\overline{\mathcal{B}}_\alpha = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq \alpha\}$ . Da  $\overline{\mathcal{B}}_\alpha$  und  $\mathcal{C}$  abgeschlossen sind und  $\overline{\mathcal{B}}_\alpha$  beschränkt ist, ist  $\overline{\mathcal{B}}_\alpha \cap \mathcal{C}$  kompakt. Daher nimmt die stetige Funktion  $\|\mathbf{x}\|$  ihr Minimum über  $\overline{\mathcal{B}}_\alpha \cap \mathcal{C}$  in einem Punkt  $\bar{\mathbf{x}} \in \overline{\mathcal{B}}_\alpha \cap \mathcal{C}$  an. Wegen  $\mathbf{0} \notin \mathcal{C}$  ist  $0 < \|\bar{\mathbf{x}}\|$  und wegen der Wahl von  $\bar{\mathbf{x}}$  mit  $\|\bar{\mathbf{x}}\| \leq \alpha$  ist

$$\bar{\mathbf{x}}^\top \bar{\mathbf{x}} = \|\bar{\mathbf{x}}\|^2 \leq \|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x} \quad \forall \mathbf{x} \in \mathcal{C}. \quad (2.3)$$

Es genügt nun zu zeigen, dass  $\bar{\mathbf{x}}^\top \mathbf{x} \geq \bar{\mathbf{x}}^\top \bar{\mathbf{x}}$  für alle  $\mathbf{x} \in \mathcal{C}$ : Wegen  $\bar{\mathbf{x}}^\top \bar{\mathbf{x}} = \|\bar{\mathbf{x}}\|^2 > 0$ , trennt die Hyperebene  $\{\mathbf{x} \in \mathbb{R}^n : \alpha^\top \mathbf{x} = \beta\}$  mit  $\alpha = \bar{\mathbf{x}}$  und  $\beta = \frac{1}{2} \bar{\mathbf{x}}^\top \bar{\mathbf{x}}$  dann strikt  $\mathcal{C}$  und  $\{\mathbf{0}\}$ .

Sei also  $\mathbf{x} \in \mathcal{C}$  beliebig gewählt. Da  $\mathcal{C}$  konvex ist, gilt für alle  $\lambda \in [0, 1]$ , dass  $(\lambda \mathbf{x} + (1-\lambda)\bar{\mathbf{x}}) \in \mathcal{C}$  und damit, wegen (2.3), dass

$$\begin{aligned} 0 &\leq \|\lambda \mathbf{x} + (1-\lambda)\bar{\mathbf{x}}\|^2 - \|\bar{\mathbf{x}}\|^2 \\ &= \|\lambda(\mathbf{x} - \bar{\mathbf{x}}) + \bar{\mathbf{x}}\|^2 - \|\bar{\mathbf{x}}\|^2 \\ &= \lambda^2 \|\mathbf{x} - \bar{\mathbf{x}}\|^2 + 2\lambda \bar{\mathbf{x}}^\top (\mathbf{x} - \bar{\mathbf{x}}). \end{aligned}$$

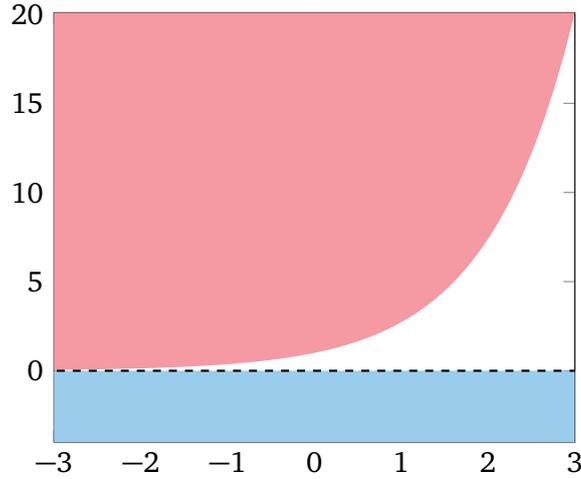
Für alle  $\lambda \in (0, 1]$  gilt also

$$\bar{\mathbf{x}}^\top (\mathbf{x} - \bar{\mathbf{x}}) \geq -\frac{\lambda}{2} \|\mathbf{x} - \bar{\mathbf{x}}\|^2.$$

Da dies für beliebig kleine  $\lambda > 0$  gelten muss, ist also  $\bar{\mathbf{x}}^\top (\mathbf{x} - \bar{\mathbf{x}}) \geq 0$  und somit  $\bar{\mathbf{x}}^\top \mathbf{x} \geq \bar{\mathbf{x}}^\top \bar{\mathbf{x}}$ , wie benötigt.  $\square$

Proposition 2.12 lässt sich von einem einzelnen Punkt auf kompakte Mengen erweitern.

**Satz 2.13 (strikt Trennungssatz).** Seien  $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{R}^n$  nichtleere, abgeschlossene, konvexe Mengen und sei  $\mathcal{C}_2$  kompakt. Wenn  $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ , dann existiert eine Hyperebene, die  $\mathcal{C}_1$  und  $\mathcal{C}_2$  strikt trennt.



**Abbildung 2.2:** Die Mengen aus Beispiel 2.14 lassen sich nicht strikt trennen.

*Beweis.* Es ist  $\mathcal{C}_1 - \mathcal{C}_2 = \{c_1 - c_2 : c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2\}$  abgeschlossen: Für jede Folge  $(z_k)_{k \in \mathbb{N}} \subseteq \mathcal{C}_1 - \mathcal{C}_2$  mit Grenzwert  $z \in \mathbb{R}^n$  existieren Folgen  $(x_k)_{k \in \mathbb{N}} \subseteq \mathcal{C}_1$  und  $(y_k)_{k \in \mathbb{N}} \subseteq \mathcal{C}_2$  mit  $z_k = x_k - y_k$  für alle  $k \in \mathbb{N}$ . Da  $\mathcal{C}_2$  kompakt ist, existiert eine konvergente Teilfolge  $(y_{j_k})_{k \in \mathbb{N}} \subseteq (y_k)_{k \in \mathbb{N}}$  mit Grenzwert  $y \in \mathcal{C}_2$ . Es konvergiert auch  $(z_{j_k})_{k \in \mathbb{N}}$  gegen  $z$  und damit muss auch  $(x_{j_k})_{k \in \mathbb{N}}$  gegen  $x := z + y$  konvergieren, mit  $x \in \mathcal{C}_1$  wegen Abgeschlossenheit von  $\mathcal{C}_1$ . Es folgt, dass  $z = x - y \in \mathcal{C}_1 - \mathcal{C}_2$ , also ist  $\mathcal{C}_1 - \mathcal{C}_2$  abgeschlossen.

Nach Proposition 2.5 ist  $\mathcal{C}_1 - \mathcal{C}_2$  konvex. Da  $\mathcal{C}_1$  und  $\mathcal{C}_2$  disjunkt sind, ist außerdem  $\mathbf{0} \notin \mathcal{C}_1 - \mathcal{C}_2$ . Nach Proposition 2.12 existiert also eine Hyperebene

$$\{z \in \mathbb{R}^n : \mathbf{a}^\top z = \beta\},$$

die  $\mathcal{C}_1 - \mathcal{C}_2$  und  $\{\mathbf{0}\}$  strikt trennt, wobei wir das Vorzeichen von  $\mathbf{a}$  so wählen können, dass

$$\mathbf{a}^\top (x - y) > \beta > 0$$

für alle  $x \in \mathcal{C}_1$  und  $y \in \mathcal{C}_2$ . Daraus folgt

$$\inf_{x \in \mathcal{C}_1} \mathbf{a}^\top x \geq \sup_{y \in \mathcal{C}_2} \mathbf{a}^\top y + \beta > \sup_{y \in \mathcal{C}_2} \mathbf{a}^\top y.$$

Damit trennt die Hyperebene

$$\{x \in \mathbb{R}^n : \mathbf{a}^\top x = \gamma\}$$

mit  $\gamma := \sup_{y \in \mathcal{C}_2} \mathbf{a}^\top y + \frac{\beta}{2}$  die Mengen  $\mathcal{C}_1$  und  $\mathcal{C}_2$  strikt. □

Das folgende Beispiel zeigt, dass die Voraussetzung der Kompaktheit in Satz 2.13 notwendig für die strikte Trennung zweier Mengen ist.

**Beispiel 2.14.** Die Mengen  $\{(\frac{\alpha}{\beta}) \in \mathbb{R}^2 : \beta \leq 0\}$  und  $\{(\frac{\alpha}{\beta}) \in \mathbb{R}^2 : \beta \geq e^\alpha\}$  sind disjunkt, konvex und abgeschlossen, trotzdem ist keine strikte Trennung möglich (siehe Abb. 2.2). Die Mengen sind aber trennbar durch die Hyperebene  $\mathcal{H} = \{x \in \mathbb{R}^2 : x_2 = 0\}$ . △

Verzichten wir auf die Striktheit der Trennung, so benötigen wir keine Abgeschlossenheit mehr.

**Proposition 2.15.** Sei  $\mathcal{D} \subseteq \mathbb{R}^n$  eine kompakte Menge und  $\mathcal{F} \subseteq 2^{\mathcal{D}}, \mathcal{F} \neq \emptyset$ , eine Familie von abgeschlossenen Teilmengen von  $\mathcal{D}$ . Der Durchschnitt aller Mengen in  $\mathcal{F}$  ist nichtleer genau dann, wenn jeder endliche Durchschnitt von Mengen in  $\mathcal{F}$  nichtleer ist.

*Beweis.* Sei  $\bigcap_{M \in \mathcal{F}} M = \emptyset$  und bezeichne  $\neg M = \mathbb{R}^n \setminus M$  das Komplement von  $M \in \mathcal{F}$ . Dann ist  $\neg M$  eine offene Menge und (wegen  $\mathcal{F} \neq \emptyset$ )  $\bigcup_{M \in \mathcal{F}} \neg M = \mathbb{R}^n$  eine offene Überdeckung von  $\mathcal{D}$ . Da  $\mathcal{D}$  kompakt ist, existiert per Definition (siehe Analysis II) eine endliche Teilüberdeckung, also eine endliche Familie  $\mathcal{F}' \subseteq \mathcal{F}$  mit  $\mathcal{D} \subseteq \bigcup_{M \in \mathcal{F}'} \neg M = \neg \bigcap_{M \in \mathcal{F}'} M$ . Damit ist  $\bigcap_{M \in \mathcal{F}'} M = \emptyset$  ein leerer endlicher Durchschnitt.  $\square$

**Proposition 2.16.** Sei  $\mathcal{C} \subseteq \mathbb{R}^n$  konvex mit  $\mathbf{0} \notin \mathcal{C} \neq \emptyset$ . Dann existiert eine Hyperebene, die  $\mathcal{C}$  und  $\{\mathbf{0}\}$  trennt.

*Beweis.* Wir suchen  $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  mit  $\mathbf{a}^\top \mathbf{x} \geq 0$  für alle  $\mathbf{x} \in \mathcal{C}$ . Wir können uns auf  $\|\mathbf{a}\| = 1$  beschränken und definieren zunächst für alle  $\mathbf{x} \in \mathcal{C}$  die Menge

$$\mathcal{Y}(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}\| = 1, \mathbf{y}^\top \mathbf{x} \geq 0\}.$$

Sei nun  $\mathcal{X}$  die konvexe Hülle einer beliebigen endlichen Teilmenge  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq \mathcal{C}$ . Wegen Satz 2.3 und 2.7 sowie Proposition 2.10 ist  $\mathcal{X} \subseteq \mathcal{C}$  konvex und kompakt. Nach Proposition 2.12 und wegen  $\mathbf{0} \notin \mathcal{C}$  existiert also (nach Skalierung)  $\bar{\mathbf{y}} \in \mathbb{R}^n$  mit  $\|\bar{\mathbf{y}}\| = 1$  und  $\bar{\mathbf{y}}^\top \mathbf{x}_i > 0$  für alle  $i \in \{1, \dots, k\}$ . Damit ist

$$\bar{\mathbf{y}} \in \bigcap_{i=1}^k \mathcal{Y}(\mathbf{x}_i)$$

und dieser Schnitt also nichtleer. Da  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  beliebig gewählt war, ist also jeder endliche Schnitt aus Mengen in der Familie  $\{\mathcal{Y}(\mathbf{x})\}_{\mathbf{x} \in \mathcal{C}}$  nichtleer.

Die Mengen  $\mathcal{Y}(\mathbf{x})$  sind jeweils Teilmenge der kompakten Einheitssphäre  $\mathcal{D} = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}\| = 1\}$ . Außerdem ist  $\mathcal{Y}(\mathbf{x})$  die Schnittmenge von  $\mathcal{D}$  mit dem abgeschlossenen Halbraum  $\{\mathbf{y} \in \mathbb{R}^n : \mathbf{x}^\top \mathbf{y} \geq 0\}$ , also wieder abgeschlossen. Mit Proposition 2.15 folgt, dass

$$\bigcap_{\mathbf{x} \in \mathcal{C}} \mathcal{Y}(\mathbf{x}) \neq \emptyset.$$

Damit existiert  $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  mit  $\mathbf{a} \in \mathcal{Y}(\mathbf{x})$  für alle  $\mathbf{x} \in \mathcal{C}$ . Es folgt  $\mathbf{a}^\top \mathbf{x} \geq 0$  für alle  $\mathbf{x} \in \mathcal{C}$ . Also trennt die Hyperebene  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = 0\}$  die Menge  $\mathcal{C}$  und  $\{\mathbf{0}\}$ .  $\square$

Wir können damit einen schwächeren Trennungssatz zeigen, der weder die Abgeschlossenheit noch die Beschränktheit der zu trennenden Mengen benötigt.

**Satz 2.17.** Seien  $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{R}^n$  nichtleer und konvex. Wenn  $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ , dann existiert eine Hyperebene, die  $\mathcal{C}_1$  und  $\mathcal{C}_2$  trennt.

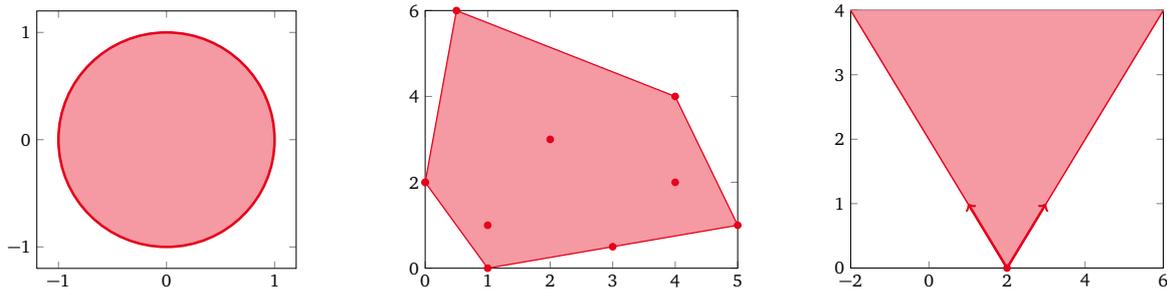


Abbildung 2.3: Die Mengen aus den Beispielen 2.19, 2.20 und 2.22.

*Beweis.* Nach Proposition 2.5 ist  $C_1 - C_2$  konvex. Da  $C_1$  und  $C_2$  disjunkt sind, ist außerdem  $0 \notin C_1 - C_2$ . Nach Proposition 2.16 existiert (ggf. durch Skalierung)  $\mathbf{a} \in \mathbb{R}^n$  mit  $\mathbf{a}^\top \mathbf{z} \geq 0$  für alle  $\mathbf{z} \in C_1 - C_2$ , beziehungsweise  $\mathbf{a}^\top (\mathbf{x} - \mathbf{y}) \geq 0$  für alle  $\mathbf{x} \in C_1$  und  $\mathbf{y} \in C_2$ . Daher können wir  $\beta \in \mathbb{R}$  finden, so dass

$$\inf_{\mathbf{x} \in C_1} \mathbf{a}^\top \mathbf{x} \geq \beta \geq \sup_{\mathbf{y} \in C_2} \mathbf{a}^\top \mathbf{y}.$$

Das heißt, dass die Hyperebene  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}$  die Mengen  $C_1$  und  $C_2$  trennt. □

### 2.3 Extrempunkte, Extremrichtungen, Stützeigenschaften

Wir werden sehen, dass wir uns zur Bestimmung von Extrema bestimmter (insbesondere linearer) Funktionen über einer konvexen Menge auf Punkte beschränken können, die sich nicht als Konvexkombination anderer Punkte ausdrücken lassen.

**Definition 2.18.** Ein Punkt  $\mathbf{x} \in C$  einer konvexen Menge  $C \subseteq \mathbb{R}^n$  heißt *Extrempunkt* von  $C$ , wenn er keine echte Konvexkombination verschiedener Punkte in  $C$  ist.

**Beispiel 2.19.** Die Extrempunkte der abgeschlossenen Kugel  $\overline{B}_1\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| \leq 1\}$  (konvex nach Proposition 2.5) sind alle Punkte ihres Randes  $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$  (siehe Abb. 2.3). Die offene Kugel  $B_1\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| < 1\}$  hat keine Extrempunkte. △

**Beispiel 2.20.** Die Extrempunkte der Menge  $\text{conv}\left(\left\{\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 6 \end{pmatrix}\right\}$  sind  $\left\{\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 6 \end{pmatrix}\right\}$  (siehe Abb. 2.3). △

In den obigen Beispielen sind die abgeschlossenen Mengen genau die konvexen Hüllen ihrer Extrempunkte. Der folgende Satz liefert eine hinreichende Bedingung für diese Eigenschaft.

**Satz 2.21.** Jede kompakte konvexe Menge  $C \subseteq \mathbb{R}^n$  ist die konvexe Hülle ihrer Extrempunkte.

*Beweis.* Sei  $\mathcal{K}$  die konvexe Hülle der Extrempunkte von  $C$ . Dann ist  $\mathcal{K} \subseteq C$ , da  $C$  konvex ist. Wir beweisen nun  $C \subseteq \mathcal{K}$  per Induktion über die Raumdimension  $n$ . Für  $n = 1$ , sind alle kompakten, konvexen Mengen abgeschlossene und beschränkte Intervalle, also die konvexe Hülle ihrer zwei Endpunkte, und damit  $C \subseteq \mathcal{K}$ .

Betrachte nun  $n \in \mathbb{N} \setminus \{1\}$ . Angenommen es gäbe  $\bar{\mathbf{x}} \in \mathcal{C} \setminus \mathcal{K}$ . Dann liefert Satz 2.17 ein  $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  und  $\beta \in \mathbb{R}$  mit  $\mathbf{a}^\top \bar{\mathbf{x}} \geq \beta \geq \mathbf{a}^\top \mathbf{x}$  für alle  $\mathbf{x} \in \mathcal{K}$ . Da  $\mathcal{C}$  kompakt ist, existiert

$$\beta^* := \max\{\mathbf{a}^\top \mathbf{x} : \mathbf{x} \in \mathcal{C}\} \geq \mathbf{a}^\top \bar{\mathbf{x}} \geq \beta.$$

Die Hyperebene  $\mathcal{H} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta^*\}$  ist ein affin linearer Teilraum der Dimension  $n - 1$ . Es ist  $\mathcal{C} \cap \mathcal{H}$  kompakt und nach Proposition 2.5 (ii) und (i) konvex. Betrachten wir also  $\mathcal{C} \cap \mathcal{H}$  als Objekt im affin linearen Teilraum  $\mathcal{H}$ , liefert uns die Induktion, dass  $\mathcal{C} \cap \mathcal{H} = \mathcal{K}'$  wobei  $\mathcal{K}'$  die konvexe Hülle der Extrempunkte von  $\mathcal{C} \cap \mathcal{H}$  bezeichnet. Per Definition von  $\mathcal{H}$  ist  $\mathcal{C} \cap \mathcal{H} \neq \emptyset$ , also auch  $\mathcal{K}' \neq \emptyset$ .

Wir behaupten, dass jeder Extrempunkt  $\mathbf{y} \in \mathcal{K}'$  von  $\mathcal{C} \cap \mathcal{H}$  auch ein Extrempunkt von  $\mathcal{C}$  ist. Das führt zu einem Widerspruch, denn es gilt  $\mathcal{C} \cap \mathcal{H} \not\subseteq \mathcal{K}$ : Für  $\beta^* > \beta$  ist sogar  $(\mathcal{C} \cap \mathcal{H}) \cap \mathcal{K} = \emptyset$  und für  $\beta^* = \beta$  ist  $\bar{\mathbf{x}} \in (\mathcal{C} \cap \mathcal{H}) \setminus \mathcal{K}$ . Es folgt, dass kein  $\bar{\mathbf{x}} \in \mathcal{C} \setminus \mathcal{K}$  existieren kann, bzw.  $\mathcal{C} \subseteq \mathcal{K}$ .

Sei nun also  $\mathbf{y}$  ein Extrempunkt von  $\mathcal{C} \cap \mathcal{H}$ . Ist  $\mathbf{y}$  kein Extrempunkt von  $\mathcal{C}$ , lässt es sich als echte Konvexkombination  $\mathbf{y} = \sum_{i=1}^p \lambda_i \mathbf{x}_i$  von Punkten  $\{\mathbf{x}_1, \dots, \mathbf{x}_p\} \subseteq \mathcal{C}$  mit  $\boldsymbol{\lambda} > \mathbf{0}$  und  $\mathbf{1}^\top \boldsymbol{\lambda} = 1$  darstellen. Da  $\mathbf{y}$  ein Extrempunkt von  $\mathcal{C} \cap \mathcal{H}$  ist, muss es  $\ell \in \{1, \dots, p\}$  geben mit  $\mathbf{x}_\ell \notin \mathcal{H}$ , bzw. mit  $\mathbf{a}^\top \mathbf{x}_\ell < \beta^*$ . Per Definition von  $\beta^*$  gilt

$$\mathbf{a}^\top \mathbf{y} = \lambda_\ell \mathbf{a}^\top \mathbf{x}_\ell + \sum_{i \in \{1, \dots, p\} \setminus \{\ell\}} \lambda_i \mathbf{a}^\top \mathbf{x}_i \leq \lambda_\ell \mathbf{a}^\top \mathbf{x}_\ell + \sum_{i \in \{1, \dots, p\} \setminus \{\ell\}} \lambda_i \beta^* \stackrel{\lambda_i > 0}{<} \beta^* \sum_{i=1}^p \lambda_i = \beta^*,$$

ein Widerspruch zu  $\mathbf{y} \in \mathcal{H}$ . Also ist  $\mathbf{y}$  ein Extrempunkt von  $\mathcal{C}$ , wie behauptet.  $\square$

Die Notwendigkeit der Abgeschlossenheit haben wir bereits in Beispiel 2.19 gesehen. Das folgende Beispiel zeigt die Notwendigkeit der Beschränktheit.

**Beispiel 2.22.** Die abgeschlossene konvexe Menge  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^2 : x_2 \geq |x_1 - 2|\}$  besteht nicht nur aus Konvexkombinationen ihres einzigen Extrempunkts  $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$  (siehe Abb. 2.3).  $\triangle$

Um den unbeschränkten Teil einer konvexen Menge beschreiben zu können, benötigen wir ihre Extremrichtungen.

**Definition 2.23.** Ein Vektor  $\mathbf{d} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  ist eine *Richtung* von  $\mathcal{D} \subseteq \mathbb{R}^n$ , wenn  $\mathbf{x} + \alpha \mathbf{d} \in \mathcal{D}$  für alle  $\mathbf{x} \in \mathcal{D}$  und  $\alpha \geq 0$ . Richtungen  $\mathbf{d}_1, \mathbf{d}_2$  heißen *verschieden*, wenn  $\mathbf{d}_1 \neq \beta \mathbf{d}_2$  für alle  $\beta > 0$ . Eine Richtung  $\mathbf{d}$  ist eine *Extremrichtung*, wenn  $\mathbf{d} \neq \mathbf{d}_1 + \mathbf{d}_2$  für alle verschiedenen Richtungen  $\mathbf{d}_1, \mathbf{d}_2$ .

**Beispiel 2.24.** Die Richtungen der Menge  $\mathcal{C}$  aus Beispiel 2.22 (siehe Abb. 2.3) sind  $\{(\frac{\alpha}{\beta}) \neq \mathbf{0} : \beta \geq |\alpha| \geq 0\}$ . Die Extremrichtungen von  $\mathcal{C}$  sind  $(\frac{-\alpha}{\alpha})$  und  $(\frac{\alpha}{\alpha})$  für alle  $\alpha > 0$ .  $\triangle$

**Beispiel 2.25.** Lineare Teilräume  $\mathcal{U} \subseteq \mathbb{R}^n$  haben keine Extremrichtungen: Jedes  $\mathbf{d} \in \mathcal{U} \setminus \{\mathbf{0}\}$  ist eine Richtung von  $\mathcal{U}$  und genauso  $\mathbf{d}' = -\mathbf{d}$ . Daher ist  $\mathbf{d} = 2\mathbf{d} - \mathbf{d}'$  eine Linearkombination verschiedener Richtungen aus  $\mathcal{U}$ .  $\triangle$

Die Beschreibung einer konvexen Menge durch ihre Extrempunkte und -richtungen nennt man ihre „innere“ Darstellung. Alternativ lässt sich eine konvexe Menge als Schnitt von Halbräumen, ihrer „äußeren“ Darstellung, beschreiben. Weitere Einzelheiten zu diesen Darstellungen und der Umrechnung werden in der Vorlesung *Diskrete Optimierung* behandelt.

**Definition 2.26.** Es heißt  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}$ , mit  $\mathbf{a} \in \mathbb{R}^n \setminus \{0\}$  und  $\beta \in \mathbb{R}$ , *Stützhyperebene* von  $\mathcal{D} \subseteq \mathbb{R}^n$  mit zugehörigem *Stützhalbraum*  $\mathcal{H}^\leq = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\}$  und *äußerer Normale*  $\mathbf{a}$ , wenn

$$\mathcal{D} \cap \mathcal{H} \neq \emptyset \quad \text{und} \quad \mathcal{D} \subseteq \mathcal{H}^\leq.$$

In Definition 2.26 ist die Bezeichnung mit ' $\leq$ ' willkürlich, da  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : -\mathbf{a}^\top \mathbf{x} = -\beta\}$ .

Wir zeigen nun, dass kompakte konvexe Mengen in jeder Richtung durch eine Hyperebene beschränkt werden können.

**Satz 2.27.** Für alle kompakten Mengen  $\mathcal{D} \subseteq \mathbb{R}^n$ ,  $\mathcal{D} \neq \emptyset$  und alle  $\mathbf{a} \in \mathbb{R}^n \setminus \{0\}$  existiert eine Stützhyperebene von  $\mathcal{D}$  mit äußerer Normale  $\mathbf{a}$ .

*Beweis.* Es ist  $\mathcal{D} \neq \emptyset$  kompakt und  $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$  stetig, daher existiert  $\beta = \max\{\mathbf{a}^\top \mathbf{y} : \mathbf{y} \in \mathcal{D}\}$ . Damit ist  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}$  Stützhyperebene von  $\mathcal{D}$ .  $\square$

**Satz 2.28.** Jede abgeschlossene, konvexe Menge  $\mathcal{C} \subsetneq \mathbb{R}^n$ ,  $\mathcal{C} \neq \emptyset$  ist die Schnittmenge ihrer Stützhalbräume.

*Beweis.* Wir zeigen zunächst, dass es für alle  $\bar{\mathbf{x}} \in \mathbb{R}^n \setminus \mathcal{C}$  einen Stützhalbraum von  $\mathcal{C}$  gibt, der  $\bar{\mathbf{x}}$  nicht enthält. Sei  $\alpha > 0$  so gewählt, dass  $\mathcal{C} \cap \bar{\mathcal{B}}_\alpha(\bar{\mathbf{x}}) \neq \emptyset$  mit  $\bar{\mathcal{B}}_\alpha(\bar{\mathbf{x}}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \alpha\}$ . Wegen Abgeschlossenheit von  $\mathcal{C}$  ist  $\mathcal{C} \cap \bar{\mathcal{B}}_\alpha(\bar{\mathbf{x}})$  kompakt und es existiert

$$\bar{\mathbf{y}} \in \arg \min\{\|\bar{\mathbf{x}} - \mathbf{x}\| : \mathbf{x} \in \mathcal{C}\} = \arg \min\{\|\bar{\mathbf{x}} - \mathbf{x}\| : \mathbf{x} \in (\mathcal{C} \cap \bar{\mathcal{B}}_\alpha(\bar{\mathbf{x}}))\}.$$

Sei  $\alpha^* = \|\bar{\mathbf{x}} - \bar{\mathbf{y}}\|$ . Dann ist  $\mathcal{C} \cap \mathcal{B}_{\alpha^*}(\bar{\mathbf{x}}) = \emptyset$  mit  $\mathcal{B}_{\alpha^*}(\bar{\mathbf{x}}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \bar{\mathbf{x}}\| < \alpha^*\}$ . Nach Satz 2.17 existieren also  $\mathbf{a} \in \mathbb{R}^n \setminus \{0\}$  und  $\beta \in \mathbb{R}$  mit  $\mathbf{a}^\top \mathbf{x} \leq \beta \leq \mathbf{a}^\top \bar{\mathbf{y}}$  für alle  $\mathbf{x} \in \mathcal{C}$  und  $\bar{\mathbf{y}} \in \mathcal{B}_{\alpha^*}(\bar{\mathbf{x}})$ . Da  $\bar{\mathbf{y}} \in \partial \mathcal{B}_{\alpha^*}(\bar{\mathbf{x}})$  existiert eine Folge  $(\bar{\mathbf{y}}_k)_{k \in \mathbb{N}} \subseteq \mathcal{B}_{\alpha^*}(\bar{\mathbf{x}})$  mit Grenzwert  $\bar{\mathbf{y}}$ . Also gilt wegen Stetigkeit von  $\mathbf{a}^\top \mathbf{x}$ , dass

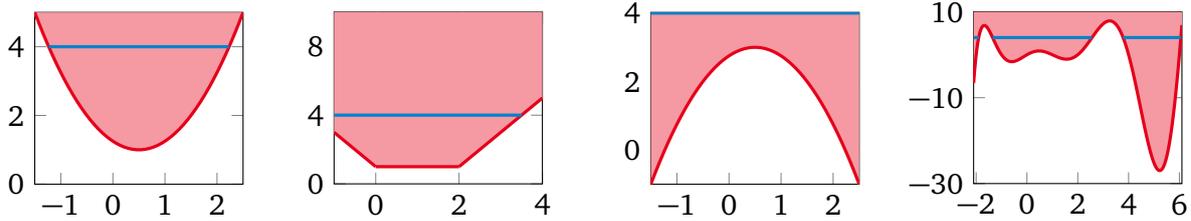
$$\mathbf{a}^\top \bar{\mathbf{y}} \leq \beta \leq \lim_{k \rightarrow \infty} \mathbf{a}^\top \bar{\mathbf{y}}_k = \mathbf{a}^\top \bar{\mathbf{y}},$$

also  $\mathbf{a}^\top \bar{\mathbf{y}} = \beta$ . Damit ist  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\} \supseteq \mathcal{C}$  der gewünschte Stützhalbraum, da  $\bar{\mathbf{y}} \in (\mathcal{C} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}) \neq \emptyset$ .

Aus  $\mathcal{C} \neq \mathbb{R}^n$  folgt jetzt die Existenz eines Stützhalbraums von  $\mathcal{C}$ . Wir können also die Schnittmenge  $\mathcal{G}$  aller Stützhalbräume von  $\mathcal{C}$  betrachten. Es ist  $\mathcal{C} \subseteq \mathcal{G}$ , da jeder Stützhalbraum per Definition  $\mathcal{C}$  enthält. Andersherum existiert nach unserer Vorüberlegung für alle  $\bar{\mathbf{x}} \in \mathbb{R}^n \setminus \mathcal{C}$  ein Stützhalbraum von  $\mathcal{C}$ , der  $\bar{\mathbf{x}}$  nicht enthält. Damit ist  $\bar{\mathbf{x}} \notin \mathcal{G}$  und somit  $\mathcal{G} \subseteq \mathcal{C}$ .  $\square$

## 2.4 Konvexe Funktionen

Eine konvexe Funktion ist eine Funktion, deren Interpolation zwischen Funktionswerten immer linear beschränkt werden kann. Anschaulich heißt das, dass die Verbindungsstrecke zwischen zwei Punkten auf dem Funktionsgraphen stets oberhalb der Funktion liegt. Wir geben die folgende formale Definition (siehe Abb. 2.4 für Beispiele).



**Abbildung 2.4:** Von links nach rechts: strikt konvex, konvex, strikt konkav, weder konvex noch konkav. Der Epigraph  $\mathcal{E}(f)$  ist ausgefüllt und die Niveaumenge  $\mathcal{L}(f, 4)$  ist blau.

**Definition 2.29.** Sei  $C \subseteq \mathbb{R}^n$  konvex. Eine Funktion  $f : C \rightarrow \mathbb{R}$  heißt *konvex*, wenn

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \forall x \in C, y \in C \setminus \{x\}, \lambda \in (0, 1),$$

und *strikt konvex*, wenn die Ungleichung für strikt gilt. Die Funktion  $f$  heißt *(strikt) konkav*, wenn  $-f$  (strikt) konvex ist.

**Beispiel 2.30.** Affin lineare Funktionen  $c^\top x + \gamma$  sind konvex und konkav. △

Ähnlich wie bei der Konvexität von Mengen, können wir die Konvexität einer Funktion auch über beliebige Konvexkombinationen von Argumenten und Funktionswerten charakterisieren.

**Beobachtung 2.31** (Übung). Sei  $f : C \rightarrow \mathbb{R}$  mit  $C \subseteq \mathbb{R}^n$  konvex. Dann ist  $f$  konvex genau dann, wenn für alle  $x_1, \dots, x_p \in C$  und  $\lambda_1, \dots, \lambda_p \geq 0$  mit  $p \in \mathbb{N}$  und  $\sum_{i=1}^p \lambda_i = 1$  gilt

$$f\left(\sum_{i=1}^p \lambda_i x_i\right) \leq \sum_{i=1}^p \lambda_i f(x_i).$$

Wir geben grundlegende Eigenschaften von konvexen Funktionen. Zu beachten ist, dass Differenzen, Produkte und Minima konvexer Funktionen im Allgemeinen nicht wieder konvex sind.

**Proposition 2.32** (Übung). Seien  $C \subseteq \mathbb{R}^n$  und  $f_1, f_2 : C \rightarrow \mathbb{R}$  jeweils konvex und sei  $\alpha > 0$ . Dann sind auch  $\alpha f_1$ ,  $f_1 + f_2$  und  $\max\{f_1, f_2\}$  konvex.

Wir wollen nun die Konvexität einer Funktion mit der Konvexität zugehöriger Mengen in Beziehung setzen. Dazu definieren wir Mengen, die es erlauben unsere Anschauung über die Geometrie des Funktionsgraphen auszudrücken (siehe Abb. 2.4).

**Definition 2.33.** Sei  $f : D \rightarrow \mathbb{R}$  mit  $D \subseteq \mathbb{R}^n$ . Der *Epigraph* von  $f$  ist gegeben durch

$$\mathcal{E}(f) := \{(x, \alpha) \in D \times \mathbb{R} : f(x) \leq \alpha\}.$$

Die (*untere*) *Niveaumenge* von  $f$  zum Niveau  $\beta \in \mathbb{R}$  ist gegeben durch

$$\mathcal{L}(f, \beta) = \{x \in D : f(x) \leq \beta\}.$$

Wir können jetzt die Konvexität einer Funktion auf die Konvexität dieser Mengen zurückführen. Zu beachten ist, dass konvexe Niveaumengen nur eine notwendige Bedingung für Konvexität einer Funktion sind. Funktionen, die dieser Bedingung genügen, nennt man *quasikonvex*.

**Satz 2.34** (Übung). Für  $f : \mathcal{C} \rightarrow \mathbb{R}$  mit  $\mathcal{C}$  konvex gilt:

- (i) Es ist  $f$  konvex genau dann, wenn  $\mathcal{E}(f)$  konvex ist.
- (ii) Ist  $f$  konvex, dann ist auch  $\mathcal{L}(f, \beta)$  konvex für alle  $\beta \in \mathbb{R}$ .

Eine wichtige Eigenschaft von konvexen Mengen sei hier genannt.

**Satz 2.35** (Übung). Sind  $\mathcal{C} \subseteq \mathbb{R}^n$  und  $f : \mathcal{C} \rightarrow \mathbb{R}$  konvex, so ist  $f$  im Inneren von  $\mathcal{C}$  stetig.

Allerdings sind konvexe Funktionen im Allgemeinen nicht differenzierbar, wie man zum Beispiel an  $f(x) = |x|$  sieht.

Wir zeigen jetzt die anschauliche Charakterisierung von Konvexität für differenzierbare Funktionen, dass die Funktion oberhalb jeder Tangente liegt.

**Satz 2.36.** Sei  $f : \mathcal{C} \rightarrow \mathbb{R}$  mit  $\mathcal{C}$  konvex und  $f \in C^1$ . Dann ist  $f$  genau dann konvex, wenn

$$f(\mathbf{y}) \geq f(\mathbf{x}) + (\mathbf{y} - \mathbf{x})^\top \nabla f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{C}, \mathbf{y} \in \mathcal{C} \setminus \{\mathbf{x}\} \quad (2.4)$$

und genau dann strikt konvex, wenn (2.4) strikt gilt.

*Beweis.* Sei  $f$  konvex und seien  $\mathbf{x} \in \mathcal{C}$ ,  $\mathbf{z} \in \mathcal{C} \setminus \{\mathbf{x}\}$  beliebig gewählt. Betrachte  $h : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$h(\lambda) = (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{z}) - f((1 - \lambda)\mathbf{x} + \lambda\mathbf{z}).$$

Da  $f$  konvex ist, gilt  $h(\lambda) \geq 0$  für alle  $\lambda \in [0, 1]$ . Wegen  $h(0) = 0$  folgt

$$0 \leq \frac{dh}{d\lambda}(0) = -f(\mathbf{x}) + f(\mathbf{z}) - (\mathbf{z} - \mathbf{x})^\top \nabla f(\mathbf{x}),$$

und damit (2.4).

Ist  $f$  strikt konvex, gilt unter Verwendung von (2.4) mit  $\mathbf{y} = \frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{z}$  sogar

$$\begin{aligned} f(\mathbf{z}) &= 2 \cdot \left[ \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{z}) - \frac{1}{2}f(\mathbf{x}) \right] \\ &> 2 \cdot \left[ f\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{z}\right) - \frac{1}{2}f(\mathbf{x}) \right] \\ &\stackrel{(2.4)}{\geq} 2 \cdot \left[ f(\mathbf{x}) + \left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{z} - \mathbf{x}\right)^\top \nabla f(\mathbf{x}) - \frac{1}{2}f(\mathbf{x}) \right] \\ &= f(\mathbf{x}) + (\mathbf{z} - \mathbf{x})^\top \nabla f(\mathbf{x}), \end{aligned}$$

und damit gilt (2.4) strikt.

Für die Rückrichtung seien  $\mathbf{x} \in \mathcal{C}$ ,  $\mathbf{y} \in \mathcal{C} \setminus \{\mathbf{x}\}$  und  $\mathbf{z} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}$  mit  $\lambda \in (0, 1)$ . Da  $\mathcal{C}$  konvex ist, ist  $\mathbf{z} \in \mathcal{C}$  und damit

$$\begin{aligned}
\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) &\stackrel{(2.4)}{\geq} \lambda[f(\mathbf{z}) + (\mathbf{x}-\mathbf{z})^\top \nabla f(\mathbf{z})] + (1-\lambda)[f(\mathbf{z}) + (\mathbf{y}-\mathbf{z})^\top \nabla f(\mathbf{z})] \\
&= f(\mathbf{z}) + [\lambda\mathbf{x} + (1-\lambda)\mathbf{y} - \mathbf{z}]^\top \nabla f(\mathbf{z}) \\
&= f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}).
\end{aligned}$$

Also ist  $f$  konvex. Gilt (2.4) strikt, dann ist  $f$  nach der gleichen Rechnung strikt konvex.  $\square$

Für zweifach stetig differenzierbare Funktionen auf ganz  $\mathbb{R}^n$ , können wir konvexe Funktionen über die intuitive Eigenschaft charakterisieren, dass sie nichtnegativ gekrümmt sind.

**Satz 2.37.** Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $f \in C^2$  ist genau dann konvex, wenn die Hesse-Matrix  $H_f(\mathbf{x}) := \left( \partial_i \partial_j f(\mathbf{x}) \right)_{i,j \in \{1, \dots, n\}}$  für alle  $\mathbf{x} \in \mathbb{R}^n$  positiv semidefinit ist.

*Beweis.* Nach dem Satz von Taylor (siehe *Analysis II*) gilt für alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , dass

$$\exists t \in [0, 1]: \quad f(\mathbf{y}) = f(\mathbf{x}) + (\mathbf{y} - \mathbf{x})^\top \nabla f(\mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^\top H_f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}). \quad (2.5)$$

Ist  $H_f(\mathbf{x})$  für alle  $\mathbf{x}$  positiv semidefinit, so ist der letzte Summand in (2.5) nichtnegativ. Mit Satz 2.36 folgt, dass  $f$  konvex ist.

Ist andersherum  $H_f(\bar{\mathbf{x}})$  nicht positiv semidefinit in  $\bar{\mathbf{x}} \in \mathbb{R}^n$ , so existiert  $\mathbf{y} \in \mathbb{R}^n$  mit

$$(\mathbf{y} - \bar{\mathbf{x}})^\top H_f(\bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{x}}) < 0.$$

Wegen der Stetigkeit von  $H_f(\mathbf{x})$  können wir  $\mathbf{y}$  nah genug an  $\bar{\mathbf{x}}$  wählen, dass

$$(\mathbf{y} - \bar{\mathbf{x}})^\top H_f(\bar{\mathbf{x}} + t(\mathbf{y} - \bar{\mathbf{x}}))(\mathbf{y} - \bar{\mathbf{x}}) < 0, \quad \forall t \in [0, 1].$$

Aus (2.5) und Satz 2.36 folgt, dass  $f$  auf  $\mathbb{R}^n$  nicht konvex ist.  $\square$

**Korollar 2.38.** Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $f \in C^2$ . Ist  $H_f(\mathbf{x})$  positiv definit für alle  $\mathbf{x} \in \mathbb{R}^n$ , so ist  $f$  strikt konvex. Die Umkehrung gilt nicht.

*Beweis.* Für die Rückwärtsrichtung beweisen wir analog, dass positive Definitheit von  $H_f(\mathbf{x})$  mit dem Satz von Taylor impliziert, dass (2.4) strikt gilt und  $f$  damit strikt konvex ist. Die Vorwärtsrichtung folgt nicht analog, da das Stetigkeitsargument einen Abstand zu 0 braucht. Die Umkehrung gilt auch nicht, wie wir anhand der konvexen, eindimensionalen Funktion  $f(x) = x^4$  einfach am Punkt 0 sehen können.  $\square$

**Beispiel 2.39.** Eine quadratische Funktion der Form  $f(\mathbf{x}) = \alpha + \mathbf{c}^\top \mathbf{x} + \frac{1}{2} \mathbf{x}^\top Q \mathbf{x}$  mit symmetrischer Matrix  $Q$  hat die Hesse-Matrix  $H_f(\mathbf{x}) = Q$  und ist daher konvex (bzw. konkav) genau dann, wenn  $Q$  positiv (bzw. negativ) semidefinit ist. Insbesondere sind alle quadratischen Funktionen über  $\mathbb{R}$  konvex und/oder konkav.  $\triangle$

Im Allgemeinen gibt es quadratische Funktionen, die weder konvex noch konkav sind.

**Beispiel 2.40.** Die Funktion  $f\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) = \alpha^2 + \beta^2 - 4\alpha\beta$  hat die indefinite Hesse-Matrix

$$H_f\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right) = \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}$$

mit Eigenwerten  $-2$  und  $6$ . Daher ist  $f$  weder konvex noch konkav.  $\triangle$

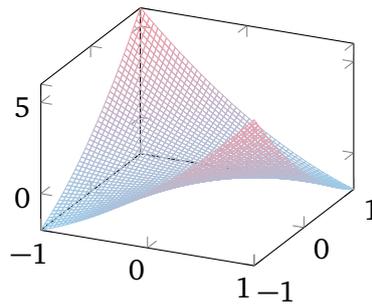


Abbildung 2.5: Die Funktion aus Beispiel 2.40.

## 2.5 Optimalitätsresultate für konvexe Optimierung

Wir verwenden nun unsere strukturellen Erkenntnisse aus den vorigen Abschnitten um Optimalitätskriterien für konvexe Funktionen abzuleiten. Wir beobachten zunächst, dass konvexe Funktionen nur globale Minima haben können.

**Satz 2.41.** Seien  $C \subseteq \mathbb{R}^n$  und  $f: C \rightarrow \mathbb{R}$  konvex. Dann ist jedes lokale Minimum von  $f$  über  $C$  bereits ein globales Minimum.

*Beweis.* Sei  $\bar{x} \in C$  ein lokaler Minimalpunkt von  $f$  über  $C$  und sei  $x \in C$  beliebig. Wegen lokaler Minimalität von  $\bar{x}$  existiert dann  $\lambda > 0$  mit  $f(\bar{x}) \leq f(\bar{x} + \lambda(x - \bar{x}))$ . Mit Konvexität von  $f$  folgt

$$f(\bar{x}) \leq f(\lambda x + (1 - \lambda)\bar{x}) \leq \lambda f(x) + (1 - \lambda)f(\bar{x}),$$

und damit  $f(\bar{x}) \leq f(x)$ . Da dies für beliebige  $x$  gilt, ist  $\bar{x}$  ein globaler Minimalpunkt.  $\square$

Weiterhin zeigen wir, dass die Menge der Minimalpunkte einer konvexen Funktion konvex ist und dass strikt konvexe Funktionen höchstens einen Minimalpunkt haben.

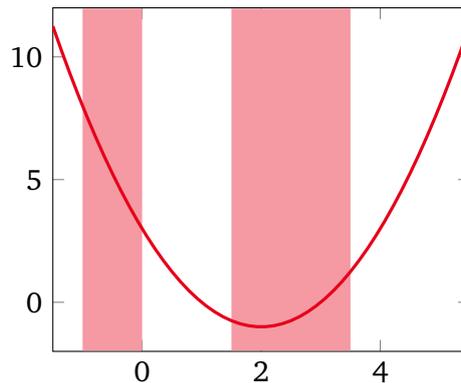
**Satz 2.42 (Übung).** Seien  $C \subseteq \mathbb{R}^n$  und  $f: C \rightarrow \mathbb{R}$  konvex. Dann ist  $\arg \min_{x \in C} f(x)$  konvex.

**Korollar 2.43.** Sei  $C \subseteq \mathbb{R}^n$  konvex und  $f: C \rightarrow \mathbb{R}$  strikt konvex. Nimmt  $f$  sein Minimum über  $C$  an, dann in einem eindeutigen Punkt.

*Beweis.* Angenommen es gäbe zwei verschiedene Punkte  $x, y \in \arg \min_{x' \in C} f(x')$  und sei  $z = \frac{1}{2}x + \frac{1}{2}y$ . Wegen strikter Konvexität von  $f$  ist  $f(z) < \frac{1}{2}f(x) + \frac{1}{2}f(y)$ , aber Satz 2.42 liefert  $f(z) = f(x) = f(y)$  – ein Widerspruch.  $\square$

Für konvexe Funktionen auf ganz  $\mathbb{R}^n$  ist die notwendige Bedingung, dass der Gradient an einem Minimalpunkt verschwindet, bereits hinreichend.

**Satz 2.44.** Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  differenzierbar und konvex. Dann ist  $\bar{x} \in \mathbb{R}^n$  ein globaler Minimalpunkt von  $f$  über  $\mathbb{R}^n$  genau dann, wenn  $\nabla f(\bar{x}) = \mathbf{0}$ .



**Abbildung 2.6:** Die Funktion  $x^2 - 4x + 3$  aus Beispiel 2.45.

*Beweis.* Es gilt  $\nabla f(\mathbf{x}) = \mathbf{0}$  für alle lokalen Extrema einer differenzierbaren Funktion  $f$  auf  $\mathbb{R}^n$ , also insbesondere  $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$  wenn  $\bar{\mathbf{x}}$  ein globaler Minimalpunkt von  $f$  ist.

Sei andersherum  $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$  und  $f$  konvex. Nach Satz 2.36 gilt

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + (\mathbf{x} - \bar{\mathbf{x}})^\top \nabla f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}}) \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

also ist  $\bar{\mathbf{x}}$  ein globaler Minimalpunkt. □

Das folgende Beispiel illustriert die obigen Optimalitätskriterien für die konvexe Minimierung. Es zeigt außerdem, dass es sehr wohl lokale Maximalpunkte einer konvexen Funktion geben kann.

**Beispiel 2.45.** Die eindimensionale Funktion  $f(x) = x^2 - 4x + 3$  nimmt ihr Minimum über  $\mathbb{R}$  im Punkt  $\bar{x} = 2$  mit  $\nabla f(\bar{x}) = 0$  an (siehe Abb. 2.6). Das Minimum über dem Intervall  $[-1, 0]$  wird am Rand des Intervalls angenommen, das Minimum über  $[1.5, 3.5]$  nicht. Das Maximum über beiden Intervallen wird an ihrem Rand angenommen und das Intervall  $[1.5, 3.5]$  besitzt ein lokales Maximum bei  $x = 1.5$ . △

Zwar kann es lokale Maximalpunkte einer konvexen Funktion geben, dafür können wir uns bei der Suche nach globalen Maxima auf Extrempunkte beschränken. Wie das vorige Beispiel zeigt, ist das für Minima nicht der Fall.

**Satz 2.46.** Seien  $\mathcal{C} \subseteq \mathbb{R}^n$  und  $f: \mathcal{C} \rightarrow \mathbb{R}$  konvex und  $\mathcal{C}$  kompakt. Nimmt  $f$  ihr Maximum über  $\mathcal{C}$  an, dann auch in einem Extrempunkt von  $\mathcal{C}$ .

*Beweis.* Sei  $\bar{\mathbf{x}} \in \mathcal{C}$  ein Maximalpunkt von  $f$  über  $\mathcal{C}$ . Nach Satz 2.21 und Satz 2.8 können wir  $\bar{\mathbf{x}}$  als Konvexkombination  $\bar{\mathbf{x}} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i$  mit  $\lambda \geq 0$  und  $\mathbf{1}^\top \lambda = 1$  darstellen, wobei  $\mathbf{x}_1, \dots, \mathbf{x}_{n+1} \in \mathcal{C}$  Extrempunkte von  $\mathcal{C}$  sind. Da  $f$  konvex ist, folgt mit Beobachtung 2.31

$$f(\bar{\mathbf{x}}) \leq \sum_{i=1}^{n+1} \lambda_i f(\mathbf{x}_i) \leq \max_{i \in \{1, \dots, n+1\}} f(\mathbf{x}_i) \cdot \sum_{i=1}^{n+1} \lambda_i = \max_{i \in \{1, \dots, n+1\}} f(\mathbf{x}_i).$$

Also existiert ein Extrempunkt  $\mathbf{x}_i$ ,  $i \in \{1, \dots, n+1\}$  in dem das Maximum angenommen wird. □

---

Besonders wichtig für uns wird sein, dass affin lineare Funktionen die Vorzüge der Minimierung und Maximierung konvexer Funktionen bezüglich aller Extrema vereinen.

**Korollar 2.47.** Jede affin lineare Funktion über jeder kompakten und konvexen Menge hat nur globale Extrema und nimmt sowohl ihr Maximum als auch ihr Minimum in einem Extrempunkt an.

*Beweis.* Sei  $f$  eine affin lineare Funktion. Da  $f$  stetig ist, nimmt sie ihr Maximum und Minimum über jeder kompakten Menge  $\mathcal{C}$  an. Außerdem sind  $f$  und  $-f$  konvex. Wegen

$$\arg \min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{C}\} = \arg \max\{-f(\mathbf{x}) : \mathbf{x} \in \mathcal{C}\}$$

impliziert Satz 2.46, dass  $f$  ihr Maximum und Minimum in einem Extrempunkt annimmt. Genauso impliziert Satz 2.41, dass nur globale Extrema existieren.  $\square$

Wir machen zuletzt noch die wichtige Beobachtung, dass wir das Maximum einer konvexen Funktion über einer kompakten Menge auf der konvexen Hülle dieser Menge suchen können.

**Satz 2.48** (Übung). Ist  $\mathcal{M} \subseteq \mathbb{R}^n$  kompakt und  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  konvex, so nimmt  $f$  ihr Maximum über  $\mathcal{M}$  an und es gilt

$$\max\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{M}\} = \max\{f(\mathbf{x}) : \mathbf{x} \in \text{conv}(\mathcal{M})\}.$$

**Beobachtung 2.49** (Bonusaufgabe). Ist  $\mathcal{M} \subseteq \mathbb{R}^n$  kompakt und  $f : \mathcal{M} \rightarrow \mathbb{R}$  konvex, so nimmt  $f$  nicht unbedingt ihr Maximum über  $\mathcal{M}$  an.

---

## 3 Polyeder

Für die Maximierung einer konvexen Funktion (auch für die Minimierung einer affinen Funktion) müssen wir nur die Extrempunkte der zulässigen Menge betrachten (Satz 2.46 und Korollar 2.47). Im Allgemeinen können das natürlich unendlich viele Punkte sein, weshalb wir zusätzliche Struktur brauchen, um auf diese Weise zu einem effizienten Lösungsverfahren zu kommen.

Aus Satz 2.28 und Proposition 2.5 wissen wir, dass eine Teilmenge des  $\mathbb{R}^n$  genau dann abgeschlossen und konvex ist, wenn sie die Schnittmenge von abgeschlossenen Halbräumen ist. Wir interessieren uns besonders für den Fall, in dem bereits eine endliche Anzahl an Halbräumen für die Beschreibung ausreicht, da dann auch die Anzahl der Extrempunkte endlich ist.

**Definition 3.1.** Ein *Polyeder* ist eine Schnittmenge endlich vieler abgeschlossener Halbräume, oder ganz  $\mathbb{R}^n$ . Ein *Polytop* ist ein beschränktes Polyeder.

Insbesondere haben Polyeder also die folgenden Eigenschaften.

**Beobachtung 3.2.** Jedes Polyeder ist abgeschlossen und konvex.

**Beobachtung 3.3.** Eine Menge  $\mathcal{P} \subseteq \mathbb{R}^n$  ist ein Polyeder genau dann, wenn  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  existieren, so dass

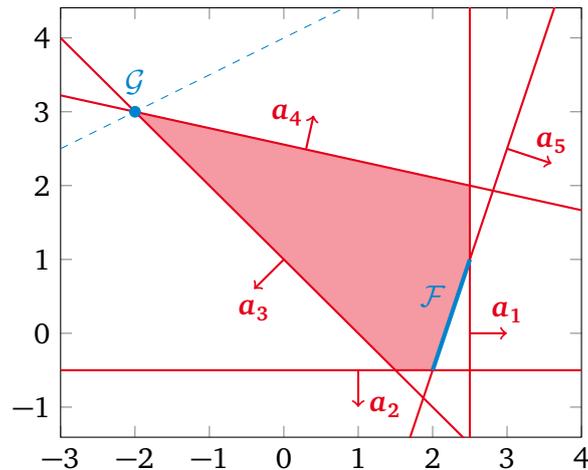
$$\mathcal{P} = \mathcal{P}(A, \mathbf{b}) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}.$$

Notation: Für eine Matrix  $A \in \mathbb{R}^{m \times n}$  bzw. einen Vektor  $\mathbf{b} \in \mathbb{R}^m$  mit Zeilenindexmenge  $M \subset \mathbb{N}$ ,  $|M| = m$  und Spaltenindexmenge  $N \subset \mathbb{N}$ ,  $|N| = n$  schreiben wir  $A \in \mathbb{R}^{M \times N}$  bzw.  $\mathbf{b} \in \mathbb{R}^M$ . Weiter bezeichnet  $A_i \in \mathbb{R}^N$  den Zeilenvektor von  $A$  mit Index  $i \in M$  (nicht notwendigerweise die  $i$ -te Zeile) und  $A_I \in \mathbb{R}^{I \times N}$  die aus den Zeilen  $I \subseteq M$  gebildete Teilmatrix. Analog bezeichnet  $A_j \in \mathbb{R}^M$  den Spaltenvektor von  $A$  mit Index  $j \in N$  und  $A_{J} \in \mathbb{R}^{M \times J}$  die aus den Spalten  $J \subseteq N$  gebildete Teilmatrix.

**Beispiel 3.4.** Das Polyeder  $\mathcal{P}(A, \mathbf{b})$  mit

$$A = \begin{pmatrix} 2 & 0 \\ 0 & -2 \\ -1 & -1 \\ 2 & 9 \\ 6 & -2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 5 \\ 1 \\ -1 \\ 23 \\ 13 \end{pmatrix}$$

ist in Abb. 3.1 dargestellt. Es ist beschränkt, also ein Polytop, daher ist es nach Beobachtung 3.2 und Satz 2.21 die konvexe Hülle seiner Extrempunkte  $\begin{pmatrix} -2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1.5 \\ -0.5 \end{pmatrix}, \begin{pmatrix} 2 \\ -0.5 \end{pmatrix}, \begin{pmatrix} 2.5 \\ 1 \end{pmatrix}, \begin{pmatrix} 2.5 \\ 2 \end{pmatrix}$ .  $\triangle$



**Abbildung 3.1:** Das Polytop aus Beispiel 3.4, wobei  $\mathbf{a}_i$  jeweils eine äußere Normale des Stützhalfbraums  $\{x \in \mathbb{R}^2 : A_i x \leq b_i\}$  bezeichnet. Die Teilmengen  $\mathcal{F}$  und  $\mathcal{G}$  sind Seitenflächen und  $\{x \in \mathbb{R}^2 : 2x_2 - x_1 = 8\}$  ist eine Stützhyperebene.

**Beobachtung 3.5.** Die Menge von Lösungen  $\begin{pmatrix} x \\ y \end{pmatrix}$  eines Systems

$$\begin{aligned} Bx + Cy &= c \\ Dx + Ey &\leq d \\ x &\geq \mathbf{0} \\ x &\in \mathbb{R}^p \\ y &\in \mathbb{R}^q \end{aligned}$$

ist ein Polyeder, mit  $B \in \mathbb{R}^{m \times p}$ ,  $C \in \mathbb{R}^{m \times q}$ ,  $D \in \mathbb{R}^{k \times p}$ ,  $E \in \mathbb{R}^{k \times q}$ ,  $c \in \mathbb{R}^m$ ,  $d \in \mathbb{R}^k$  und  $m, k, q, p \in \mathbb{N}$ .

*Beweis.* Die Lösungsmenge des Systems ist  $\mathcal{P}(A, \mathbf{b})$  mit

$$A = \begin{pmatrix} B & C \\ -B & -C \\ D & E \\ -\mathbb{I} & \mathbb{0} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} c \\ -c \\ d \\ \mathbf{0} \end{pmatrix},$$

wobei  $\mathbb{I}$  die Einheitsmatrix und  $\mathbb{0}$  die Nullmatrix bezeichnet. □

Wir haben also ein System von Gleichungen und Ungleichungen in die Form  $Ax \leq \mathbf{b}$  gebracht. Insbesondere haben wir dabei Gleichungen und Nichtnegativitätsbedingungen eliminiert. Wir können andersherum ein Ungleichungssystem in ein Gleichungssystem mit Nichtnegativitätsbedingungen transformieren.

**Beobachtung 3.6.** Die Lösungsmenge eines Systems von (Un)gleichungen ist invariant unter folgenden Transformationen (bzgl. der ursprünglichen Variablen):

(i) Ersetze eine Ungleichung der Form

$$\mathbf{a}^\top \mathbf{x} \leq \beta \quad \text{durch} \quad \mathbf{a}^\top \mathbf{x} + y = \beta, y \geq 0,$$

mit einer neuen *Schlupfvariablen*  $y \in \mathbb{R}$ .

(ii) Ersetze eine Variable

$$x \in \mathbb{R} \quad \text{durch} \quad x = x^+ - x^-, x^+ \geq 0, x^- \geq 0,$$

mit zwei neuen vorzeichenbeschränkten Variablen  $x^+, x^-$ .

Diese Transformationen erlauben es uns zwischen verschiedenen Beschreibungen von Polyedern zu wechseln, je nachdem welche Beschreibung im Kontext natürlicher ist.

**Korollar 3.7.** Jedes Polyeder  $\mathcal{P} \subseteq \mathbb{R}^n$  kann beschrieben werden durch ein Polyeder  $\mathcal{P}' \subseteq \mathbb{R}^{n'}$  mit  $n' \geq n$ ,  $A \in \mathbb{R}^{m \times n'}$ ,  $\mathbf{b} \in \mathbb{R}^m$  und

$$\mathcal{P}' = \mathcal{P}^=(A, \mathbf{b}) := \{\mathbf{x} \in \mathbb{R}^{n'} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}.$$

Natürlich können wir auch auf überflüssige Ungleichungen in der Beschreibung des Polyeders verzichten.

**Definition 3.8.** Sei  $A\mathbf{x} \leq \mathbf{b}$  mit  $A \in \mathbb{R}^{M \times N}$  und  $\mathbf{b} \in \mathbb{R}^M$ . Existiert  $I \subseteq M$  mit  $I \neq \emptyset$  und  $\mathcal{P}(A, \mathbf{b}) = \mathcal{P}(A_{(M \setminus I)}, \mathbf{b}_{M \setminus I})$ , dann heißt  $A\mathbf{x} \leq \mathbf{b}$  *redundant* und das Teilsystem  $A_I \mathbf{x} \leq \mathbf{b}_I$  sowie alle Ungleichungen  $A_i \mathbf{x} \leq b_i$  mit  $i \in I$  *redundant* bezüglich  $A\mathbf{x} \leq \mathbf{b}$ .

Eine wichtige Klasse von Polyedern sind solche, die aus Ursprungsstrahlen bestehen.

**Definition 3.9.** Ein Menge  $\mathcal{K} \subseteq \mathbb{R}^n$  ist ein *Kegel*, wenn  $\lambda \mathbf{x} \in \mathcal{K}$  für alle  $\mathbf{x} \in \mathcal{K}$ ,  $\lambda \geq 0$ . Ein Kegel ist *polyedrisch*, wenn er ein Polyeder ist.

Die folgende Charakterisierung von polyedrischen Kegeln ist sehr nützlich.

**Beobachtung 3.10.** Eine nichtleere Menge  $\mathcal{K} \subseteq \mathbb{R}^n$  ist genau dann ein polyedrischer Kegel, wenn eine Matrix  $A$  existiert mit

$$\mathcal{K} = \mathcal{P}(A, \mathbf{0}).$$

*Beweis.* Ist  $\mathcal{K} = \mathcal{P}(A, \mathbf{0})$ , dann ist  $\mathcal{K}$  offensichtlich ein Polyeder (Beobachtung 3.3). Für alle  $\mathbf{x} \in \mathcal{K}$  und  $\lambda \geq 0$  gilt außerdem

$$A(\lambda \mathbf{x}) = \lambda A\mathbf{x} \leq \mathbf{0},$$

also ist  $\lambda \mathbf{x} \in \mathcal{K}$  und  $\mathcal{K}$  somit ein Kegel.

Ist andersrum  $\mathcal{K} \neq \emptyset$  ein polyedrischer Kegel, dann existieren  $A \in \mathbb{R}^{m \times n}$  und  $\mathbf{b} \in \mathbb{R}^m$  mit  $\mathcal{K} = \mathcal{P}(A, \mathbf{b})$  (Beobachtung 3.3). Wegen  $\mathbf{0} \in \mathcal{K}$  gilt  $\mathbf{b} \geq \mathbf{0}$  und damit  $\mathcal{P}(A, \mathbf{0}) \subseteq \mathcal{P}(A, \mathbf{b}) = \mathcal{K}$ . Wir behaupten, dass  $\mathcal{K} = \mathcal{P}(A, \mathbf{0})$ .

Sei dazu  $\bar{\mathbf{x}} \in \mathbb{R}^n \setminus \mathcal{P}(A, \mathbf{0})$ . Dann existiert also  $i \in \{1, \dots, m\}$  mit  $A_i \cdot \bar{\mathbf{x}} > 0$ . Für beliebiges  $\lambda > b_i / (A_i \cdot \bar{\mathbf{x}})$  gilt damit  $A_i \cdot (\lambda \bar{\mathbf{x}}) = \lambda A_i \cdot \bar{\mathbf{x}} > b_i$ , also  $\lambda \bar{\mathbf{x}} \notin \mathcal{P}(A, \mathbf{b}) = \mathcal{K}$ . Da  $\mathcal{K}$  ein Kegel ist und  $\lambda > 0$  wegen  $\mathbf{b} \geq \mathbf{0}$ , folgt  $\bar{\mathbf{x}} \notin \mathcal{K}$ . Damit ist  $\mathcal{K} \subseteq \mathcal{P}(A, \mathbf{0})$  wie behauptet.  $\square$

### 3.1 Seitenflächen, Ecken und Facetten

Wir hatten bereits gesehen (Satz 2.28), dass abgeschlossene konvexe Mengen genau durch die Ungleichungen gegeben sind, die für alle Punkte der Menge erfüllt sind.

**Definition 3.11.** Die Ungleichung  $\mathbf{a}^\top \mathbf{x} \leq \beta$  heißt *gültig* für  $\mathcal{M} \subseteq \mathbb{R}^n$ , wenn  $\mathbf{a} \in \mathbb{R}^n$ ,  $\beta \in \mathbb{R}$  und

$$\mathcal{M} \subseteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\}.$$

Dabei erlauben wir  $\mathbf{a} = \mathbf{0}$ , also muss die rechte Seite kein Halbraum sein. Die Punkte eines Polyeders, für die eine gültige Ungleichung mit Gleichheit erfüllt sind, bilden eine Seitenfläche.

**Definition 3.12.** Ist  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine gültige Ungleichung für ein Polyeder  $\mathcal{P} \subseteq \mathbb{R}^n$ , dann heißt

$$\mathcal{F} = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}$$

Seitenfläche von  $\mathcal{P}$ . Die Seitenfläche  $\mathcal{F}$  wird von  $\mathbf{a}^\top \mathbf{x} \leq \beta$  induziert, sie ist *echt*, wenn  $\mathcal{F} \neq \mathcal{P}$  und *trivial*, wenn  $\mathcal{F} \in \{\emptyset, \mathcal{P}\}$ .

**Beispiel 3.13.** Das Polytop  $\mathcal{P}$  aus Beispiel 3.4 hat  $\mathcal{F} = \{\mathbf{x} \in \mathcal{P} : 6x_1 - 2x_2 = 13\}$  als Seitenfläche (siehe Abb. 3.1). Die Ungleichung  $2x_2 - x_1 \leq 8$  ist gültig für  $\mathcal{P}$  und induziert die Seitenfläche  $\mathcal{G} = \left\{ \begin{pmatrix} -2 \\ 3 \end{pmatrix} \right\}$ . Dieselbe Seitenfläche wird auch von  $-x_1 \leq 2$  induziert. Die gültigen Ungleichungen  $2x_2 - x_1 \leq 9$  und  $\mathbf{0}^\top \mathbf{x} \leq 0$  induzieren triviale Seitenflächen  $\triangle$

Wir machen einfache Beobachtungen zur Definition von Seitenflächen und setzen sie in Bezug zu Stützhyperebenen.

**Beobachtung 3.14.** Sei  $\mathcal{P}$  ein Polyeder. Dann gilt:

- (i)  $\mathcal{P}$  ist eine Seitenfläche von sich selbst.
- (ii)  $\emptyset$  ist eine Seitenfläche von  $\mathcal{P}$ .
- (iii) Induziert  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine nichttriviale Seitenfläche von  $\mathcal{P}$ , so ist  $\mathbf{a} \neq \mathbf{0}$ .
- (iv) Es ist  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = \beta\}$  eine Stützhyperebene von  $\mathcal{P}$  mit äußerer Normale  $\mathbf{a}$  genau dann, wenn  $\mathbf{a} \neq \mathbf{0}$  und  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine nichtleere Seitenfläche von  $\mathcal{P}$  induziert.

*Beweis.*

(i) Es ist  $\mathbf{0}^\top \mathbf{x} \leq 0$  eine gültige Ungleichung für  $\mathcal{P}$  und somit ist

$$\mathcal{P} = \mathcal{P} \cap \mathbb{R}^n = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{0}^\top \mathbf{x} = 0\}$$

eine Seitenfläche.

(ii) Es ist  $\mathbf{0}^\top \mathbf{x} \leq \beta$  für  $\beta > 0$  eine gültige Ungleichung für  $\mathcal{P}$  und somit ist

$$\emptyset = \mathcal{P} \cap \emptyset = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{0}^\top \mathbf{x} = \beta\}$$

eine Seitenfläche.

(iii) Damit  $\mathbf{a}^\top \mathbf{x} \leq \beta$  für  $\mathbf{a} = \mathbf{0}$  gültig ist, muss  $\beta \geq 0$  gelten. Damit wäre nach (i) und (ii)  $\mathcal{P} \cap \mathcal{H} \in \{\mathcal{P}, \emptyset\}$  eine triviale Seitenfläche.

(iv) Sei  $\mathcal{H}$  eine Stützhyperebene von  $\mathcal{P}$ . Per Definition einer Hyperebene ist dann  $\mathbf{a} \neq \mathbf{0}$ . Da  $\mathcal{H}$  Stützhyperebene mit äußerer Normale  $\mathbf{a}$  ist, gilt  $\mathcal{P} \subseteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\}$ , also ist  $\mathbf{a}^\top \mathbf{x} \leq \beta$  gültig. Da  $\mathcal{H}$  Stützhyperebene ist, gilt außerdem  $\mathcal{P} \cap \mathcal{H} \neq \emptyset$ . Damit induziert  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine nichtleere Seitenfläche von  $\mathcal{P}$ .

Sei andersherum  $\mathbf{a} \neq \mathbf{0}$  und induziere  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine nichtleere Seitenfläche von  $\mathcal{P}$ . Wegen  $\mathbf{a} \neq \mathbf{0}$  ist  $\mathcal{H}$  eine Hyperebene. Außerdem ist  $\mathcal{P} \cap \mathcal{H} \neq \emptyset$ , da diese von  $\mathbf{a}^\top \mathbf{x} \leq \beta$  induzierte Seitenfläche nichtleer ist. Schließlich ist  $\mathcal{P} \subseteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} \leq \beta\}$ , da die Ungleichung  $\mathbf{a}^\top \mathbf{x} \leq \beta$  gültig für  $\mathcal{P}$  ist.

□

Eine Seitenfläche ist also trivial oder der Schnitt des Polyeders mit einer Hyperebene, die wiederum der Schnitt zweier abgeschlossener Halbräume ist.

**Beobachtung 3.15.** Jede Seitenfläche eines Polyeders ist wieder ein Polyeder.

Von zentraler Bedeutung für die lineare Optimierung ist, dass lineare Zielfunktionen ihre Extrema genau auf Seitenflächen von Polyedern annehmen.

**Proposition 3.16.** Für alle Polyeder  $\mathcal{P} \subseteq \mathbb{R}^n$  und alle  $\mathbf{c} \in \mathbb{R}^n$  sind  $\operatorname{argmin}\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$  und  $\operatorname{argmax}\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$  jeweils Seitenflächen von  $\mathcal{P}$ .

*Beweis.* Wir zeigen die Aussage für  $\operatorname{argmax}$ , das Argument für  $\operatorname{argmin}$  ist identisch. Falls  $\operatorname{argmax}\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\} \in \{\emptyset, \mathcal{P}\}$ , also insbesondere für  $\mathbf{c} = \mathbf{0}$  oder wenn das Maximum nicht angenommen wird, gilt die Behauptung nach Beobachtung 3.14. Ansonsten ist die Ungleichung  $\mathbf{c}^\top \mathbf{x} \leq \gamma^*$  mit  $\gamma^* = \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$  gültig für  $\mathcal{P}$  und induziert die nichtleere Seitenfläche  $\mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} = \gamma^*\} = \operatorname{argmax}\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ . □

Wir werden sehen, dass die Seitenflächen eines Polyeders der Form  $\mathcal{P}(\mathbf{A}, \mathbf{b})$  genau über Teilmengen von Ungleichungen im System  $\mathbf{A}\mathbf{x} \leq \mathbf{b}$  charakterisiert sind, die mit Gleichheit gelten. Die folgende Definition bezieht sich auf eine feste Beschreibung eines Polyeders.

**Definition 3.17.** Sei  $\mathcal{P} = \mathcal{P}(\mathbf{A}, \mathbf{b})$  ein Polyeder mit  $\mathbf{A} \in \mathbb{R}^{M \times N}$  und  $\mathbf{b} \in \mathbb{R}^M$ . Die Menge der *aktiven Ungleichungen* für alle  $\bar{\mathbf{x}} \in \mathcal{F} \subseteq \mathcal{P}$  ist gegeben durch

$$\operatorname{eq}(\mathcal{F}) := \{i \in M : \mathbf{A}_i \bar{\mathbf{x}} = b_i \forall \bar{\mathbf{x}} \in \mathcal{F}\}.$$

Die von  $I \subseteq M$  induzierte *Seitenfläche* von  $\mathcal{P}$  ist gegeben durch

$$\operatorname{fa}(I) := \{\mathbf{x} \in \mathcal{P} : \mathbf{A}_I \mathbf{x} = \mathbf{b}_I\}.$$

**Beobachtung 3.18.** Sei  $\mathcal{P} = \mathcal{P}(A, \mathbf{b})$  ein Polyeder mit  $A \in \mathbb{R}^{M \times N}$  und  $\mathbf{b} \in \mathbb{R}^M$ . Für alle  $I \subseteq M$  ist  $\text{fa}(I)$  eine Seitenfläche von  $\mathcal{P}$  gemäß Definition 3.12.

*Beweis.* Mit  $\mathbf{a}^\top = \sum_{i \in I} \mathbf{A}_i$  und  $\beta = \sum_{i \in I} b_i$  ist  $\mathbf{a}^\top \mathbf{x} \leq \beta$  eine gültige Ungleichung für  $\mathcal{P}$  und damit  $\mathcal{F} = \{\mathbf{x} \in \mathcal{P} : \mathbf{a}^\top \mathbf{x} = \beta\}$  eine Seitenfläche von  $\mathcal{P}$ . Offensichtlich gilt  $\text{fa}(I) \subseteq \mathcal{F}$ . Sei andersherum  $\bar{\mathbf{x}} \in \mathcal{F}$ . Wegen  $\bar{\mathbf{x}} \in \mathcal{P}$  gilt  $\mathbf{a}^\top \bar{\mathbf{x}} = \sum_{i \in I} \mathbf{A}_i \bar{\mathbf{x}} \leq \sum_{i \in I} b_i = \beta$  und wegen  $\bar{\mathbf{x}} \in \mathcal{F}$  gilt  $\mathbf{a}^\top \bar{\mathbf{x}} = \beta$ . Es folgt  $\mathbf{A}_i \bar{\mathbf{x}} = b_i$  für alle  $i \in I$  und damit  $\bar{\mathbf{x}} \in \text{fa}(I)$ . Also ist  $\text{fa}(I) = \mathcal{F}$ .  $\square$

**Beispiel 3.19.** In Beispiel 3.4 bzw. 3.13 (siehe Abb. 3.1) gilt  $\text{eq}(\mathcal{F}) = \{5\}$  und  $\text{eq}(\mathcal{G}) = \{3, 4\}$ , sowie  $\mathcal{F} = \text{fa}(\{5\})$  und  $\mathcal{G} = \text{fa}(\{3, 4\})$ .  $\triangle$

Zur Beschreibung des Rands eines Polyeders können wir uns auf seine Facetten beschränken.

**Definition 3.20.** Eine echte Seitenfläche  $\mathcal{F}$  von  $\mathcal{P} = \mathcal{P}(A, \mathbf{b})$  heißt *Facette* von  $\mathcal{P}$ , falls  $\mathcal{F}$  in keiner anderen echten Seitenfläche von  $\mathcal{P}$  enthalten ist.

Während Facetten die äußere Beschreibung eines Polyeders liefern, brauchen wir für innere Beschreibung seine Ecken.

**Definition 3.21.** Ein Punkt  $\mathbf{x} \in \mathcal{P}$  ist eine *Ecke* des Polyeders  $\mathcal{P}$ , wenn  $\{\mathbf{x}\}$  eine Seitenfläche von  $\mathcal{P}$  ist.

**Beispiel 3.22.** In Beispiel 3.4 bzw. 3.13 (siehe Abb. 3.1) ist  $\mathcal{F}$  eine Facette, und der einzige Punkt in  $\mathcal{G}$  ist eine Ecke von  $\mathcal{P}$ . Aber  $\mathcal{G}$  ist in der echten Seitenfläche  $\mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{A}_3 \mathbf{x} = b_3\}$  enthalten, also ist  $\mathcal{G}$  keine Facette.  $\triangle$

Wir charakterisieren nun die Ecken eines Polyeders und zeigen insbesondere, dass sie genau seine Extrempunkte sind.

**Satz 3.23.** Sei  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \subseteq \mathbb{R}^n$  ein Polyeder und  $\hat{\mathbf{x}} \in \mathcal{P}$ . Die folgenden Aussagen sind äquivalent:

- (i) Der Punkt  $\hat{\mathbf{x}}$  ist eine Ecke von  $\mathcal{P}$ .
- (ii) Es gilt  $\text{rank}(A_{\text{eq}(\{\hat{\mathbf{x}}\})}) = n$ .
- (iii) Der Punkt  $\hat{\mathbf{x}}$  ist ein Extrempunkt von  $\mathcal{P}$ .
- (iv) Es existiert  $\mathbf{c} \in \mathbb{R}^n$  mit  $\{\hat{\mathbf{x}}\} = \arg \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ .

*Beweis.*

Aus (i) folgt (iv): Sei  $\hat{\mathbf{x}}$  Ecke von  $\mathcal{P}$ . Per Definition ist  $\{\hat{\mathbf{x}}\}$  eine Seitenfläche von  $\mathcal{P}$ , also existiert eine gültige Ungleichung  $\mathbf{c}^\top \mathbf{x} \leq \gamma$  für  $\mathcal{P}$  mit  $\{\hat{\mathbf{x}}\} = \{\mathbf{x} \in \mathcal{P} : \mathbf{c}^\top \mathbf{x} = \gamma\}$ . Damit ist  $\{\hat{\mathbf{x}}\} = \arg \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ .

Aus (iv) folgt (iii): Sei  $\mathbf{c} \in \mathbb{R}^n$  so dass  $\{\hat{\mathbf{x}}\} = \arg \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ . Dann gilt  $\mathbf{c}^\top \hat{\mathbf{x}} > \mathbf{c}^\top \mathbf{x}$  für alle  $\mathbf{x} \in \mathcal{P} \setminus \{\hat{\mathbf{x}}\}$ . Damit ist  $\hat{\mathbf{x}} \neq \lambda \mathbf{x} + (1 - \lambda) \mathbf{y}$  für alle  $\mathbf{x} \in \mathcal{P}$ ,  $\mathbf{y} \in \mathcal{P} \setminus \{\hat{\mathbf{x}}\}$  und  $\lambda \in (0, 1)$ , denn

$$\mathbf{c}^\top \hat{\mathbf{x}} = \lambda \mathbf{c}^\top \hat{\mathbf{x}} + (1 - \lambda) \mathbf{c}^\top \hat{\mathbf{x}} > \lambda \mathbf{c}^\top \mathbf{x} + (1 - \lambda) \mathbf{c}^\top \mathbf{y} = \mathbf{c}^\top (\lambda \mathbf{x} + (1 - \lambda) \mathbf{y})$$

wegen  $\mathbf{x} \neq \hat{\mathbf{x}}$  oder  $\mathbf{y} \neq \hat{\mathbf{x}}$ .

Aus (iii) folgt (ii): Ist  $\text{rank}(A_{\text{eq}(\{\hat{\mathbf{x}}\})}) < n$ , dann existiert  $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  mit  $A_{\text{eq}(\{\hat{\mathbf{x}}\})} \cdot \mathbf{x} = \mathbf{0}$ . Damit existiert, wegen Stetigkeit der Abbildung  $A\mathbf{x}$ , ein  $\varepsilon > 0$  für das inaktive Ungleichungen für  $\{\hat{\mathbf{x}}\}$  auch für  $\{\hat{\mathbf{x}} \pm \varepsilon \mathbf{x}\}$  inaktiv sind, also

$$A(\hat{\mathbf{x}} \pm \varepsilon \mathbf{x}) \leq \mathbf{b}.$$

Aber dann ist  $\mathbf{y} := (\hat{\mathbf{x}} + \varepsilon \mathbf{x}) \in \mathcal{P}$  und  $\mathbf{z} := (\hat{\mathbf{x}} - \varepsilon \mathbf{x}) \in \mathcal{P}$  und somit  $\hat{\mathbf{x}} = \frac{1}{2}\mathbf{y} + \frac{1}{2}\mathbf{z}$  kein Extrempunkt von  $\mathcal{P}$ .

Aus (ii) folgt (i): Ist  $\text{rank}(A_{\text{eq}(\{\hat{\mathbf{x}}\})}) = n$ , hat also  $A_{\text{eq}(\{\hat{\mathbf{x}}\})}$  vollen Spaltenrang, so hat  $A_{\text{eq}(\{\hat{\mathbf{x}}\})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\{\hat{\mathbf{x}}\})}$  höchstens eine Lösung (siehe *Lineare Algebra I*). Per Definition von  $\text{eq}(\cdot)$  ist  $\hat{\mathbf{x}}$  also die einzige Lösung. Damit ist

$$\text{fa}(\text{eq}(\{\hat{\mathbf{x}}\})) = \{\mathbf{x} \in \mathcal{P} : A_{\text{eq}(\{\hat{\mathbf{x}}\})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\{\hat{\mathbf{x}}\})}\} = \{\hat{\mathbf{x}}\}$$

und nach Beobachtung 3.18 ist  $\text{fa}(\text{eq}(\{\hat{\mathbf{x}}\})) = \{\hat{\mathbf{x}}\}$  eine Seitenfläche von  $\mathcal{P}$ . □

Wir geben eine weitere Charakterisierung von Ecken für Polyeder der Form  $\mathcal{P}^=(A, \mathbf{b})$ . Dazu erinnern wir zunächst an folgende Notation.

**Definition 3.24.** Der Träger (engl: support) eines Vektors  $\mathbf{x} \in \mathbb{R}^N$  ist

$$\text{supp}(\mathbf{x}) := \{i \in N : x_i \neq 0\}.$$

**Satz 3.25** (Übung). Sei  $\mathcal{P} = \mathcal{P}^=(A, \mathbf{b})$  ein Polyeder und  $\hat{\mathbf{x}} \in \mathcal{P}$ . Die folgenden Aussagen sind äquivalent:

- (i) Der Punkt  $\hat{\mathbf{x}}$  ist eine Ecke von  $\mathcal{P}$ .
- (ii) Es gilt  $\text{rank}(A_{\text{supp}(\hat{\mathbf{x}})}) = |\text{supp}(\hat{\mathbf{x}})|$ .
- (iii) Die Spalten von  $A_{\text{supp}(\hat{\mathbf{x}})}$  sind linear unabhängig.

Wir erhalten insbesondere eine Beziehung zwischen Ecken des Polyeders  $\mathcal{P}^=(A, \mathbf{b})$  und Spalten des Systems  $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$ .

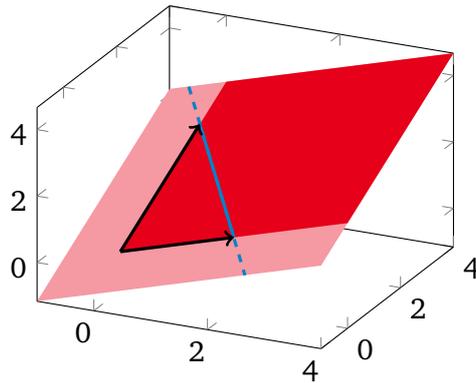
**Korollar 3.26.** Habe  $A \in \mathbb{R}^{M \times N}$  vollen Zeilenrang und sei  $\hat{\mathbf{x}}$  eine Ecke des Polyeders  $\mathcal{P}^=(A, \mathbf{b})$ . Dann existiert eine Partition  $B \uplus N_0 = N$  mit

$$\text{supp}(\hat{\mathbf{x}}) \subseteq B, \quad |B| = m := |M|, \quad A_{\cdot B} \text{ invertierbar.}$$

Für jede solche Partition gilt

$$\hat{\mathbf{x}}_{N_0} = \mathbf{0}, \quad \hat{\mathbf{x}}_B = A_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}.$$

*Beweis.* Nach Satz 3.25 besteht  $A_{\text{supp}(\hat{\mathbf{x}})}$  aus linear unabhängigen Spalten. Wegen  $\text{rank}(A) = m$  hat der Spaltenraum von  $A$  Dimension  $m$ . Es folgt  $|\text{supp}(\hat{\mathbf{x}})| \leq m$  und dass wir durch Basisergänzung  $m - |\text{supp}(\hat{\mathbf{x}})|$  zusätzliche Spalten von  $A$  finden können, die mit den Spalten aus  $A_{\text{supp}(\hat{\mathbf{x}})}$



**Abbildung 3.2:** Zwei Punkte  $(2, 0, 1)^T$  und  $(0, 3, 2)^T$  in  $\mathbb{R}^3$  mit ihrer linearen (hellrot), affinen (blau), konvexen (blau durchgezogen) und konischen (rot) Hülle.

linear unabhängig sind (*Lineare Algebra I*). Also können wir  $B \supseteq \text{supp}(\hat{x})$  finden mit  $|B| = m$  und  $\text{rank}(A_B) = m$ . Damit ist  $A_B \in \mathbb{R}^{m \times m}$  invertierbar.

Offensichtlich ist  $\hat{x}_{N_0} = \mathbf{0}$ , da  $\text{supp}(\hat{x}) \subseteq B = N \setminus N_0$ . Mit  $\hat{x} \in \mathcal{P}^=(A, \mathbf{b})$  folgt  $\hat{x} \geq \mathbf{0}$  und

$$\mathbf{b} = A\hat{x} = A_B \hat{x}_B + A_{N_0} \hat{x}_{N_0} = A_B \hat{x}_B.$$

Da  $A_B$  invertierbar ist, gilt also  $\mathbf{0} \leq \hat{x}_B = A_B^{-1} \mathbf{b}$ . □

### 3.2 Dimension und Darstellung von Seitenflächen

Wir beschäftigen uns nun mit der Dimension von Seitenflächen eines Polyeders im Zusammenhang mit ihren aktiven Ungleichungen. Wir führen zunächst die Dimension einer beliebigen Punktmenge auf die Dimension eines affinen Raums zurück. Dazu definieren wir den Abschluss einer Menge in Bezug auf die Bildung von Linearkombinationen, affinen Kombinationen und konischen Kombinationen von Punkten (vgl. Satz 2.7 für die konvexe Hülle).

**Definition 3.27.** Die *lineare Hülle*  $\text{lin}(\mathcal{M})$  von  $\mathcal{M} \subseteq \mathbb{R}^n$ ,  $\mathcal{M} \neq \emptyset$  ist der kleinste lineare Raum, der  $\mathcal{M}$  enthält, also

$$\text{lin}(\mathcal{M}) := \{A\boldsymbol{\lambda} : A \in \mathbb{R}^{n \times k}, \{\mathbf{A}_1, \dots, \mathbf{A}_k\} \subseteq \mathcal{M}, k \in \mathbb{N}, \boldsymbol{\lambda} \in \mathbb{R}^k\}.$$

Die *affine Hülle*  $\text{aff}(\mathcal{M})$  von  $\mathcal{M} \subseteq \mathbb{R}^n$  ist der kleinste affine Raum, der  $\mathcal{M}$  enthält, also

$$\text{aff}(\mathcal{M}) := \{A\boldsymbol{\lambda} : A \in \mathbb{R}^{n \times k}, \{\mathbf{A}_1, \dots, \mathbf{A}_k\} \subseteq \mathcal{M}, k \in \mathbb{N}, \boldsymbol{\lambda} \in \mathbb{R}^k, \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

Die *konische Hülle*  $\text{cone}(\mathcal{M})$  von  $\mathcal{M} \subseteq \mathbb{R}^n$  ist der kleinste konvexe Kegel, der  $\mathcal{M}$  enthält, also

$$\text{cone}(\mathcal{M}) := \{A\boldsymbol{\lambda} : A \in \mathbb{R}^{n \times k}, \{\mathbf{A}_1, \dots, \mathbf{A}_k\} \subseteq \mathcal{M}, k \in \mathbb{N}, \boldsymbol{\lambda} \in \mathbb{R}^k, \boldsymbol{\lambda} \geq \mathbf{0}\}.$$

**Beispiel 3.28.** Seien  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  mit  $\mathbf{x} \neq \mathbf{y}$  (siehe Abb. 3.2). Sind  $\mathbf{x}$  und  $\mathbf{y}$  linear unabhängig, dann ist  $\text{lin}(\{\mathbf{x}, \mathbf{y}\})$  durch die Ebene gegeben, die  $\mathbf{x}, \mathbf{y}, \mathbf{0}$  enthält. Die affine Hülle von  $\mathbf{x}$  und  $\mathbf{y}$  ist durch die Gerade gegeben, die durch  $\mathbf{x}$  und  $\mathbf{y}$  verläuft, denn

$$\text{aff}(\{\mathbf{x}, \mathbf{y}\}) = \{\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} : \lambda \in \mathbb{R}\} = \{\mathbf{y} + \lambda(\mathbf{x} - \mathbf{y}) : \lambda \in \mathbb{R}\}.$$

Die konische Hülle wird von den zwei Ursprungsstrahlen, die  $\mathbf{x}$  bzw.  $\mathbf{y}$  enthalten, aufgespannt.  $\triangle$

Wir können damit die Dimension einer Menge von Punkten als die kleinste Dimension eines affinen Raums definieren, in den die Menge eingebettet werden kann.

**Definition 3.29.** Die Dimension  $\dim(\mathcal{M})$  einer nichtleeren Menge  $\mathcal{M} \subseteq \mathbb{R}^n$  ist definiert durch

$$\dim(\mathcal{M}) := \dim(\text{aff}(\mathcal{M})).$$

Die Menge  $\mathcal{M}$  heißt *volldimensional*, wenn  $\dim(\mathcal{M}) = n$ , bzw.  $\text{aff}(\mathcal{M}) = \mathbb{R}^n$ .

Wir können die Dimension natürlich auch auf die Dimension von linearen Räumen zurückführen.

**Beobachtung 3.30.** Für  $\mathcal{M} \subseteq \mathbb{R}^n$  gilt  $\text{lin}(\mathcal{M}) = \text{aff}(\mathcal{M} \cup \{\mathbf{0}\})$  und  $\text{aff}(\mathcal{M}) = \{\mathbf{x}_0\} + \text{lin}(\{\mathbf{x} - \mathbf{x}_0 : \mathbf{x} \in \mathcal{M} \setminus \{\mathbf{x}_0\}\})$  für beliebiges  $\mathbf{x}_0 \in \mathcal{M}$ , also insbesondere

$$\dim(\mathcal{M}) = \max\{d \in \mathbb{N} \cup \{0\} : \exists \mathbf{x}_0, \dots, \mathbf{x}_d \in \mathcal{M} \text{ mit } \mathbf{x}_1 - \mathbf{x}_0, \mathbf{x}_2 - \mathbf{x}_0, \dots, \mathbf{x}_d - \mathbf{x}_0 \text{ linear unabh.}\}.$$

Wir definieren relativ innere Punkte eines Polyeders als innere Punkte bezüglich der affinen Hülle. Ist das Polyeder nichtleer und nicht volldimensional, liegen alle seine Punkte auf dem Rand des Polyeders, es gibt also keine inneren Punkte. Wir werden dennoch sehen, dass mindestens ein *relativ innerer* Punkt existiert. Anders gesagt, ist ein Punkt ein relativ innerer Punkt, wenn eine Kugel der Dimension des Polyeders um den Punkt existiert, die im Polyeder liegt. Insbesondere ist  $\mathbf{x}$  relativ innerer Punkt des nulldimensionalen Polyeders  $\{\mathbf{x}\}$ .

**Definition 3.31.** Ein Punkt  $\mathbf{x}$  eines Polyeders  $\mathcal{P}$  heißt *relativ innerer Punkt* von  $\mathcal{P}$ , falls  $\mathbf{x}$  in keiner echten Seitenfläche von  $\mathcal{P}$  enthalten ist.

Wir zeigen, dass jedes Polyeder einen relativ inneren Punkt besitzt.

**Proposition 3.32.** Sei  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \neq \emptyset$  mit  $A \in \mathbb{R}^{M \times N}$  und  $\mathbf{b} \in \mathbb{R}^M$ . Dann existiert  $\bar{\mathbf{x}} \in \mathcal{P}$  mit

$$\begin{aligned} A_i \bar{\mathbf{x}} &= b_i \quad \forall i \in \text{eq}(\mathcal{P}), \\ A_j \bar{\mathbf{x}} &< b_j \quad \forall j \in M \setminus \text{eq}(\mathcal{P}). \end{aligned}$$

Jeder solche Punkt ist relativ innerer Punkt von  $\mathcal{P}$ .

*Beweis.* Wir zeigen zunächst die Existenz eines Punkts  $\bar{\mathbf{x}}$  mit der gewünschten Eigenschaft. Ist  $\text{eq}(\mathcal{P}) = M$ , können wir jedes  $\bar{\mathbf{x}} \in \mathcal{P} \neq \emptyset$  wählen.

Sei also  $I := \text{eq}(\mathcal{P}) \neq M$  und  $J := M \setminus \text{eq}(\mathcal{P}) \neq \emptyset$ . Für alle  $j \in J$  existiert ein  $\mathbf{x}_j \in \mathcal{P}$  mit  $A_j \mathbf{x}_j < \mathbf{b}_j$  per Definition von  $\text{eq}(\mathcal{P})$ . Da  $\mathcal{P}$  konvex ist, gilt

$$\bar{\mathbf{x}} := \frac{1}{|J|} \sum_{j \in J} \mathbf{x}_j \in \mathcal{P}$$

und damit  $A_I \bar{\mathbf{x}} = \mathbf{b}_I$ . Also hat  $\bar{\mathbf{x}}$  die gewünschte Eigenschaft, denn für alle  $i \in J$  gilt

$$A_i \bar{\mathbf{x}} = \frac{1}{|J|} A_i \mathbf{x}_i + \frac{1}{|J|} \sum_{j \in J \setminus \{i\}} A_i \mathbf{x}_j < \frac{1}{|J|} \sum_{j \in J} \mathbf{b}_i = \mathbf{b}_i.$$

Es bleibt zu zeigen, dass  $\bar{\mathbf{x}}$  relativ innerer Punkt von  $\mathcal{P}$  ist. Sei dazu  $\mathbf{d}^\top \mathbf{x} \leq \delta$  eine gültige Ungleichung für  $\mathcal{P}$  und  $\mathcal{F}$  die zugehörige induzierte Seitenfläche, so dass  $\bar{\mathbf{x}} \in \mathcal{F}$ . Wir zeigen, dass  $\mathcal{F} = \mathcal{P}$ .

Für alle  $\mathbf{y} \in \ker(A_I) = \{\mathbf{x} \in \mathbb{R}^n : A_I \mathbf{x} = \mathbf{0}\}$  gilt

$$A_I(\bar{\mathbf{x}} + \mathbf{y}) = \mathbf{b}_I.$$

Da  $A_J \mathbf{x}$  eine stetige Abbildung ist, existiert  $\varepsilon > 0$ , so dass für alle  $\mathbf{y} \in \ker(A_I)$  mit  $\|\mathbf{y}\| \leq \varepsilon$  gilt

$$A_J(\bar{\mathbf{x}} + \mathbf{y}) < \mathbf{b}_J.$$

Es folgt

$$\{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_I), \|\mathbf{y}\| \leq \varepsilon\} \subseteq \mathcal{P} \subseteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{d}^\top \mathbf{x} \leq \delta\}.$$

Damit gilt für alle  $\mathbf{y} \in \ker(A_I)$  mit  $\|\mathbf{y}\| \leq \varepsilon$ , dass

$$\delta \geq \mathbf{d}^\top (\bar{\mathbf{x}} + \mathbf{y}) = \delta + \mathbf{d}^\top \mathbf{y}.$$

Da  $\ker(A_I)$  ein linearer Teilraum ist, folgt  $\mathbf{d}^\top \mathbf{y} = 0$  für alle  $\mathbf{y} \in \ker(A_I)$ . Also ist

$$\mathcal{P} \cap \{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_I)\} \subseteq \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{d}^\top \mathbf{x} = \delta\} = \mathcal{F}. \quad (3.1)$$

Andererseits ist  $A_I \mathbf{x} = \mathbf{b}_I$  für  $\mathbf{x} \in \mathbb{R}^n$  genau dann, wenn  $A_I(\mathbf{x} - \bar{\mathbf{x}}) = \mathbf{0}$ , bzw. wenn  $\mathbf{y} = \mathbf{x} - \bar{\mathbf{x}} \in \ker(A_I)$ . Daraus folgt

$$\mathcal{P} = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : A_I \mathbf{x} = \mathbf{b}_I\} = \mathcal{P} \cap \{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_I)\}. \quad (3.2)$$

Da  $\mathcal{F}$  Seitenfläche von  $\mathcal{P}$  ist, also  $\mathcal{F} \subseteq \mathcal{P}$ , ergibt sich aus (3.1) und (3.2), dass  $\mathcal{F} = \mathcal{P}$ .  $\square$

Aus unseren Überlegungen lässt sich eine Dimensionsformel ableiten.

**Korollar 3.33.** Für jedes Polyeder  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \neq \emptyset$  gilt

$$\dim(\mathcal{P}) = n - \text{rank}(A_{\text{eq}(\mathcal{P})}) = \dim(\ker(A_{\text{eq}(\mathcal{P})})).$$

*Beweis.* Wir haben im Beweis von Proposition 3.32 gesehen, dass

$$\{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_{\text{eq}(\mathcal{P})}), \|\mathbf{y}\| \leq \varepsilon\} \subseteq \mathcal{P} \subseteq \{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_{\text{eq}(\mathcal{P})})\}.$$

Aus  $\text{aff}(\{\bar{\mathbf{x}} + \mathbf{y} : \mathbf{y} \in \ker(A_{\text{eq}(\mathcal{P})}), \|\mathbf{y}\| \leq \varepsilon\}) = \text{aff}(\{\bar{\mathbf{x}}\} + \ker(A_{\text{eq}(\mathcal{P})}))$  folgt damit

$$\dim(\ker(A_{\text{eq}(\mathcal{P})})) \leq \dim(\mathcal{P}) \leq \dim(\ker(A_{\text{eq}(\mathcal{P})})),$$

beziehungsweise  $\dim(\mathcal{P}) = \dim(\ker(A_{\text{eq}(\mathcal{P})}))$ . Aus dem Dimensionssatz über lineare Abbildungen (*Lineare Algebra I*) wissen wir außerdem

$$n = \text{rank}(A_{\text{eq}(\mathcal{P})}) + \dim(\ker(A_{\text{eq}(\mathcal{P})})).$$

Beide Tatsachen implizieren zusammen die Aussage. □

Eine unmittelbare Folgerung ist, dass aktive Ungleichungen nötig sind, um die Dimension einzuschränken. Die Existenz aktiver Ungleichungen ist allerdings nicht hinreichend, da die Ungleichung  $\mathbf{0}^\top \mathbf{x} \leq 0$  zugelassen ist.

**Korollar 3.34.** Ist  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \neq \emptyset$  ein Polyeder mit  $\text{eq}(\mathcal{P}) = \emptyset$ , dann ist  $\mathcal{P}$  volldimensional.

Wir verallgemeinern die Dimensionsformel weiter für Seitenflächen. Dabei ist zu beachten, dass sich die folgende Dimensionsformel über  $\text{eq}(\cdot)$  auf die Beschreibung des zugrunde liegenden Polytops bezieht!

**Satz 3.35** (Dimensionsformel). Ist  $\mathcal{F} \neq \emptyset$  eine Seitenfläche des Polyeders  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \subseteq \mathbb{R}^n$ , dann gilt

$$\dim(\mathcal{F}) = n - \text{rank}(A_{\text{eq}(\mathcal{F})}). \quad (3.3)$$

Ist  $\mathcal{F}$  eine echte Seitenfläche von  $\mathcal{P}$ , dann gilt  $\dim(\mathcal{F}) < \dim(\mathcal{P})$ .

*Beweis.* Sei  $A = \mathbb{R}^{M \times N}$  und definiere  $I = \text{eq}(\mathcal{F})$  und  $J = M \setminus \text{eq}(\mathcal{F})$ . Sei  $\mathbf{d}^\top \mathbf{x} \leq \delta$  eine gültige Ungleichung für  $\mathcal{P}$ , die  $\mathcal{F}$  induziert. Dann ist

$$\begin{aligned} \mathcal{F} &= \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{d}^\top \mathbf{x} = \delta\} \\ &= \{\mathbf{x} \in \mathbb{R}^n : -\mathbf{d}^\top \mathbf{x} \leq -\delta, A\mathbf{x} \leq \mathbf{b}\}. \end{aligned}$$

Es ist also  $\mathcal{F}$  ein Polyeder (Beobachtung 3.15) mit den aktiven Nebenbedingungen

$$-\mathbf{d}^\top \mathbf{x} = -\delta \quad \text{und} \quad A_I \mathbf{x} = \mathbf{b}_I.$$

Nach Proposition 3.32 existiert also ein  $\bar{\mathbf{x}} \in \mathcal{F}$  mit

$$\mathbf{d}^\top \bar{\mathbf{x}} = \delta, \quad A_I \bar{\mathbf{x}} = \mathbf{b}_I, \quad A_J \bar{\mathbf{x}} < \mathbf{b}_J$$

und nach Korollar 3.33 für das Polyeder  $\mathcal{F}$  gilt

$$\dim(\mathcal{F}) = \dim(\ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right)).$$

Wegen dem Dimensionssatz über lineare Abbildungen (*Lineare Algebra I*) ist

$$\dim(\ker(A_I)) = n - \text{rank}(A_I).$$

Um (3.3) zu zeigen, genügt also zu zeigen, dass  $\ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right) = \ker(A_I)$ . Es ist klar, dass  $\ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right) \subseteq \ker(A_I)$ . Sei nun  $\mathbf{y} \in \ker(A_I)$ . Wegen Stetigkeit der Abbildung  $A_I \cdot \mathbf{x}$ , existiert  $\varepsilon > 0$  mit

$$A_I(\bar{\mathbf{x}} \pm \varepsilon \mathbf{y}) = \mathbf{b}_I, \quad A_I(\bar{\mathbf{x}} \pm \varepsilon \mathbf{y}) < \mathbf{b}_I.$$

Also sind  $\bar{\mathbf{x}} \pm \varepsilon \mathbf{y} \in \mathcal{P}$  und da  $\mathbf{d}^\top \mathbf{x} \leq \delta$  eine gültige Ungleichung ist, gilt

$$\delta \geq \mathbf{d}^\top(\bar{\mathbf{x}} \pm \varepsilon \mathbf{y}) = \delta \pm \varepsilon \mathbf{d}^\top \mathbf{y}.$$

Daraus folgt  $\mathbf{d}^\top \mathbf{y} = 0$ , also ist  $\mathbf{y} \in \ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right)$  und  $\ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right) = \ker(A_I)$  bzw. (3.3) ist gezeigt.

Wegen  $\ker\left(\begin{pmatrix} A_I \\ \mathbf{d}^\top \end{pmatrix}\right) = \ker(A_I)$  und  $\mathcal{F} \neq \emptyset$  gilt

$$\mathcal{F} = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}, \mathbf{d}^\top \mathbf{x} = \delta\} = \mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\} = \text{fa}(\text{eq}(\mathcal{F})) \quad (3.4)$$

Ist  $\mathcal{F}$  eine echte Seitenfläche von  $\mathcal{P}$ , dann gilt damit

$$\mathcal{P} \cap \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\} = \mathcal{F} \subsetneq \mathcal{P} = \{\mathbf{x} \in \mathcal{P} : A_{\text{eq}(\mathcal{P})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{P})}\}.$$

Insbesondere ist, wegen  $\mathcal{F} \neq \emptyset$ ,

$$\emptyset \neq \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\} \subsetneq \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{P})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{P})}\}$$

und damit  $\text{rank}(A_{\text{eq}(\mathcal{F})}) > \text{rank}(A_{\text{eq}(\mathcal{P})})$ . Aus (3.3) folgt damit  $\dim(\mathcal{F}) < \dim(\mathcal{P})$ .  $\square$

Wir können insbesondere ableiten, dass Seitenflächen eindeutig durch ihre aktiven Ungleichungen induziert werden.

**Korollar 3.36.** Ist  $\mathcal{F} \neq \emptyset$  eine Seitenfläche des Polyeders  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \subseteq \mathbb{R}^n$ , dann ist

$$\begin{aligned} \mathcal{F} &= \{\mathbf{x} \in \mathcal{P} : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\} = \text{fa}(\text{eq}(\mathcal{F})), \\ \text{aff}(\mathcal{F}) &= \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\}. \end{aligned}$$

*Beweis.* Den ersten Teil der Aussage haben wir im Beweis von Satz 3.35 (Gleichung (3.4)) gezeigt.

Aus  $\mathcal{F} \subseteq \mathcal{T} := \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}(\mathcal{F})}\}$  folgt

$$\text{aff}(\mathcal{F}) \subseteq \text{aff}(\mathcal{T}) = \mathcal{T},$$

da  $\mathcal{T}$  ein affiner Raum ist. Weiter gilt wegen  $\mathcal{F} \neq \emptyset$  nach Satz 3.35 und dem Dimensionssatz über lineare Abbildungen (*Lineare Algebra I*), dass

$$\dim(\mathcal{F}) = n - \text{rank}(A_{\text{eq}(\mathcal{F})}) = \dim(\ker(A_{\text{eq}(\mathcal{F})})) = \dim(\mathcal{T}).$$

Da  $\text{aff}(\mathcal{F})$  und  $\mathcal{T}$  affine Räume sind mit  $\text{aff}(\mathcal{F}) \subseteq \mathcal{T}$  sowie  $\dim(\text{aff}(\mathcal{F})) = \dim(\mathcal{T})$ , folgt, dass  $\text{aff}(\mathcal{F}) = \mathcal{T}$ .  $\square$

Wir haben schon beobachtet (Beobachtung 3.15), dass Seitenflächen wieder Polyeder sind. Wir stellen weiter fest, dass sich die Seitenflächen übertragen.

**Beobachtung 3.37.** Sei  $\mathcal{P} = \mathcal{P}(A, \mathbf{b}) \subseteq \mathbb{R}^n$  ein Polyeder mit Seitenfläche  $\mathcal{F} \subseteq \mathcal{P}$ . Dann ist  $\mathcal{G} \subseteq \mathcal{F}$  Seitenfläche von  $\mathcal{P}$  genau dann, wenn  $\mathcal{G}$  Seitenfläche von  $\mathcal{F}$  ist.

*Beweis.* Nach Korollar 3.36 können wir  $\mathcal{F}$  als Polyeder der Form

$$\mathcal{F} = \mathcal{P}(D, \mathbf{d}) := \{\mathbf{x} \in \mathbb{R}^n : A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} \leq \mathbf{b}_{\text{eq}(\mathcal{F})}, -A_{\text{eq}(\mathcal{F})} \cdot \mathbf{x} \leq -\mathbf{b}_{\text{eq}(\mathcal{F})}, A\mathbf{x} \leq \mathbf{b}\}$$

betrachten. Bezeichne mit  $\text{eq}_{\mathcal{F}}(\cdot)$  und  $\text{fa}_{\mathcal{F}}(\cdot)$  die aktiven Ungleichungen bzw. die induzierten Seitenflächen bezüglich der Beschreibung  $\mathcal{P}(D, \mathbf{d})$ . Wegen  $\mathcal{G} \subseteq \mathcal{F}$  ist  $\text{eq}(\mathcal{F}) \subseteq \text{eq}(\mathcal{G})$ . Damit gilt

$$\begin{aligned} \text{fa}_{\mathcal{F}}(\text{eq}_{\mathcal{F}}(\mathcal{G})) &= \{\mathbf{x} \in \mathcal{F} : D_{\text{eq}_{\mathcal{F}}(\mathcal{G})} \cdot \mathbf{x} = \mathbf{d}_{\text{eq}_{\mathcal{F}}(\mathcal{G})}\} \\ &= \mathcal{F} \cap \{\mathbf{x} \in \mathcal{P} : D_{\text{eq}_{\mathcal{F}}(\mathcal{G})} \cdot \mathbf{x} = \mathbf{d}_{\text{eq}_{\mathcal{F}}(\mathcal{G})}\} \\ &= \{\mathbf{x} \in \mathcal{P} : A_{\text{eq}(\mathcal{G})} \cdot \mathbf{x} = \mathbf{d}_{\text{eq}(\mathcal{G})}\} \\ &= \text{fa}(\text{eq}(\mathcal{G})), \end{aligned}$$

also ist  $\mathcal{G}$  nach Beobachtung 3.18 und Korollar 3.36 Seitenfläche von  $\mathcal{F}$  genau dann, wenn es Seitenfläche von  $\mathcal{P}$  ist.  $\square$

Wir können nun Ecken und Facetten über ihre Dimension charakterisieren.

**Proposition 3.38.** Eine Seitenfläche  $\mathcal{F}$  eines Polyeders  $\mathcal{P} \subseteq \mathbb{R}^n$  ist eine Ecke genau dann, wenn  $\dim(\mathcal{F}) = 0$  und eine Facette genau dann, wenn  $\dim(\mathcal{F}) = \dim(\mathcal{P}) - 1$ .

*Beweis.* Die Aussage für Ecken gilt per Definition. Ist  $\mathcal{F}$  keine Facette, dann existiert eine Seitenfläche  $\mathcal{G}$  mit  $\mathcal{F} \subsetneq \mathcal{G} \subsetneq \mathcal{P}$ . Aus Satz 3.35 und Beobachtung 3.37 folgt

$$\dim(\mathcal{F}) \leq \dim(\mathcal{G}) - 1 \leq \dim(\mathcal{P}) - 2.$$

Sei andersherum  $\dim(\mathcal{F}) \neq \dim(\mathcal{P}) - 1$ . Nach Satz 3.35 kann  $\mathcal{F}$  also nur eine Facette, und damit eine echte Seitenfläche, sein, wenn  $\dim(\mathcal{F}) \leq \dim(\mathcal{P}) - 2$  bzw.  $\text{rank}(A_{\text{eq}(\mathcal{F})}) \geq \text{rank}(A_{\text{eq}(\mathcal{P})}) + 2$ . Sei  $I \subset \text{eq}(\mathcal{F})$  so gewählt, dass  $\text{rank}(A_I) = \text{rank}(A_{\text{eq}(\mathcal{F})}) - 1$ . Nach Beobachtung 3.18 ist  $\mathcal{G} = \text{fa}(I)$  eine Seitenfläche von  $\mathcal{P}$  und nach Korollar 3.36 ist  $\mathcal{F} = \text{fa}(\text{eq}(\mathcal{F}))$ . Damit gilt

$$\mathcal{G} = \text{fa}(I) \supseteq \text{fa}(\text{eq}(\mathcal{F})) = \mathcal{F}.$$

Nach Satz 3.35 ist

$$\dim(\mathcal{G}) = n - \text{rank}(A_I) = n - \text{rank}(A_{\text{eq}(\mathcal{F})}) + 1 = \dim(\mathcal{F}) + 1 < \dim(\mathcal{P}).$$

Also ist  $\mathcal{G} \neq \mathcal{F}$  eine echte Seitenfläche, die  $\mathcal{F}$  enthält. Damit ist  $\mathcal{F}$  keine Facette.  $\square$

Wichtig sind außerdem eindimensionale Seitenflächen.

---

**Definition 3.39.** Eine Seitenfläche  $\mathcal{F}$  eines Polyeders  $\mathcal{P}$  ist eine *Kante* wenn  $\dim(\mathcal{F}) = 1$ . Existieren  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $\mathbf{y} \neq \mathbf{0}$  mit  $\mathcal{F} = \{\mathbf{x}\} + \text{lin}(\mathbf{y}) \equiv \text{aff}(\{\mathbf{x}, \mathbf{y}\})$ , so heißt  $\mathcal{F}$  *Extremalgerade* von  $\mathcal{P}$ . Existieren  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $\mathbf{y} \neq \mathbf{0}$  mit  $\mathcal{F} = \{\mathbf{x}\} + \text{cone}(\mathbf{y})$ , so heißt  $\mathcal{F}$  *Extremalstrahl* von  $\mathcal{P}$ .

**Beobachtung 3.40.** Jede Kante eines Polyeders, die keine Extremalgerade und kein Extremalstrahl ist, ist die Verbindungsstrecke  $\text{conv}(\{\mathbf{x}, \mathbf{y}\})$  zweier Ecken  $\mathbf{x}, \mathbf{y}$ . Diese Ecken heißen dann *adjazent* auf  $\mathcal{P}$ .

Wir können auch relativ innere Punkte über ihre aktiven Ungleichungen charakterisieren.

**Satz 3.41.** Sei  $\mathcal{F}$  eine Seitenfläche eines Polyeders  $\mathcal{P}(A, \mathbf{b})$  und  $\bar{\mathbf{x}} \in \mathcal{F}$ . Dann ist  $\bar{\mathbf{x}}$  genau dann ein relativ innerer Punkt von  $\mathcal{F}$ , wenn  $\text{eq}(\{\bar{\mathbf{x}}\}) = \text{eq}(\mathcal{F})$ .

*Beweis.* Sei  $\mathcal{G}$  eine Seitenfläche von  $\mathcal{F}$  mit  $\bar{\mathbf{x}} \in \mathcal{G}$ . Aus  $\bar{\mathbf{x}} \in \mathcal{G} \subseteq \mathcal{F}$  folgt  $\text{eq}(\bar{\mathbf{x}}) \supseteq \text{eq}(\mathcal{G}) \supseteq \text{eq}(\mathcal{F})$ . Ist  $\text{eq}(\{\bar{\mathbf{x}}\}) = \text{eq}(\mathcal{F})$ , dann auch  $\text{eq}(\mathcal{G}) = \text{eq}(\mathcal{F})$  und damit, nach Korollar 3.36, auch

$$\mathcal{G} = \text{fa}(\text{eq}(\mathcal{G})) = \text{fa}(\text{eq}(\mathcal{F})) = \mathcal{F}.$$

Also ist  $\bar{\mathbf{x}}$  relativ innerer Punkt von  $\mathcal{F}$ .

Ist  $\text{eq}(\{\bar{\mathbf{x}}\}) \neq \text{eq}(\mathcal{F})$ , dann gilt  $\text{eq}(\mathcal{F}) \subsetneq \text{eq}(\{\bar{\mathbf{x}}\})$  wegen  $\bar{\mathbf{x}} \in \mathcal{F}$ . Aber dann ist  $\bar{\mathbf{x}} \in \text{fa}(\text{eq}(\{\bar{\mathbf{x}}\}))$  und nach Beobachtung 3.18 ist  $\text{fa}(\text{eq}(\{\bar{\mathbf{x}}\}))$  eine Seitenfläche von  $\mathcal{P}$ . Wegen  $\text{eq}(\mathcal{F}) \neq \text{eq}(\{\bar{\mathbf{x}}\})$  gilt  $\text{fa}(\text{eq}(\{\bar{\mathbf{x}}\})) \neq \mathcal{F}$  und damit

$$\text{fa}(\text{eq}(\{\bar{\mathbf{x}}\})) \subsetneq \text{fa}(\text{eq}(\mathcal{F})) = \mathcal{F}.$$

Nach Beobachtung 3.37 ist die Seitenfläche  $\text{fa}(\text{eq}(\{\bar{\mathbf{x}}\})) \subseteq \mathcal{F}$  von  $\mathcal{P}$  auch Seitenfläche von  $\mathcal{F}$  und damit  $\bar{\mathbf{x}}$  kein relativ innerer Punkt von  $\mathcal{F}$ . □

---

## 4 Grundlagen der Linearen Optimierung

Wir beschäftigen uns nun speziell mit linearen Optimierungsproblemen, die also eine affin lineare Zielfunktion haben und deren Nebenbedingungen mittels affin linearer Funktionen beschrieben werden können. Wir werden die zulässige Menge solcher Probleme oft als Polyeder der Form  $\mathcal{P}(A, \mathbf{b})$  ausdrücken.

**Definition 4.1.** Ein *lineares Programm (LP)* ist ein lineares Optimierungsproblem, das in der Form

$$\max \mathbf{c}^\top \mathbf{x} \quad \text{s.t.} \quad A\mathbf{x} \leq \mathbf{b},$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$  für  $n, m \in \mathbb{N}$  geschrieben werden kann. Diese Form heißt die *natürliche Form* des LPs.

---

### 4.1 Duales Problem und schwache Dualität

---

Betrachte beispielsweise das folgende lineare Programm in natürlicher Form:

$$\begin{aligned} \max \quad & 5x_1 + 3x_2 + 4x_3 \\ \text{s.t.} \quad & 3x_1 + 2x_2 + x_3 \leq 6, \\ & x_1 + x_2 + x_3 \leq 5, \\ & 2x_1 + 2x_3 \leq 4, \\ & -x_i \leq 0 \quad \text{für } i \in \{1, 2, 3\}. \end{aligned}$$

Die Lösung  $x_1 = 1, x_2 = 1, x_3 = 1$  ist offensichtlich zulässig und erreicht einen Zielfunktionswert von 12. Wir können einfach beweisen, dass diese Lösung nicht optimal ist: Dazu brauchen wir lediglich eine Lösung mit besserem Zielfunktionswert, z.B.  $x_1 = 0, x_2 = 2, x_3 = 2$  mit Wert 14. Die Lösung  $\mathbf{x} = (0, 2, 2)^\top$  ist sogar optimal. Aber wie können wir das formal beweisen? Wir können versuchen den optimalen Zielfunktionswert mit Hilfe der Nebenbedingungen zu beschränken. Da alle Variablen nicht-negativ sind (!), impliziert die erste Nebenbedingung zum Beispiel

$$5x_1 + 3x_2 + 4x_3 \leq 4 \cdot (3x_1 + 2x_2 + x_3) \leq 24.$$

Mit Hilfe der zweiten Nebenbedingung bekommen wir

$$5x_1 + 3x_2 + 4x_3 \leq 1 \cdot (3x_1 + 2x_2 + x_3) + 3 \cdot (x_1 + x_2 + x_3) \leq 21.$$

Nehmen wir die dritte Nebenbedingung hinzu sogar beispielsweise

$$5x_1 + 3x_2 + 4x_3 \leq 1 \cdot (3x_1 + 2x_2 + x_3) + 1 \cdot (x_1 + x_2 + x_3) + 1 \cdot (2x_1 + 2x_3) \leq 15.$$

Wir haben also gezeigt, dass keine zulässige Lösung unseres linearen Programms einen besseren Wert als 15 erreichen kann. Damit haben wir aber immer noch nicht die Optimalität unserer Lösung mit Wert 14 gezeigt. Was ist die niedrigste obere Schranke, die wir durch Kombination von Ungleichungen erhalten können? Wir suchen also eine Gewichtung der drei Ungleichungen mit Faktoren  $y_1, y_2, y_3$ , so dass  $x_1$  in Summe mit Gewicht mindestens 5 vorkommt,  $x_2$  mit Gewicht mindestens 3 und  $x_3$  mit Gewicht mindestens 4. Wir können diese Frage wiederum als lineares Programm formulieren:

$$\begin{aligned} \min \quad & 6y_1 + 5y_2 + 4y_3 \\ \text{s.t.} \quad & 3y_1 + y_2 + 2y_3 \geq 5, \\ & 2y_1 + y_2 \geq 3, \\ & y_1 + y_2 + 2y_3 \geq 4, \\ & y_i \geq 0 \quad \text{für } i \in \{1, 2, 3\}. \end{aligned}$$

Beachte, dass wir die Gewichte der Ungleichungen nicht-negativ wählen müssen, damit die Ungleichungen gültig bleiben.

Jede zulässige Lösung des Minimierungsproblems ist per Konstruktion eine obere Schranke an die optimale Lösung des Maximierungsproblems. Die Lösung  $y_1 = 3/2, y_2 = 0, y_3 = 5/4$  ist zulässig und hat Wert 14. Wir haben also einen kurzen Beweis, dass unsere obige Lösung in der Tat optimal ist, denn

$$5x_1 + 3x_2 + 4x_3 \leq \frac{3}{2} \cdot (3x_1 + 2x_2 + x_3) + \frac{5}{4} (2x_1 + 2x_3) \leq \frac{3}{2} \cdot 6 + \frac{5}{4} \cdot 4 = 14.$$

Daraus folgt ebenfalls, dass  $y_1 = 3/2, y_2 = 0, y_3 = 5/4$  eine Optimallösung des Minimierungsproblems ist.

Im Allgemeinen müssen wir Variablen zulassen, die sowohl positiv als auch negativ werden können. In diesem Fall, können wir die Zielfunktion nicht durch eine Linearkombination von Ungleichungen von oben beschränken. Stattdessen müssen wir die Zielfunktion durch eine exakte Linearkombination ausdrücken. Wir führen das *duale LP* entsprechend wie folgt ein.

**Definition 4.2.** Für das (*primale*) LP

$$\begin{aligned} \text{(LP)} \quad & \max \quad \mathbf{c}^\top \mathbf{x} \\ & \text{s.t.} \quad \mathbf{Ax} \leq \mathbf{b}, \end{aligned}$$

ist das *duale* LP gegeben durch

$$\begin{aligned} \text{(DLP)} \quad & \min \quad \mathbf{b}^\top \mathbf{y} \\ & \text{s.t.} \quad \mathbf{A}^\top \mathbf{y} = \mathbf{c}, \\ & \quad \mathbf{y} \geq \mathbf{0}. \end{aligned}$$

Das Duale in der obigen Definition befindet sich in sogenannter *Standardform*, wobei die zulässige Menge in der Form  $\mathcal{P}^\bullet(\mathbf{A}, \mathbf{b})$  ausgedrückt ist. Diese Form wird im nächsten Kapitel eine zentrale Rolle spielen. Wir haben bereits in Korollar 3.7 festgestellt, dass sich die zulässige Menge immer durch Anwendung von Beobachtung 3.6 in diese Form bringen lässt.

**Beobachtung 4.3.** Jedes lineare Programm kann in Standardform

$$\min \mathbf{c}^\top \mathbf{x} \quad \text{s.t.} \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0},$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$  für  $n, m \in \mathbb{N}$ , überführt werden.

Wir zeigen, dass das primale LP dual zum dualen LP ist.

**Satz 4.4.** Das duale LP von (DLP) ist wieder (LP).

*Beweis.* Negieren wir den Zielfunktionswert, lässt sich (DLP) in natürlicher Form schreiben als

$$\begin{aligned} \max & -\mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & \begin{pmatrix} A^\top \\ -A^\top \\ -\mathbb{I} \end{pmatrix} \mathbf{y} \leq \begin{pmatrix} \mathbf{c} \\ -\mathbf{c} \\ \mathbf{0} \end{pmatrix}. \end{aligned}$$

Das hierzu duale LP ist per Definition

$$\begin{aligned} \min & \begin{pmatrix} \mathbf{c} \\ -\mathbf{c} \\ \mathbf{0} \end{pmatrix}^\top \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} \\ \text{s.t.} & (A \quad -A \quad -\mathbb{I}) \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{pmatrix} = -\mathbf{b}, \\ & \mathbf{u}, \mathbf{v}, \mathbf{w} \geq \mathbf{0}. \end{aligned}$$

Durch die Substitution  $\mathbf{x} := \mathbf{v} - \mathbf{u}$  und Eliminieren der Schlupfvariablen  $\mathbf{w}$  (siehe Beobachtung 3.6), erhalten wir das äquivalente Programm

$$\begin{aligned} \min & -\mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & -\mathbf{Ax} \geq -\mathbf{b}, \end{aligned}$$

was nach erneuter Negierung des Zielfunktionswerts genau (LP) ist. □

Unsere Konstruktion impliziert sofort, dass das duale LP immer eine Schranke an das primale LP liefert.

**Satz 4.5** (schwache Dualität). Für alle zulässigen Lösungen  $\mathbf{x}, \mathbf{y}$  eines primalen Maximierungsproblems der Form (LP) und des dazugehörigen dualen LPs (DLP) gilt  $\mathbf{c}^\top \mathbf{x} \leq \mathbf{b}^\top \mathbf{y}$ .

*Beweis.* Es gilt

$$\mathbf{c}^\top \mathbf{x} = (A^\top \mathbf{y})^\top \mathbf{x} = \mathbf{y}^\top \mathbf{Ax} \stackrel{\mathbf{y} \geq \mathbf{0}}{\leq} \mathbf{y}^\top \mathbf{b}. \quad \square$$

**Korollar 4.6.** Sind  $\mathbf{x}, \mathbf{y}$  zulässige Lösungen von (LP) bzw. (DLP) mit  $\mathbf{c}^\top \mathbf{x} = \mathbf{b}^\top \mathbf{y}$ , dann sind  $\mathbf{x}, \mathbf{y}$  Optimallösungen von (LP) bzw. (DLP).

Wir zeigen nun, dass das Duale LP sich immer genau als das lineare Programm ergibt, dass die beste Schranke an den primalen Zielfunktionswert mittels einer Linearkombination der primalen Ungleichungen bestimmt.

**Satz 4.7.** Gegeben seien dimensionsverträgliche Matrizen  $A, B, C, D$  und Vektoren  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ . Dann ist

$$(PA) \quad \begin{array}{ll} \min & \mathbf{a}^\top \mathbf{x} + \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & A\mathbf{x} + B\mathbf{y} \geq \mathbf{c}, \\ & C\mathbf{x} + D\mathbf{y} = \mathbf{d}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

ein lineares Programm. Das zu (PA) duale LP kann geschrieben werden in der Form

$$(DA) \quad \begin{array}{ll} \max & \mathbf{c}^\top \mathbf{u} + \mathbf{d}^\top \mathbf{v} \\ \text{s.t.} & A^\top \mathbf{u} + C^\top \mathbf{v} \leq \mathbf{a}, \\ & B^\top \mathbf{u} + D^\top \mathbf{v} = \mathbf{b}, \\ & \mathbf{u} \geq \mathbf{0}. \end{array}$$

*Beweis.* Wir können (PA) durch Anwendung von Beobachtung 3.6 in Standardform schreiben als

$$\begin{array}{ll} \min & \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ -\mathbf{b} \\ \mathbf{0} \end{pmatrix}^\top \begin{pmatrix} \mathbf{x} \\ \mathbf{y}^+ \\ \mathbf{y}^- \\ \mathbf{z} \end{pmatrix} \\ \text{s.t.} & \begin{pmatrix} A & B & -B & -\mathbb{1} \\ C & D & -D & \mathbb{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y}^+ \\ \mathbf{y}^- \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}, \\ & \mathbf{x}, \mathbf{y}^+, \mathbf{y}^-, \mathbf{z} \geq \mathbf{0}, \end{array}$$

wobei  $\mathbb{1}$  die Identitätsmatrix und  $\mathbb{0}$  die 0-Matrix, beide mit geeigneten Dimensionen, bezeichnet. Das duale Programm hierzu ist nach Satz 4.4

$$\begin{array}{ll} \max & \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}^\top \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \\ \text{s.t.} & \begin{pmatrix} A^\top & C^\top \\ B^\top & D^\top \\ -B^\top & -D^\top \\ -\mathbb{1} & \mathbb{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \leq \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ -\mathbf{b} \\ \mathbf{0} \end{pmatrix}. \end{array}$$

Dieses Programm ist äquivalent zu (DA). □

Insbesondere haben wir im einleitenden Beispiel genau das Duale konstruiert.

**Korollar 4.8.** Das zu

$$\max \mathbf{c}^\top \mathbf{x} \quad \text{s.t.} \quad \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

duale LP ist

$$\min \mathbf{b}^\top \mathbf{y} \quad \text{s.t.} \quad \mathbf{A}^\top \mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0}.$$

## 4.2 Beispiel: Batterierecycling

Wir betrachten ein prototypisches Beispiel, um die Bedeutung der Dualität zu veranschaulichen. Nehmen wir dazu an, wir kontrollieren ein Unternehmen, das sich darauf spezialisiert hat, Altbatterien zu verwerten. Der Einfachheit halber gebe es nur zwei unterschiedliche Akkupacks mit unterschiedlicher (stark vereinfachter) Zellchemie zu berücksichtigen:

Akkupack	Eisen [kg]	Cobalt [kg]	sonstige Metalle [kg]	Gesamtkosten [€]
NiMH	2	1	2	300
Li-ion	2	4	1	500

Wir haben einen Auftrag zur Lieferung von 500kg Eisen, 400kg Cobalt und 300kg sonstige Metalle, den wir so kostengünstig wie möglich bedienen wollen.

Sei dazu  $x_1$  eine Variable, die die Anzahl zu verwertender NiMH-Packs bezeichnet und sei  $x_2$  eine entsprechende Variable für Li-ion-Packs. Wegen der großen Anzahl benötigter Akkupacks, können wir näherungsweise fraktionale  $x_1, x_2 \geq 0$  zulassen. Dann lässt sich das Problem wie folgt als LP modellieren (siehe Abb. 4.1):

$$\begin{aligned} \text{(BR)} \quad & \min \quad 300x_1 + 500x_2 && \text{(Gesamtkosten in €)} \\ & \text{s.t.} \quad 2x_1 + 2x_2 \geq 500, && \text{(benötigtes Eisen in kg)} \\ & \quad \quad x_1 + 4x_2 \geq 400, && \text{(benötigtes Cobalt in kg)} \\ & \quad \quad 2x_1 + x_2 \geq 300, && \text{(benötigte sonstige Metalle in kg)} \\ & \quad \quad x_1 \geq 0, \\ & \quad \quad x_2 \geq 0. \end{aligned}$$

Wir können das Problem wie folgt grafisch lösen. Die Menge der zulässigen Lösungen mit Zielfunktionswert  $\gamma$  ist durch den Schnitt der zulässigen Menge  $\mathcal{P}$  mit der Geraden

$$\mathcal{G}_\gamma = \{\mathbf{x} \in \mathbb{R}^2 : 300x_1 + 500x_2 = \gamma\}$$

gegeben. Wir suchen den kleinsten Zielfunktionswert  $\gamma^*$  für den dieser Schnitt nicht leer ist. Dazu können wir von einer Geraden mit nichtleerem Schnitt ausgehen, zum Beispiel  $\mathcal{G}_{230.000}$  und diese in Richtung  $-\mathbf{c} = -\begin{pmatrix} 300 \\ 500 \end{pmatrix}$  parallel verschieben, bis sie zu einer Tangente der zulässigen Menge wird. In unserem Fall ist diese Tangente  $\mathcal{G}_{85.000}$  und ihr Schnitt mit der zulässigen Menge ist  $\left\{\begin{pmatrix} 200 \\ 50 \end{pmatrix}\right\}$  (siehe Abb. 4.1). In anderen Worten ist die optimale Lösung  $\mathbf{x}^*$  unseres Problems, 200 NiMH-Packs und 50 Li-ion-Packs zu verwerten. Die Kosten dieser Lösung sind 85.000€.

In höheren Dimensionen ist unser grafisches Lösungsverfahren nicht mehr praktikabel. Wie können wir im Allgemeinen vorgehen, um eine optimale Lösung zu bestimmen?

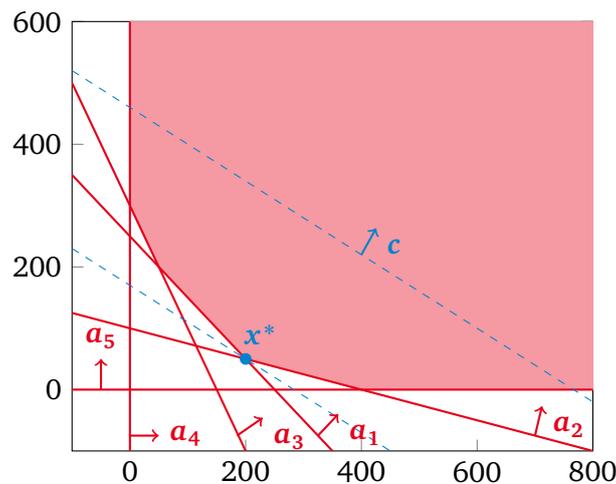


Abbildung 4.1: Grafische Darstellung des LPs (BR).

Wir stellen zunächst fest, dass die Optimallösung in unserem Beispiel eine Ecke bzw. ein Extrempunkt der zulässigen Menge ist. Nach Korollar 2.47 muss das der Fall sein, sofern ein Optimum existiert. Können wir also nachweisen, dass ein Optimum existiert, genügt es alle Ecken der zulässigen Menge abzusuchen.

In unserem Fall können wir wie folgt zeigen, dass eine optimale Lösung existiert. Zuerst zeigen wir, dass die zulässige Menge nicht leer ist, indem wir eine beliebige zulässige Lösung angeben, zum Beispiel  $\begin{pmatrix} 500 \\ 0 \end{pmatrix}$  oder  $\begin{pmatrix} 0 \\ 500 \end{pmatrix}$ . Der Punkt  $\begin{pmatrix} 500 \\ 0 \end{pmatrix}$  hat Kosten bzw. einen Zielfunktionswert von 150.000. Also wird die Optimallösung auf dem Schnitt der zulässigen Menge mit der kompakten Menge  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^2 : 300x_1 + 500x_2 \leq 150.000, \mathbf{x} \geq \mathbf{0}\}$  angenommen.

Damit wissen wir also, dass wir nur die Ecken absuchen müssen. Nach Satz 3.35 sind in unserem Fall (Raumdimension  $n = 2$ ) Ecken jeweils der Schnitt zweier Geraden der Form  $\{\mathbf{x} \in \mathbb{R}^2 : A_i \mathbf{x} = b_i\}$  für  $i \in \{1, \dots, 5\}$ . Es genügt also die  $\binom{5}{2} = 10$  entsprechenden Schnittpunkte zu berechnen, jeweils zu prüfen, ob sie zulässig sind, und den zulässigen Schnittpunkt auszuwählen, der den besten Zielfunktionswert hat. Und tatsächlich hat  $\mathbf{x}^*$  unter den 4 zulässigen Schnittpunkten den besten Zielfunktionswert.

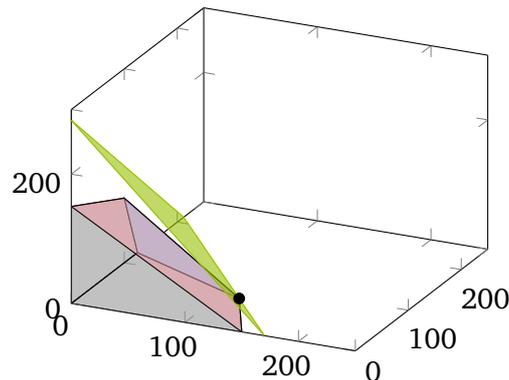
Im Allgemeinen sind Ecken über Schnittpunkte von  $\binom{m}{n}$  Hyperebenen definiert und auch dieses Lösungsverfahren unpraktikabel. Wir werden in Kapitel 5 ein systematischeres Verfahren kennenlernen, um die Ecken abzusuchen: Der Simplexalgorithmus geht von Ecke zu Ecke entlang Kanten der zulässigen Menge, so dass sich der Zielfunktionswert in jedem Schritt verbessert. Wir werden sehen, dass man auf diese Weise eine optimale Lösung findet, sofern eine existiert. Wären wir in unserem Beispiel so verfahren, hätten wir, angefangen von der Ecke  $\begin{pmatrix} 500 \\ 0 \end{pmatrix}$  des Polytops  $\mathcal{P} \cap \mathcal{C}$ , nur zwei weitere Ecken untersuchen müssen.

---

## Duales Problem

---

Um zu beweisen, dass  $\mathbf{x}^* = \begin{pmatrix} 200 \\ 50 \end{pmatrix}$  tatsächlich optimal ist, können wir das duale LP zu (BR) betrachten. Dazu müssen wir das Problem formulieren, die beste Schranke an (BR) durch Linearkombination der Ungleichungen zu finden. Wir führen Variablen  $\mathbf{y} \in \mathbb{R}^3$  ein, die beschreiben wie oft die drei ersten Ungleichungen von (BR) jeweils in die Schranke eingehen. Wir müssen



**Abbildung 4.2:** Illustration der zulässigen Menge und Zielfunktion (grün) von (DBR).

sicherstellen, dass  $x_1$  in der Linearkombination höchstens mit Vorfaktor 300 auftaucht und  $x_2$  höchstens mit Vorfaktor 500, damit wir eine untere Schranke erhalten (wegen  $\mathbf{x} \geq \mathbf{0}$  genügt das). Außerdem müssen wir  $\mathbf{y} \geq \mathbf{0}$  sicherstellen, damit unsere Ungleichungen mit diesen Faktoren multipliziert gültig bleiben. Die untere Schranke ergibt sich dann als Linearkombination der rechten Seiten. Wir erhalten also das duale LP

$$\begin{array}{ll}
 \text{(DBR)} & \max \quad 500y_1 + 400y_2 + 300y_3 \\
 & \text{s.t.} \quad 2y_1 + y_2 + 2y_3 \leq 300, \\
 & \quad \quad 2y_1 + 4y_2 + y_3 \leq 500, \\
 & \quad \quad \mathbf{y} \geq \mathbf{0}.
 \end{array}$$

Wir können auch hier wieder grafisch eine Optimallösung bestimmen (siehe Abb. 4.2), oder alle  $\binom{5}{3} = 10$  Schnittpunkte der Ungleichungen absuchen. Wir erhalten die Optimallösung  $\mathbf{y}^* = \frac{1}{6}(700, 400, 0)^\top$  mit Zielfunktionswert

$$500y_1^* + 400y_2^* + 300y_3^* = 85.000 = 300x_1^* + 500x_2^*.$$

Wir haben also eine straffe untere Schranke gefunden, die die Optimalität von  $\mathbf{x}^*$  (und  $\mathbf{y}^*$ ) beweist (vgl. Korollar 4.6). Wir werden das Prinzip der *starken Dualität* zeigen (Satz 4.17), das besagt, dass wir immer eine straffe untere Schranke als Linearkombination der Ungleichungen bestimmen können.

---

### Interpretation der dualen Lösung

---

Wir können die dualen Variablen wie folgt interpretieren. Angenommen, ein Rohstofflager möchte ein Konkurrenzangebot für die Lieferung machen. Welchen Preis  $y_1$  kann es für Eisen, welchen Preis  $y_2$  für Cobalt und welchen Preis  $y_3$  für sonstige Metalle verlangen, um konkurrenzfähig zu sein und dabei maximalen Profit zu erzielen? Die Nebenbedingungen von (DBR) fordern, dass die Preise nicht teurer sind als das Recycling von NiMH- bzw. Li-ion-Packs und die Zielfunktion maximiert den Profit. Der Profit der Optimallösung  $\mathbf{y}^*$  entspricht genau den Selbstkosten von  $\mathbf{x}^*$  aus (BR). Einen höheren Preis kann das Rohstofflager nicht fordern, ohne teurer als ein Recycling-Prozess zu werden. Die Werte der dualen Optimallösung  $\mathbf{y}^*$  werden auch *Schattenpreise* genannt.

In unserem Fall ist  $y_1^* = 700/6$  und  $y_2^* = 400/6$ , was den Kosten entspricht, die zusätzlich pro Einheit entstehen, wenn wir die Auslieferungsmenge von Eisen bzw. Cobalt (infinitesimal) erhöhen wollen. Anders gesagt, sind dies unsere Kosten pro produzierter Einheit, wobei wir keinen Überschuss haben. Im Gegensatz dazu ist  $y_3^* = 0$ , was bedeutet, dass wir einen Überschuss an sonstigen Metallen produzieren, also, dass die entsprechende Ungleichung nicht aktiv ist. Mehr davon zu liefern würde uns zunächst gar nichts kosten. Wie wir sehen werden, besteht immer eine solche Komplementarität zwischen inaktiven primalen Nebenbedingungen in der primalen Optimallösung und dualen Variablen im Träger der dualen Optimallösung (siehe Satz 4.20).

### 4.3 Zulässige Richtungen und Lösungen

Um starke Dualität herzuleiten, charakterisieren wir zunächst optimale Lösungen darüber, dass Sie sich in keiner Richtung lokal verbessern lassen. Wir definieren dazu zunächst Richtungen in denen wir lokal zulässig bleiben. Achtung, das entspricht nicht den Richtungen der konvexen zulässigen Menge gemäß Definition 2.23.

**Definition 4.9.** Ein Vektor  $s \in \mathbb{R}^n$  heißt *zulässige Richtung* in einer zulässige Lösung  $\bar{x} \in \mathbb{R}^n$  eines LPs, wenn  $\lambda > 0$  existiert, so dass  $\bar{x} + \lambda s$  eine zulässige Lösung des LPs ist. Die Menge aller zulässigen Richtungen in  $\bar{x}$  bezeichnen wir mit  $\mathcal{Z}(\bar{x}) \subseteq \mathbb{R}^n$ .

Für zulässige Mengen der Form  $\mathcal{P}^=(A, b)$ , müssen zulässige Richtungen in  $\ker(A)$  liegen und die Nichtnegativität erhalten. Wir beobachten, dass diese Bedingungen auch hinreichend sind.

**Beobachtung 4.10.** Sei  $A \in \mathbb{R}^{M \times N}$ ,  $b \in \mathbb{R}^M$  und  $c \in \mathbb{R}^N$  und  $\bar{x} \in \mathbb{R}^N$  eine zulässige Lösung von

$$\min c^\top x \quad \text{s.t.} \quad Ax = b, x \geq 0.$$

Dann ist die Menge der zulässigen Richtungen für  $\bar{x}$  gegeben durch

$$\mathcal{Z}(\bar{x}) = \{s \in \ker(A) : s_{N \setminus \text{supp}(\bar{x})} \geq 0\}.$$

*Beweis.* Für alle  $s \in \ker(A)$  und  $\lambda > 0$  gilt  $A(\bar{x} + \lambda s) = b$ . Außerdem existiert ein  $\lambda > 0$  mit  $\bar{x}_{\text{supp}(\bar{x})} + \lambda s_{\text{supp}(\bar{x})} \geq 0$ . Ist  $s_{N \setminus \text{supp}(\bar{x})} \geq 0$ , dann ist  $\bar{x} + \lambda s \geq 0$  also zulässige Lösung und somit  $s$  zulässige Richtung.

Ist andersherum  $s$  zulässige Richtung, dann existiert  $\lambda > 0$ , so dass  $\bar{x} + \lambda s$  zulässige Lösung ist. Dann ist  $A(\bar{x} + \lambda s) = b$  und damit  $s \in \ker(A)$ . Außerdem ist dann  $(\bar{x} + \lambda s) \geq 0$ , also  $s_{N \setminus \text{supp}(\bar{x})} \geq 0$ .  $\square$

Damit können wir zeigen, dass eine Lösung optimal ist genau dann, wenn sie keine zulässige Verbesserungsrichtung hat.

**Satz 4.11.** Sei  $A \in \mathbb{R}^{M \times N}$ ,  $b \in \mathbb{R}^M$  und  $c \in \mathbb{R}^N$ . Dann ist  $\bar{x} \in \mathcal{P}^=(A, b)$  genau dann Optimallösung von

$$\min c^\top x \quad \text{s.t.} \quad Ax = b, x \geq 0,$$

wenn

$$c^\top s \geq 0 \quad \forall s \in \mathcal{Z}(\bar{x}).$$

*Beweis.* Sei  $\bar{x}$  Optimallösung. Für alle  $s \in \mathcal{Z}(\bar{x})$  existiert  $\lambda > 0$ , so dass  $\bar{x} + \lambda s$  zulässig ist. Wegen Optimalität von  $\bar{x}$  gilt also  $c^\top \bar{x} \leq c^\top (\bar{x} + \lambda s)$ , also  $c^\top s \geq 0$ .

Sei  $\bar{x} \in \mathcal{P}^\circ(A, \mathbf{b})$  keine Optimallösung. Dann existiert eine zulässige Lösung  $\mathbf{x}^*$  mit  $c^\top \mathbf{x}^* < c^\top \bar{x}$ . Dann ist  $s := \mathbf{x}^* - \bar{x}$  zulässige Richtung für  $\bar{x}$  per Definition mit  $\lambda = 1$  (oder nach Beobachtung 4.10 wegen  $A s = \mathbf{0}$  und  $s_{N \setminus \text{supp}(\bar{x})} = \mathbf{x}_{N \setminus \text{supp}(\bar{x})}^* \geq \mathbf{0}$ ). Für diese Richtung gilt  $c^\top s = c^\top \mathbf{x}^* - c^\top \bar{x} < 0$ .  $\square$

Um diese Optimalitätsbedingung in eine nützlichere Form zu bringen, müssen wir den Trennungssatz (Satz 2.13) auf Polyeder übertragen. Dazu beobachten wir zunächst, dass die linke Seite der Nebenbedingungen einen Kegel definiert.

**Lemma 4.12.** Für alle  $A \in \mathbb{R}^{m \times n}$  ist die Menge  $\mathcal{K} = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq \mathbf{0}\}$  ein abgeschlossener, konvexer Kegel.

*Beweis.* Aus  $\mathbf{y} = A\mathbf{x} \in \mathcal{K}$  folgt  $\lambda \mathbf{y} = A(\lambda \mathbf{x}) \in \mathcal{K}$  mit  $\lambda \geq 0$ , daher ist  $\mathcal{K}$  ein Kegel. Für  $A\mathbf{x}_1 \in \mathcal{K}$ ,  $A\mathbf{x}_2 \in \mathcal{K}$  mit  $\mathbf{x}_1, \mathbf{x}_2 \geq \mathbf{0}$  und  $\lambda \in (0, 1)$  gilt  $\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \geq \mathbf{0}$ , also

$$\lambda A\mathbf{x}_1 + (1 - \lambda)A\mathbf{x}_2 = A(\lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \in \mathcal{K},$$

daher ist  $\mathcal{K}$  konvex. Wir zeigen, dass  $\mathcal{K}$  abgeschlossen ist.

Seien  $\mathcal{U} := \{I \subseteq \{1, \dots, n\} : \text{die Spalten von } A_I \text{ sind linear unabhängig}\}$  und  $\mathcal{K}_I := \{A_I \mathbf{x}_I : \mathbf{x}_I \in \mathbb{R}^I, \mathbf{x}_I \geq \mathbf{0}\}$  für  $I \in \mathcal{U}$ . Wir behaupten, dass

$$\mathcal{K} = \bigcup_{I \in \mathcal{U}} \mathcal{K}_I.$$

Offensichtlich gilt  $\mathcal{K}_I \subseteq \mathcal{K}$  für alle  $I \in \mathcal{U}$ . Für die Rückrichtung beobachten wir, dass sich jeder Vektor  $\mathbf{y} \in \mathcal{K}$  als konische Kombination der Spalten von  $A$  darstellen lässt. Nach Korollar 2.9 existiert  $I' \in \mathcal{U}$ , so dass sich  $\mathbf{y}$  als konische Kombination der linear unabhängigen Spalten in  $A_{I'}$  darstellen lässt. Damit ist  $\mathbf{y} \in \mathcal{K}_{I'}$  und somit  $\mathcal{K} \subseteq \bigcup_{I \in \mathcal{U}} \mathcal{K}_I$ .

Es bleibt zu zeigen, dass alle  $\mathcal{K}_I$  abgeschlossen sind, und damit auch ihre endliche Vereinigung  $\mathcal{K}$ . Sei dazu  $(\mathbf{y}_k)_{k \in \mathbb{N}} \subset \mathcal{K}_I$  eine konvergente Folge in  $\mathcal{K}_I$  mit Grenzwert  $\mathbf{y}$ . Wir müssen zeigen, dass  $\mathbf{y} \in \mathcal{K}_I$ . Es existiert eine Folge  $(\mathbf{x}_k)_{k \in \mathbb{N}} \subset \mathbb{R}^I$  mit  $\mathbf{x}_k \geq \mathbf{0}$  und  $\mathbf{y}_k = A_I \mathbf{x}_k$ . Da  $A_I$  vollen Spaltenrang hat, können wir eine invertierbare Teilmatrix  $A_{JI}$  finden. Also ist  $\mathbf{x}_k := A_{JI}^{-1}(\mathbf{y}_k)_J$  eindeutig bestimmt. Damit ist die Folge  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  das Bild einer stetigen Abbildung und konvergiert somit, mit Grenzwert  $\mathbf{x} \geq \mathbf{0}$  wegen  $\mathbf{x}_k \geq \mathbf{0}$ . Damit gilt

$$\mathbf{y} = \lim_{k \rightarrow \infty} \mathbf{y}_k = \lim_{k \rightarrow \infty} A_I \mathbf{x}_k = A_I \mathbf{x} \in \mathcal{K}_I. \quad \square$$

Der Trennungssatz (Satz 2.13) besagt nun, dass ein Vektor  $\mathbf{b}$  entweder im Kegel liegt, der von  $A_1, \dots, A_n$  aufgespannt wird, oder dass es eine Hyperebene gibt, die  $\mathbf{b}$  von  $A_1, \dots, A_n$  trennt. Das erlaubt folgende Charakterisierung von Zulässigkeit für Nebenbedingungen der Form  $\mathcal{P}^\circ(A, \mathbf{b})$ .

**Satz 4.13 (Farkas-Lemma).** Für  $A \in \mathbb{R}^{m \times n}$  und  $\mathbf{b} \in \mathbb{R}^m$  gilt genau eine der folgenden Aussagen:

- (i) Es existiert  $\mathbf{x} \geq \mathbf{0}$  mit  $A\mathbf{x} = \mathbf{b}$ .
- (ii) Es existiert  $\mathbf{y} \in \mathbb{R}^m$  mit  $\mathbf{y}^\top A \geq \mathbf{0}^\top$  und  $\mathbf{y}^\top \mathbf{b} < 0$ .

*Beweis.* Es können nicht beide Aussagen gelten: Existiert  $\mathbf{x} \geq \mathbf{0}$  mit  $A\mathbf{x} = \mathbf{b}$  und  $\mathbf{y} \in \mathbb{R}^m$  mit  $\mathbf{y}^\top A \geq \mathbf{0}^\top$ , dann gilt

$$0 \leq (\mathbf{y}^\top A)\mathbf{x} = \mathbf{y}^\top (A\mathbf{x}) = \mathbf{y}^\top \mathbf{b}.$$

Existiert kein  $\mathbf{x} \geq \mathbf{0}$  mit  $A\mathbf{x} = \mathbf{b}$ , dann ist  $\mathbf{b} \notin \mathcal{K}$ , wobei  $\mathcal{K} = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq \mathbf{0}\}$  den abgeschlossenen, konvexen Kegel aus Lemma 4.12 bezeichnet. Da  $\mathbf{0} \in \mathcal{K}$  ist  $\mathcal{K} \neq \emptyset$  und nach Satz 2.13 existiert eine Hyperebene, die  $\mathcal{K}$  und  $\{\mathbf{b}\}$  strikt trennt, also ein  $\mathbf{y} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$  und  $\delta \in \mathbb{R}$  mit

$$\mathbf{y}^\top \mathbf{z} \geq \delta \quad \forall \mathbf{z} \in \mathcal{K}, \quad \mathbf{y}^\top \mathbf{b} < \delta.$$

Wegen  $\mathbf{0} \in \mathcal{K}$  gilt  $\delta \leq 0$  und damit  $\mathbf{y}^\top \mathbf{b} < 0$ . Wir behaupten, dass  $\mathbf{y}^\top \mathbf{z} \geq 0$  für alle  $\mathbf{z} \in \mathcal{K}$ . Ist nämlich  $\mathbf{y}^\top \mathbf{z} < 0$  für  $\mathbf{z} \in \mathcal{K}$ , dann existiert  $\lambda \geq 0$  mit  $\mathbf{y}^\top (\lambda \mathbf{z}) < \delta$  aber  $\lambda \mathbf{z} \in \mathcal{K}$ , da  $\mathcal{K}$  ein Kegel ist.

Da  $A_i \in \mathcal{K}$  für alle  $i \in \{1, \dots, n\}$  folgt also  $\mathbf{y}^\top A_i \geq 0$  und damit

$$\mathbf{y}^\top A \geq \mathbf{0}^\top, \quad \mathbf{y}^\top \mathbf{b} < 0. \quad \square$$

Wir können das Farkas-Lemma natürlich auch für Nebenbedingungen der Form  $\mathcal{P}(A, \mathbf{b})$  formulieren.

**Korollar 4.14** (Übung). Für alle  $A \in \mathbb{R}^{m \times n}$  und  $\mathbf{b} \in \mathbb{R}^m$  gilt genau eine der folgenden Aussagen:

- (i) Es existiert  $\mathbf{x}$  mit  $A\mathbf{x} \leq \mathbf{b}$ .
- (ii) Es existiert  $\mathbf{y} \geq \mathbf{0}$  mit  $A^\top \mathbf{y} = \mathbf{0}$  und  $\mathbf{b}^\top \mathbf{y} < 0$ .

Im Allgemeinen erhalten wir die folgende Aussage.

**Korollar 4.15** (Übung). Für dimensionsverträgliche Matrizen  $A, B, C, D$  sowie Vektoren  $\mathbf{a}, \mathbf{b}$  hat genau eines der beiden folgenden Systeme eine Lösung:

$$\begin{array}{lcl} Ax + By \leq \mathbf{a}, & & \mathbf{u}^\top A + \mathbf{v}^\top C \geq \mathbf{0}, \\ Cx + Dy = \mathbf{b}, & \text{oder} & \mathbf{u}^\top B + \mathbf{v}^\top D = \mathbf{0}, \\ \mathbf{x} \geq \mathbf{0}. & & \mathbf{u} \geq \mathbf{0}, \\ & & \mathbf{u}^\top \mathbf{a} + \mathbf{v}^\top \mathbf{b} < 0. \end{array}$$

## 4.4 Optimalitätsbedingungen

Wir betrachten ein Paar von einem primalem Problem und dem dazugehörigen dualen Problem in der Form

$$\begin{array}{ll} \text{(P)} & \min \quad \mathbf{c}^\top \mathbf{x} \\ & \text{s.t.} \quad A\mathbf{x} = \mathbf{b}, \\ & \quad \quad \mathbf{x} \geq \mathbf{0}, \\ \text{(D)} & \max \quad \mathbf{b}^\top \mathbf{y} \\ & \text{s.t.} \quad A^\top \mathbf{y} \leq \mathbf{c}, \end{array}$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$ . Da jedes lineare Problem in Standardform überführt werden kann, gelten die entsprechenden Aussagen genauso für Probleme der allgemeineren Form (PA) und (DA).

Mit Hilfe der Farkas-Lemmas können wir jetzt die Optimalitätsbedingung aus Satz 4.11 in eine nützlichere Form, über die sogenannten *KKT-Bedingungen*, bringen.

**Satz 4.16** (Karush-Kuhn-Tucker-Bedingungen). Es ist  $\bar{\mathbf{x}} \in \mathbb{R}^n$  genau dann Optimallösung von (P), wenn  $\bar{\mathbf{y}} \in \mathbb{R}^m$  existiert, so dass

$$\begin{aligned} \text{(KKT)} \quad & A\bar{\mathbf{x}} = \mathbf{b}, \bar{\mathbf{x}} \geq \mathbf{0}, && \text{(primale Zulässigkeit)} \\ & A^\top \bar{\mathbf{y}} \leq \mathbf{c}, && \text{(duale Zulässigkeit)} \\ & \bar{x}_j = 0 \vee A_{\cdot j}^\top \bar{\mathbf{y}} = c_j \quad \forall j \in \{1, \dots, n\}. && \text{(Komplementarität)} \end{aligned}$$

Dann ist jedes solche  $\bar{\mathbf{y}}$  Optimallösung von (D) mit demselben Optimalwert wie (P).

*Beweis.* Sei  $\bar{\mathbf{x}}$  Optimallösung von (P). Nach Satz 4.11 existiert dann keine zulässige Richtung  $\mathbf{s} \in \mathcal{Z}(\bar{\mathbf{x}})$  für  $\bar{\mathbf{x}}$  mit  $\mathbf{c}^\top \mathbf{s} < 0$ . Nach Beobachtung 4.10 ist  $\mathcal{Z}(\bar{\mathbf{x}}) = \{\mathbf{s} \in \ker(A) : \mathbf{s}_Z \geq \mathbf{0}\}$  für  $A \in \mathbb{R}^{M \times N}$  und  $Z = N \setminus \text{supp}(\bar{\mathbf{x}})$ . Also existiert keine Lösung  $\mathbf{s} \in \mathbb{R}^n$  des Systems  $\mathbf{s}^\top D \geq \mathbf{0}^\top$  bzw.  $D^\top \mathbf{s} \geq \mathbf{0}$  mit  $\mathbf{s}^\top \mathbf{c} = \mathbf{c}^\top \mathbf{s} < 0$ , wobei

$$D^\top := \begin{pmatrix} A \\ -A \\ \mathbb{1}_Z \end{pmatrix} \quad \text{bzw.} \quad D = \begin{pmatrix} A^\top & -A^\top & \mathbb{1}_Z \end{pmatrix}$$

und  $\mathbb{1} \in \mathbb{R}^{n \times n}$  die Einheitsmatrix bezeichnet. Nach dem Farkas-Lemma (Satz 4.13) existiert also ein  $\mathbf{x} = (\mathbf{y}^+, \mathbf{y}^-, \mathbf{z}) \geq \mathbf{0}$  mit

$$\mathbf{c} = D\mathbf{x} = A^\top \mathbf{y}^+ - A^\top \mathbf{y}^- + \mathbb{1}_Z \mathbf{z}.$$

Also ist  $\bar{\mathbf{y}} := \mathbf{y}^+ - \mathbf{y}^-$  dual zulässig. Für  $j \in Z$  ist  $\bar{x}_j = 0$  und für  $j \notin Z$  ist  $(\mathbb{1}_Z)_j = \mathbf{0}^\top$  und damit

$$A_{\cdot j}^\top \bar{\mathbf{y}} = (A^\top)_j \cdot \bar{\mathbf{y}} = (A^\top)_j \cdot \mathbf{y}^+ - (A^\top)_j \cdot \mathbf{y}^- = c_j.$$

Also gilt (KKT).

Existiert umgekehrt für  $\bar{\mathbf{x}} \in \mathbb{R}^n$  ein  $\bar{\mathbf{y}} \in \mathbb{R}^m$ , so dass (KKT) gilt, dann ist  $\bar{\mathbf{x}}$  zulässige Lösung für (P) und  $\bar{\mathbf{y}}$  zulässige Lösung für (D). Weiter gilt dann wegen Komplementarität, dass

$$\mathbf{b}^\top \bar{\mathbf{y}} - \mathbf{c}^\top \bar{\mathbf{x}} = \bar{\mathbf{y}}^\top A\bar{\mathbf{x}} - \mathbf{c}^\top \bar{\mathbf{x}} = (\bar{\mathbf{y}}^\top A - \mathbf{c}^\top) \bar{\mathbf{x}} = 0.$$

Nach schwacher Dualität (Korollar 4.6) ist damit  $\bar{\mathbf{x}}$  Optimallösung von (P) und  $\bar{\mathbf{y}}$  Optimallösung von (D). □

Damit können wir nun die starke Dualität nachweisen, die insbesondere besagt, dass wir immer eine straffe Schranke an den primalen Zielfunktionswert aus einer Linearkombination der Nebenbedingungen herleiten können. Dieses zentrale Resultat ist invariant unter den Transformationsregeln aus Beobachtung 3.6 und gilt daher für primal-duale Paare in beliebiger Form.

**Satz 4.17** (starke Dualität). Folgende Aussagen sind äquivalent:

- (i) (P) und (D) sind beide zulässig, haben also zulässige Lösungen.
- (ii) (P) hat eine endliche Optimallösung.
- (iii) (P) und (D) haben Optimallösungen mit übereinstimmendem Zielfunktionswert.
- (iv) (D) hat eine endliche Optimallösung.
- (v) (P) ist zulässig und beschränkt oder (D) ist zulässig und beschränkt.

*Beweis.* Aus (i) folgt (ii): Da (D) zulässig ist, existiert eine zulässige Lösung  $\mathbf{z} \in \mathcal{P}(A^\top, \mathbf{c})$  für (D). Wegen schwacher Dualität (Satz 4.5) gilt  $\mathbf{c}^\top \mathbf{x} \geq \mathbf{b}^\top \mathbf{z}$  für alle  $\mathbf{x} \in \mathcal{P}^\circ(A, \mathbf{b})$ . Da (P) zulässig ist, ist  $\mathcal{P}^\circ(A, \mathbf{b}) \neq \emptyset$  und es existiert also ein endliches  $p^* := \inf\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}^\circ(A, \mathbf{b})\} \geq \mathbf{b}^\top \mathbf{z}$ . Wir zeigen, dass das Infimum angenommen wird, also eine primale Optimallösung existiert. Angenommen, das sei nicht der Fall. Dann existiert kein  $\mathbf{x} \geq \mathbf{0}$  mit

$$\begin{pmatrix} A \\ \mathbf{c}^\top \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{b} \\ p^* \end{pmatrix}.$$

Nach dem Farkas-Lemma (Satz 4.13) existiert dann aber  $\begin{pmatrix} \mathbf{y} \\ \gamma \end{pmatrix} \in \mathbb{R}^{m+1}$  mit

$$\begin{pmatrix} \mathbf{y}^\top & \gamma \end{pmatrix} \begin{pmatrix} A \\ \mathbf{c}^\top \end{pmatrix} = \mathbf{y}^\top A + \gamma \mathbf{c}^\top \geq \mathbf{0}^\top \quad (4.1)$$

und

$$\begin{pmatrix} \mathbf{y}^\top & \gamma \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ p^* \end{pmatrix} = \mathbf{y}^\top \mathbf{b} + \gamma p^* < 0. \quad (4.2)$$

Per Definition von  $p^*$  existiert eine Folge  $(\mathbf{x}_k) \subset \mathcal{P}^\circ(A, \mathbf{b})$  für die  $\lim_{k \rightarrow \infty} \mathbf{c}^\top \mathbf{x}_k = p^*$  gilt. Wegen  $\mathbf{x}_k \geq \mathbf{0}$  und  $A\mathbf{x}_k = \mathbf{b}$  folgt

$$0 \stackrel{(4.1)}{\leq} \gamma \mathbf{c}^\top \mathbf{x}_k + \mathbf{y}^\top A\mathbf{x}_k = \gamma \mathbf{c}^\top \mathbf{x}_k + \mathbf{y}^\top \mathbf{b} \stackrel{(4.2)}{<} \gamma \mathbf{c}^\top \mathbf{x}_k - \gamma p^* = \gamma(\mathbf{c}^\top \mathbf{x}_k - p^*) \rightarrow_{k \rightarrow \infty} 0,$$

ein Widerspruch zur Annahme, dass das Infimum  $p^*$  nicht angenommen wird. Also existiert  $\mathbf{x}^* \in \mathcal{P}^\circ(A, \mathbf{b})$  mit  $\mathbf{c}^\top \mathbf{x}^* = p^*$  und  $\mathbf{x}^*$  ist somit eine Optimallösung von (P).

Aus (ii) folgt (iii) nach Satz 4.16. Aus (iii) folgt trivialerweise (iv) und daraus trivialerweise (v). Aus (v) folgt (i): Da wir die zulässige Menge  $\mathcal{P}(A^\top, \mathbf{c})$  von (D) auf die Form  $\mathcal{P}^\circ(D, \mathbf{d})$  bringen können (Beobachtung 3.6), dürfen wir, ohne Beschränkung der Allgemeinheit, annehmen (P) habe eine zulässige Lösung  $\bar{\mathbf{x}} \in \mathcal{P}^\circ(A, \mathbf{b})$  und sei beschränkt. Wir müssen nun zeigen, dass auch (D) zulässig ist.

Angenommen (D) sei unzulässig, also  $\mathcal{P}(A^\top, \mathbf{c}) = \emptyset$ . Nach dem Farkas-Lemma (Korollar 4.14) existiert dann ein  $\mathbf{u} \geq \mathbf{0}$  mit  $A\mathbf{u} = \mathbf{0}$  und  $\mathbf{c}^\top \mathbf{u} < 0$ . Für alle  $\lambda > 0$  ist dann  $\bar{\mathbf{x}} + \lambda \mathbf{u} \in \mathcal{P}^\circ(A, \mathbf{b})$  zulässig für (P) mit

$$\lim_{\lambda \rightarrow \infty} \mathbf{c}^\top (\bar{\mathbf{x}} + \lambda \mathbf{u}) = \mathbf{c}^\top \bar{\mathbf{x}} + \lim_{\lambda \rightarrow \infty} \lambda \mathbf{c}^\top \mathbf{u} = -\infty.$$

Das ist ein Widerspruch zur Beschränktheit von (P). Also ist (D) zulässig.  $\square$

Insbesondere kann ein LP keine Lösung haben, wenn das Duale unbeschränkt ist, da wir dann jede Schranke an das Primale herleiten können. Andererseits bedeutet ein unzulässiges Duales, dass wir gar keine Schranke an das Primale herleiten können, was nur sein kann, wenn das Primale unzulässig oder unbeschränkt ist. Zusammengefasst gilt Folgendes – natürlich auch für primal-duale Paare in beliebiger Form und insbesondere umgekehrt.

**Korollar 4.18.** Ist (D) unbeschränkt, so ist (P) unzulässig. Ist (D) unzulässig, so ist (P) unzulässig oder unbeschränkt.

*Beweis.* Die erste Aussage folgt unmittelbar aus der schwachen Dualität (Satz 4.5). Ist (P) zulässig und beschränkt, dann folgt aus der starken Dualität (Satz 4.17 (v) und (i)), dass (D) zulässig sein muss. Daher gilt auch die zweite Aussage.  $\square$

**Beispiel 4.19.** Die beiden linearen Programme

$$\begin{array}{ll} \min & -x_2 \\ \text{s.t.} & x_1 - x_2 = 1, \\ & -x_1 + x_2 = 1, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \quad \text{und} \quad \begin{array}{ll} \max & y_1 + y_2 \\ \text{s.t.} & y_1 - y_2 \leq 0, \\ & -y_1 + y_2 \leq -1, \end{array}$$

sind dual zueinander und beide unzulässig.  $\triangle$

Paare aus einer primalen und dualen Optimallösung lassen sich durch Komplementarität charakterisieren. Allgemein formuliert, ist für jede inaktive primale/duale Ungleichung die zugehörige duale/primale Variable Null.

**Satz 4.20** (schwacher komplementärer Schlupf). Seien  $\bar{\mathbf{x}} \in \mathbb{R}^n$  und  $\bar{\mathbf{y}} \in \mathbb{R}^m$  zulässig für (P) bzw. (D). Dann sind folgende Aussagen äquivalent:

- (i)  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{y}}$  sind Optimallösungen
- (ii)  $(\mathbf{c} - \mathbf{A}^\top \bar{\mathbf{y}})^\top \bar{\mathbf{x}} = 0$
- (iii)  $\bar{x}_j = 0 \vee \bar{\mathbf{y}}^\top \mathbf{A}_{\cdot j} = c_j \quad \forall j \in \{1, \dots, n\}$

*Beweis.* Nach starker Dualität (Satz 4.17 (i) und (iii)) ist (i) äquivalent zu  $\mathbf{c}^\top \bar{\mathbf{x}} = \mathbf{b}^\top \bar{\mathbf{y}}$ , was wegen  $\mathbf{b} = \mathbf{A}\bar{\mathbf{x}}$  wiederum äquivalent ist zu (ii), denn

$$0 = \mathbf{c}^\top \bar{\mathbf{x}} - \bar{\mathbf{y}}^\top \mathbf{b} = (\mathbf{c}^\top - \bar{\mathbf{y}}^\top \mathbf{A})\bar{\mathbf{x}} = (\mathbf{c} - \mathbf{A}^\top \bar{\mathbf{y}})^\top \bar{\mathbf{x}}.$$

Das ist schließlich, wegen  $\bar{\mathbf{x}} \geq \mathbf{0}$  und  $\mathbf{A}^\top \bar{\mathbf{y}} \leq \mathbf{c}$ , bzw.  $\mathbf{c} - \mathbf{A}^\top \bar{\mathbf{y}} \geq \mathbf{0}$ , äquivalent zu (iii).  $\square$

Wir können jetzt die Gültigkeit der KKT-Bedingungen für (D), also die direkte Umkehrung von Satz 4.16, einfach beweisen. Für primal-duale Paare in anderer Form, müssen wir die KKT-Bedingungen im Allgemeinen umformulieren.

**Korollar 4.21.** Es ist  $\bar{\mathbf{y}} \in \mathbb{R}^m$  genau dann Optimallösung von (D), wenn  $\bar{\mathbf{x}} \in \mathbb{R}^n$  existiert, so dass (KKT) gilt. Dann ist  $\bar{\mathbf{x}}$  Optimallösung von (P) mit demselben Optimalwert wie (D).

*Beweis.* Ist  $\bar{\mathbf{y}} \in \mathbb{R}^m$  Optimallösung von (D), dann ist (iv) aus Satz 4.17 erfüllt. Also existiert nach Satz 4.17 (iii) eine Optimallösung  $\bar{\mathbf{x}} \in \mathbb{R}^n$  von (P) mit demselben Optimalwert wie (D). Nach Satz 4.20 und wegen der Zulässigkeit von  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{y}}$  gilt damit (KKT).  $\square$

Die schwachen komplementären Schlupfbedingungen lassen sich wieder allgemein formulieren, wie oben erwähnt.

**Korollar 4.22** (Übung). Seien  $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$  und  $\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix}$  zulässige Lösungen für (PA) bzw. (DA). Dann sind folgende Aussagen äquivalent:

(i)  $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$  ist optimal für (PA) und  $\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix}$  ist optimal für (DA).

(ii)  $(\mathbf{a} - (A^\top \bar{\mathbf{u}} + C^\top \bar{\mathbf{v}}))^\top \bar{\mathbf{x}} + \bar{\mathbf{u}}^\top ((A\bar{\mathbf{x}} + B\bar{\mathbf{y}}) - \mathbf{c}) = 0$ .

(iii) Für alle Komponenten  $\bar{u}_i$  gilt

$$\bar{u}_i = 0 \quad \text{oder} \quad (A\bar{\mathbf{x}} + B\bar{\mathbf{y}})_i = c_i$$

und für alle Komponenten  $\bar{x}_j$  gilt

$$\bar{x}_j = 0 \quad \text{oder} \quad (A^\top \bar{\mathbf{u}} + C^\top \bar{\mathbf{v}})_j = a_j.$$

In Satz 4.20 können  $\bar{x}_j$  und  $(\mathbf{y}^\top \mathbf{A}_{\cdot j} - c_j)$  gleichzeitig verschwinden. Es gibt jedoch immer auch Paare von *strikt komplementären* Optimallösungen, bei denen immer nur genau eine Komponente verschwindet:

**Satz 4.23** (starker komplementärer Schlupf, ohne Beweis). Sind (P) und (D) zulässig, so existieren Optimallösungen  $\bar{\mathbf{x}} \in \mathbb{R}^n$  und  $\bar{\mathbf{y}} \in \mathbb{R}^m$  von (P) bzw. (D), so dass für alle  $j \in \{1, \dots, n\}$  gilt

$$\bar{x}_j > 0 \quad \text{genau dann, wenn} \quad \mathbf{A}_{\cdot j}^\top \bar{\mathbf{y}} = c_j.$$

Der Satz vom starken komplementären Schlupf überträgt sich analog auf (PA) und (DA).

**Beispiel 4.24.** Die linearen Programme

$$\begin{array}{ll} \min & x_1 + 2x_2 \\ \text{s.t.} & x_1 + 2x_2 = 1, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \quad \text{und} \quad \begin{array}{ll} \max & y \\ \text{s.t.} & y \leq 1, \\ & 2y \leq 2, \end{array}$$

sind zueinander dual. Die Optimallösungen  $\mathbf{x}^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  und  $y^* = 1$  haben schwachen komplementären Schlupf (Satz 4.20 (iii)), aber keinen starken komplementären Schlupf (Satz 4.23), wegen  $x_2^* = 0$  aber  $2y^* \not\leq 2$ . Es existieren aber Optimallösungen  $\mathbf{x}^{**} = \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix}$  und  $y^{**} = 1$  mit starkem komplementären Schlupf. △

# 5 Der Simplex-Algorithmus

Wir betrachten in diesem Kapitel wieder das Paar von primalem Problem in Standardform und dem dazugehörigen dualen Problem in natürlicher Form

$$\begin{array}{ll}
 \text{(P)} & \min \quad \mathbf{c}^\top \mathbf{x} \\
 & \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \\
 & \quad \quad \mathbf{x} \geq \mathbf{0}, \\
 \text{(D)} & \max \quad \mathbf{b}^\top \mathbf{y} \\
 & \text{s.t.} \quad \mathbf{A}^\top \mathbf{y} \leq \mathbf{c},
 \end{array}$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$ . Wir nehmen in diesem Kapitel  $n \geq m$  an, sowie (vorerst)  $\text{rank}(A) = m$ , das heißt,  $A$  hat vollen Zeilenrang.

## 5.1 Basen, Basislösungen und Degeneriertheit

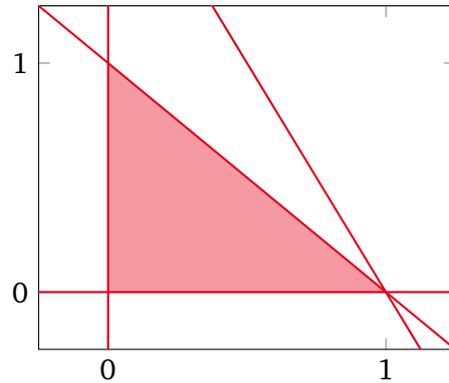
**Definition 5.1.** Habe  $A \in \mathbb{R}^{m \times n}$  vollen Zeilenrang. Sei  $B \subseteq \{1, \dots, n\}$  mit  $|B| = m$  und  $N_0 = \{1, \dots, n\} \setminus B$ , so dass  $A_{\cdot B}$  regulär ist. Dann ist  $B$  eine *Basis* von (P) und wir führen die folgenden Definitionen bezüglich  $B$  und (P) ein.

- (i) Die Basis  $B$  ist *primal zulässig*, falls  $A_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}$ .
- (ii) Die Basis  $B$  ist *dual zulässig*, falls  $\mathbf{z}_{N_0} := \mathbf{c}_{N_0} - A_{\cdot N_0}^\top A_{\cdot B}^{-\top} \mathbf{c}_B \geq \mathbf{0}$ . Dabei sind  $\mathbf{z}_{N_0}$  die *reduzierten Kosten*.
- (iii) Die Matrix  $A_{\cdot B}$  heißt *Basismatrix* und  $A_{\cdot N_0}$  heißt *Nichtbasismatrix*. Eine Variable  $x_j$ ,  $j \in \{1, \dots, n\}$  heißt *Basisvariable* falls  $j \in B$  und *Nichtbasisvariable* falls  $j \in N_0$ .
- (iv) Die zu  $B$  gehörige *Basislösung*  $\bar{\mathbf{x}} \in \mathbb{R}^n$  ist gegeben durch  $\bar{\mathbf{x}}_B = A_{\cdot B}^{-1} \mathbf{b}$ ,  $\bar{\mathbf{x}}_{N_0} = \mathbf{0}$ .
- (v) Die Basislösung  $\bar{\mathbf{x}}$  ist *degeneriert*, falls  $\text{supp}(\bar{\mathbf{x}}) \neq B$ .
- (vi) Die Basislösung  $\bar{\mathbf{x}}$  ist *zulässig*, falls  $B$  primal zulässig ist.
- (vii) Die Basis  $B$  ist *optimal*, falls die zugehörige Basislösung optimal ist.

Satz 3.25 impliziert, dass wir Ecken eines Polyeders der Form  $\mathcal{P}^=(A, \mathbf{b})$ , wobei  $A$  vollen Zeilenrang hat, über zulässige Basislösungen charakterisieren können.

**Satz 5.2.** Habe  $A \in \mathbb{R}^{m \times n}$  vollen Zeilenrang. Dann ist  $\hat{\mathbf{x}} \in \mathbb{R}^n$  eine Ecke von  $\mathcal{P}^=(A, \mathbf{b})$  genau dann, wenn  $\hat{\mathbf{x}}$  eine zulässige Basislösung von (P) ist.

*Beweis.* Sei  $\hat{\mathbf{x}}$  eine Ecke von  $\mathcal{P}^=(A, \mathbf{b})$ . Nach Satz 3.25 besteht  $A_{\cdot \text{supp}(\hat{\mathbf{x}})}$  aus linear unabhängigen Spalten. Wegen  $\text{rank}(A) = m$  hat der Spaltenraum von  $A$  Dimension  $m$ . Es folgt  $|\text{supp}(\hat{\mathbf{x}})| \leq m$  und dass wir durch Basisergänzung  $m - |\text{supp}(\hat{\mathbf{x}})|$  zusätzliche Spalten von  $A$  finden können, die mit den Spalten aus  $A_{\cdot \text{supp}(\hat{\mathbf{x}})}$  linear unabhängig sind (*Lineare Algebra I*). Also können wir



**Abbildung 5.1:** Zulässige Menge des Systems aus Beispiel 5.3.

$B \supseteq \text{supp}(\hat{\mathbf{x}})$  finden mit  $|B| = m$  und  $\text{rank}(A_B) = m$ . Damit ist  $A_B \in \mathbb{R}^{m \times m}$  invertierbar und somit  $B$  eine Basis von  $(P)$ .

Offensichtlich ist  $\hat{\mathbf{x}}_{N_0} = \mathbf{0}$ , da  $\text{supp}(\hat{\mathbf{x}}) \subseteq B = N \setminus N_0$ . Mit  $\hat{\mathbf{x}} \in \mathcal{P}^=(A, \mathbf{b})$  folgt  $\hat{\mathbf{x}} \geq \mathbf{0}$  und

$$\mathbf{b} = A\hat{\mathbf{x}} = A_B \hat{\mathbf{x}}_B + A_{N_0} \hat{\mathbf{x}}_{N_0} = A_B \hat{\mathbf{x}}_B.$$

Da  $A_B$  invertierbar ist, gilt also  $\mathbf{0} \leq \hat{\mathbf{x}}_B = A_B^{-1} \mathbf{b}$ . Also ist  $\hat{\mathbf{x}}$  Basislösung und  $B$  primal zulässig.

Ist andersherum  $\hat{\mathbf{x}}$  eine zulässige Basislösung mit zugehöriger Basis  $B$ , dann gilt  $\text{supp}(\hat{\mathbf{x}}) \subseteq B$  und die Spalten von  $A_B$  sind linear unabhängig. Also sind insbesondere die Spalten von  $A_{\text{supp}(\hat{\mathbf{x}})}$  linear unabhängig und  $\hat{\mathbf{x}}$  nach Satz 3.25 eine Ecke von  $\mathcal{P}^=(A, \mathbf{b})$ .  $\square$

Offensichtlich ist die zu einer Basis  $B$  gehörige Ecke über die Basislösung  $\mathbf{x}_B = A_B^{-1} \mathbf{b}$ ,  $\mathbf{x}_{N_0} = \mathbf{0}$  eindeutig bestimmt. Andersherum kann dieselbe Ecke  $\hat{\mathbf{x}}$  zu verschiedenen Basen gehören.

**Beispiel 5.3.** Betrachte das System

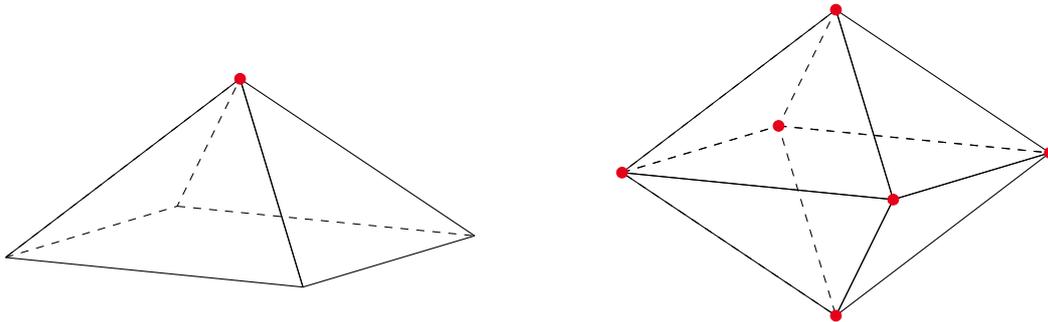
$$\begin{aligned} x_1 + x_2 &\leq 1, \\ 2x_1 + x_2 &\leq 2, \\ x_1, x_2 &\geq 0. \end{aligned}$$

Das zugehörige Polytop hat drei Ecken (siehe Abb. 5.1). Wir können das Polytop in die Form  $\mathcal{P}^=(A, \mathbf{b})$  bringen, mit

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

Zur Ecke  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  des ursprünglichen Systems gehört dann die degenerierte Ecke  $(1, 0, 0, 0)^\top \in \mathcal{P}^=(A, \mathbf{b})$ , die jeweils zugehörige Basislösung der Basen  $\{1, 2\}$ ,  $\{1, 3\}$  und  $\{1, 4\}$  ist. Die Degeneriertheit der Ecke resultiert aus der Redundanz der zweiten Ungleichung im ursprünglichen System. Im  $\mathbb{R}^2$  ist dies der einzige mögliche Grund für Degeneriertheit, da der Schnitt von mehr als zwei Halbräumen, die einen Randpunkt teilen, auf den Schnitt von nur zwei Halbräumen zurückgeführt werden kann.  $\triangle$

**Beispiel 5.4.** In höheren Dimensionen kann Degeneriertheit auch ohne redundante Nebenbedingungen auftreten (siehe Abb. 5.2).  $\triangle$



**Abbildung 5.2:** Beispiel für degenerierte Ecken (rot) zweier Polytope in  $\mathbb{R}^3$  auch in Beschreibungen ohne Redundanz. Degenerierte Ecken sind der Schnittpunkt von mehr als 3 Ebenen.

Nicht-degenerierte Ecken haben allerdings eine eindeutige Basis.

**Beobachtung 5.5.** Ist  $\bar{x}$  eine nicht-degenerierte Basislösung von (P), dann gehört  $\bar{x}$  zu einer eindeutigen Basis von (P). Die Umkehrung gilt im Allgemeinen nicht.

*Beweis.* Für eine zu  $\bar{x}$  gehörige Basis  $B$  gilt  $B \supseteq \text{supp}(\bar{x})$ . Ist  $\bar{x}$  nicht degeneriert, ist  $|\text{supp}(\bar{x})| = m$  und daher  $B = \text{supp}(\bar{x})$  eindeutig bestimmt.

Sei andersherum

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

Dann hat (P) genau die zwei Basen  $B = \{1, 3\}$  und  $B' = \{2, 3\}$ . Die zugehörigen Basislösungen  $\bar{x}, \bar{x}'$  sind gegeben durch  $\bar{x}_B = A_{\cdot B}^{-1} \mathbf{b} = \mathbf{b}$  und  $\bar{x}'_{B'} = A_{\cdot B'}^{-1} \mathbf{b} = \mathbf{b}$ , also  $\bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \bar{x}'$ . Damit gehören  $\bar{x}, \bar{x}'$  jeweils zu einer eindeutigen Basis, sind aber degeneriert, wegen  $\text{supp}(\bar{x}) = \{1\} \neq B$  und  $\text{supp}(\bar{x}') = \{2\} \neq B'$ .  $\square$

## 5.2 Die Grundversion des Simplex-Verfahrens

Das Grundprinzip des Simplexverfahrens besteht darin Ecken der zulässigen Menge systematisch nach dem Optimum abzusuchen. Wir haben bereits gesehen (Satz 4.11), dass eine Ecke genau dann optimal ist, wenn wir uns lokal nicht weiter verbessern können. Es genügt sich auf Verbesserungen entlang von Kanten zu beschränken, so dass wir das Optimum mittels lokaler Suche über die Ecken finden können. Der Simplexalgorithmus implementiert diese lokale Suche besonders elegant, indem er Ecken implizit über eine zugehörige Basis darstellt.

**Beobachtung 5.6** (Übung). Zwei Ecken von  $\mathcal{P}^=(A, \mathbf{b})$  sind genau dann adjazent, wenn sie Basislösungen zu Basen sind, die sich nur um ein Element unterscheiden.

Wir nehmen zunächst an, dass wir eine primal zulässige Basis  $B$  gegeben haben. Die zugehörige Basislösung  $\bar{x}$  ist dann zulässig für (P) mit

$$\bar{x}_B = A_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}, \quad \bar{x}_{N_0} = \mathbf{0},$$

wobei  $N_0 = \{1, \dots, n\} \setminus B$ . Da  $\bar{x}$  eindeutig von  $B$  bestimmt wird, können wir  $\bar{x}$  nur lokal ändern und zulässig bleiben, indem wir eine Nichtbasisvariable  $x_j$ ,  $j \in N_0$  erhöhen. Die resultierenden Werte der Basisvariablen  $x_B$  sind wegen

$$\mathbf{b} = A\mathbf{x} = A_{\cdot B}\mathbf{x}_B + A_{\cdot N_0}\mathbf{x}_{N_0} = A_{\cdot B}\mathbf{x}_B + A_{\cdot j}x_j$$

gegeben durch

$$\mathbf{x}_B = A_{\cdot B}^{-1}(\mathbf{b} - A_{\cdot j}x_j). \quad (5.1)$$

Der Einfluss auf die Zielfunktion ist damit gegeben durch

$$\begin{aligned} \mathbf{c}^\top \mathbf{x} &= \mathbf{c}_B^\top \mathbf{x}_B + \mathbf{c}_{N_0}^\top \mathbf{x}_{N_0} \\ &\stackrel{(5.1)}{=} \mathbf{c}_B^\top A_{\cdot B}^{-1}(\mathbf{b} - A_{\cdot j}x_j) + c_j x_j \\ &= \mathbf{c}_B^\top A_{\cdot B}^{-1} \mathbf{b} + (c_j - \mathbf{c}_B^\top A_{\cdot B}^{-1} A_{\cdot j})x_j. \end{aligned} \quad (5.2)$$

Wir können die Zielfunktion also genau dann mittels einer Erhöhung von  $\bar{x}_j$  verbessern, wenn die reduzierten Kosten  $\bar{z}_j = (c_j - \mathbf{c}_B^\top A_{\cdot B}^{-1} A_{\cdot j})$  der Variable  $x_j$  negativ sind. Existiert keine Variable  $x_j$ ,  $j \in N_0$  für die das der Fall ist, dann ist  $B$  per Definition dual zulässig. Dann ist also keine lokale Verbesserung von  $\bar{x}$  möglich und  $\bar{x}$  tatsächlich optimal.

**Proposition 5.7.** Ist eine Basis  $B$  von (P) primal und dual zulässig, dann ist die zugehörige Basislösung optimal und  $\bar{\mathbf{y}} = A_{\cdot B}^{-\top} \mathbf{c}_B$  eine Optimallösung von (D).

*Beweis.* Sei  $\bar{\mathbf{x}}$  die zu  $B$  gehörige Basislösung und sei  $\bar{\mathbf{y}} = A_{\cdot B}^{-\top} \mathbf{c}_B$  die Lösung von  $A_{\cdot B}^\top \bar{\mathbf{y}} = \mathbf{c}_B$ . Wegen primaler Zulässigkeit ist  $\bar{\mathbf{x}}$  zulässig für (P). Wegen dualer Zulässigkeit von  $B$ , gilt  $\mathbf{c}_{N_0} - A_{\cdot N_0}^\top \bar{\mathbf{y}} = \mathbf{c}_{N_0} - A_{\cdot N_0}^\top A_{\cdot B}^{-\top} \mathbf{c}_B \geq \mathbf{0}$ . Es folgt, dass  $A^\top \bar{\mathbf{y}} \leq \mathbf{c}$ , also ist  $\bar{\mathbf{y}}$  zulässig für (D). Außerdem gilt  $\mathbf{c}_B - A_{\cdot B}^\top \bar{\mathbf{y}} = \mathbf{0}$  und  $\bar{\mathbf{x}}_{N_0} = \mathbf{0}$ . Damit ist die komplementäre-Schlupf-Bedingung erfüllt und  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{y}}$  nach Satz 4.20 jeweils optimal.  $\square$

Nehmen wir also an, es existiert eine Variable  $x_j$ ,  $j \in N_0$  mit  $\bar{z}_j < 0$ . Also können wir  $\bar{x}$  durch Erhöhung von  $\bar{x}_j$  verbessern. Die resultierende Veränderung von  $x_B$  ist nach (5.1) genau

$$-\mathbf{w}_B x_j := -A_{\cdot B}^{-1} A_{\cdot j} x_j.$$

Wir können  $\bar{x}_j$  maximal so erhöhen, dass  $\bar{\mathbf{x}}$  mit  $\bar{\mathbf{x}}_B \geq \mathbf{0}$  zulässig bleibt. Ist nun  $\mathbf{w}_B \leq \mathbf{0}$ , können wir  $\bar{x}_j$  also beliebig erhöhen und die Zielfunktion nach (5.2) und wegen  $\bar{z}_j < 0$  beliebig verringern.

**Beobachtung 5.8.** Sei  $B$  eine primal zulässige Basis von (P). Existiert  $j \in \{1, \dots, n\}$  mit  $c_j - \mathbf{c}_B^\top A_{\cdot B}^{-1} A_{\cdot j} < 0$  und  $A_{\cdot B}^{-1} A_{\cdot j} \leq \mathbf{0}$ , dann ist (P) unbeschränkt.

Existiert andererseits  $i \in B$  mit  $w_i > 0$ , so ist wegen der Notwendigkeit von  $\bar{x}_i \geq 0$  die Erhöhung von  $\bar{x}_j$  nach (5.1) beschränkt durch  $\bar{x}_j \leq (A_{\cdot B}^{-1} \mathbf{b})_i / w_i$ . Insgesamt ist die Erhöhung von  $\bar{x}_j$  also durch eine Variable  $x_i$ ,  $i \in B$  beschränkt, die diesen Ausdruck minimiert. Wenn wir dann  $\bar{x}_j \leftarrow (A_{\cdot B}^{-1} \mathbf{b})_i / w_i$  setzen, verlässt  $i$  den Träger  $\text{supp}(\bar{\mathbf{x}})$  und wir können in der Basis  $i$  durch  $j$  ersetzen. Der Simplexalgorithmus geht wie beschrieben vor und lässt sich also wie folgt formulieren.

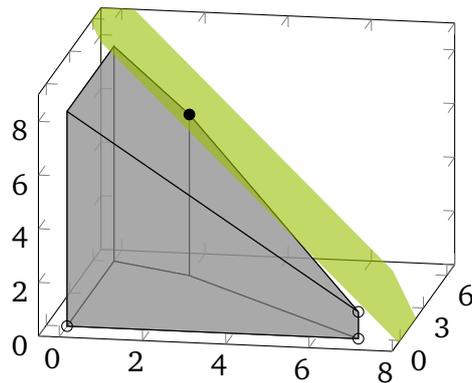


Abbildung 5.3: Zulässige Menge, Zielfunktion und Zwischenlösungen des Simplexalgorithmus in Beispiel 5.9.

---

**Algorithmus** :  $\text{SIMPLEX}(A, \mathbf{b}, \mathbf{c}, B, \bar{\mathbf{x}})$

---

**input:** Matrix  $A \in \mathbb{R}^{m \times n}$  mit vollem Zeilenrang,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$ ,  
 primal zulässige Basis  $B$  mit zugehöriger Basislösung  $\bar{\mathbf{x}}$

**output:** Optimallösungen  $\bar{\mathbf{x}}$  für (P) und  $\bar{\mathbf{y}}$  für (D) mit einer zugehörigen Basis  $B$ ,  
 oder  $-\infty$  falls (P) unbeschränkt

---

**repeat**

/\* Backward Transformation \*/

$\bar{\mathbf{y}} \leftarrow A_{.B}^{-\top} \mathbf{c}_B$  (Lösung von  $A_{.B}^{\top} \bar{\mathbf{y}} = \mathbf{c}_B$ )

/\* Pricing \*/

$N_0 \leftarrow \{1, \dots, n\} \setminus B$

$\bar{\mathbf{z}}_{N_0} \leftarrow \mathbf{c}_{N_0} - A_{.N_0}^{\top} \bar{\mathbf{y}}$  (reduzierte Kosten)

**if**  $\bar{\mathbf{z}}_{N_0} \geq \mathbf{0}$ :

└ **return**  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, B)$

/\* Entering Rule \*/

$j := \arg \min \{\bar{z}_\ell : \ell \in N_0, \bar{z}_\ell < 0\}$

/\* Forward Transformation \*/

$\mathbf{w}_B \leftarrow A_{.B}^{-1} \mathbf{A}_{.j}$  (Lösung von  $A_{.B} \mathbf{w}_B = \mathbf{A}_{.j}$ )

/\* Ratio Test \*/

**if**  $\mathbf{w}_B \leq \mathbf{0}$ :

└ **return**  $-\infty$

/\* Leaving Rule \*/

$i := \arg \min \{\frac{\bar{x}_\ell}{w_\ell} : \ell \in B, w_\ell > 0\}$

$\gamma \leftarrow \frac{\bar{x}_i}{w_i}$  (Schrittweite)

/\* Update \*/

$\bar{\mathbf{x}}_B \leftarrow \bar{\mathbf{x}}_B - \gamma \mathbf{w}_B$

$B \leftarrow B \cup \{j\} \setminus \{i\}$

$\bar{x}_j \leftarrow \gamma$

---

**Beispiel 5.9.** Wir betrachten das folgende lineare Programm (siehe Abb. 5.3):

$$\begin{aligned}
 \text{(LP)} \quad & \max \quad 3x_1 + 2x_2 + 2x_3 \\
 & \text{s.t.} \quad x_1 + x_3 \leq 8, \\
 & \quad \quad x_1 + x_2 \leq 7, \\
 & \quad \quad x_1 + 2x_2 \leq 12, \\
 & \quad \quad \mathbf{x} \geq \mathbf{0}.
 \end{aligned}$$

Durch Negieren und Einführungen von Schlupfvariablen (nach Bemerkung (3.6)) können wir das LP in Standardform überführen und erhalten

$$\begin{aligned}
 \text{(LP')} \quad & \min \quad (-3 \quad -2 \quad -2 \quad 0 \quad 0 \quad 0) \mathbf{x} \\
 & \text{s.t.} \quad \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 8 \\ 7 \\ 12 \end{pmatrix}, \\
 & \quad \quad \mathbf{x} \geq \mathbf{0}.
 \end{aligned}$$

Eine zulässige Basis ist  $B = \{4, 5, 6\}$  mit Basismatrix  $A_B = \mathbb{I}$  und zugehöriger Basislösung  $\bar{\mathbf{x}}$  gegeben durch  $\bar{\mathbf{x}}_B = A_B^{-1} \mathbf{b} = \begin{pmatrix} 8 \\ 7 \\ 12 \end{pmatrix}$ . Wir fassen die Schritte des Simplexalgorithmus tabellarisch zusammen.

step	$\bar{\mathbf{y}}$	$A_{\cdot N_0}^\top$	$\bar{\mathbf{z}}_{N_0}$	$j$	$\mathbf{w}_B$	$i$	$\gamma$	$B$	$A_B$	$A_B^{-1}$	$\bar{\mathbf{x}}_B$	$\mathbf{c}^\top \bar{\mathbf{x}}$
								$\{4, 5, 6\}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 7 \\ 12 \end{pmatrix}$	0
1	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} -3 \\ -2 \\ -2 \end{pmatrix}$	1	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	5	7	$\{1, 4, 6\}$	$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}$	-21
2	$\begin{pmatrix} 0 \\ -3 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$	3	$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$	4	1	$\{1, 3, 6\}$	$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}$	-23
3	$\begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$	2	$\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$	6	5	$\{1, 2, 3\}$	$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 2 & -1 \\ 0 & -1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}$	-28
4	$\begin{pmatrix} -2 \\ 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$									

In Iteration 4 terminiert der Algorithmus im Pricing-Schritt und gibt  $\bar{\mathbf{x}} = (2, 5, 6, 0, 0, 0)^\top$ ,  $\bar{\mathbf{y}} = \begin{pmatrix} -2 \\ 0 \\ -1 \end{pmatrix}$  und  $B = \{1, 2, 3\}$  zurück, was der Optimallösung  $\begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix}$  von (LP) mit Wert 28 entspricht.  $\triangle$

**Satz 5.10.** Terminiert SIMPLEX, so liefert der Algorithmus ein korrektes Ergebnis.

*Beweis.* Wir zeigen, dass der Algorithmus die Invariante erhält, dass  $B$  eine primal zulässige Basis mit zugehöriger Basislösung  $\bar{\mathbf{x}}$  ist. Seien dazu  $B$ ,  $\bar{\mathbf{x}}$  und  $B'$ ,  $\bar{\mathbf{x}}'$  die Werte vor und nach dem Update Schritt einer Iteration.

Wir weisen zuerst nach, dass  $B'$  eine Basis ist. Offensichtlich ist  $B' \subseteq \{1, \dots, n\}$  und  $|B'| = m$ . Angenommen  $A_{B'}$  wäre nicht regulär. Dann sind die Spalten von  $A_{B'}$  linear abhängig, es existiert also  $\boldsymbol{\lambda} \in \mathbb{R}^{m-1}$  so dass

$$A_{B' \setminus \{j\}} \boldsymbol{\lambda} = \mathbf{A}_{\cdot j}.$$

Damit ist

$$A_{B \setminus \{i\}} \boldsymbol{\lambda} = A_{B' \setminus \{j\}} \boldsymbol{\lambda} = \mathbf{A}_{\cdot j} = A_{\cdot B} \mathbf{w}_B = w_i \mathbf{A}_{\cdot i} + \sum_{\ell \in B \setminus \{i\}} w_\ell \mathbf{A}_{\cdot \ell}.$$

Da die Spalten von  $A_{\cdot B}$  nach der Invariante linear unabhängig sind, muss  $w_i = 0$  gelten. Das steht im Widerspruch zur Wahl von  $i$ . Also ist  $A_{\cdot B'}$  regulär und  $B'$  eine Basis.

Wegen der Wahl von  $i$ ,  $j$  und  $\gamma$  sowie  $\bar{\mathbf{x}} \geq \mathbf{0}$  gilt

$$\begin{aligned} \bar{x}'_i &= \bar{x}_i - \gamma w_i = 0, \\ \bar{x}'_j &= \gamma \geq 0, \\ \bar{x}'_\ell &= \bar{x}_\ell - \gamma w_\ell \geq 0 \quad \forall \ell \in B. \end{aligned}$$

Also ist  $\text{supp}(\bar{\mathbf{x}}') \subseteq B'$  und  $\bar{\mathbf{x}}' \geq \mathbf{0}$ . Deswegen gilt

$$A_{\cdot B'} \bar{\mathbf{x}}'_{B'} = A_{\cdot B \cup B'} \bar{\mathbf{x}}'_{B \cup B'} = A_{\cdot B} (\bar{\mathbf{x}}_B - \gamma \mathbf{w}_B) + \mathbf{A}_{\cdot j} \bar{x}'_j = \mathbf{b} - \gamma A_{\cdot B} A_{\cdot B}^{-1} \mathbf{A}_{\cdot j} + \gamma \mathbf{A}_{\cdot j} = \mathbf{b}$$

und  $\bar{\mathbf{x}}'$  ist also zugehörige zulässige Basislösung zu  $B'$ . Die Invariante ist somit erhalten.

Terminiert SIMPLEX weil die reduzierten Kosten nichtnegativ sind, dann ist  $B$  nach der Invariante eine primal und dual zulässige Basis. Nach der Invariante ist  $\bar{\mathbf{x}}$  die zugehörige Basislösung und nach Proposition 5.7 sind  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{y}}$  somit optimal. Also gibt der Algorithmus ein korrektes Ergebnis zurück.

Terminiert SIMPLEX wegen des Ratio-Tests, dann ist (P) nach Beobachtung 5.8 unbeschränkt. Also gibt der Algorithmus ein korrektes Ergebnis zurück.  $\square$

Es kann im Allgemeinen sein, dass SIMPLEX nicht terminiert.

**Beispiel 5.11** (Übung). SIMPLEX terminiert nicht für jede Wahl von eintretender und austretender Basisvariable.  $\triangle$

Sofern keine Ecke zu mehr als einer Basis gehört, können wir sicher sein, dass wir in jeder Iteration die aktuelle Ecke verlassen und somit die Zielfunktion strikt verbessern. Insbesondere können wir keine Ecke mehr als einmal besuchen.

**Satz 5.12.** Ist keine zulässige Basislösung von (P) degeneriert, so terminiert SIMPLEX.

*Beweis.* Wir haben im Beweis von Satz 5.10 gesehen, dass SIMPLEX die Invariante erhält, dass  $B$  eine primal zulässige Basis mit Basislösung  $\bar{\mathbf{x}}$  ist. Ist keine Basislösung von (P) degeneriert, so gilt also zu jeder Zeit  $\bar{\mathbf{x}}_B > \mathbf{0}$  während der Ausführung von SIMPLEX. Damit ist  $\gamma > 0$  und wegen  $\bar{z}_j < 0$  verbessert sich die Zielfunktion um  $\gamma \bar{z}_j < 0$ . Da die Basislösung zu einer Basis eindeutig ist, kann jede Basis somit höchstens einmal auftreten. Somit terminiert SIMPLEX nach höchstens  $\binom{n}{m}$  Iterationen.  $\square$

---

---

### 5.3 Pivotregeln

---

Das Simplex-Verfahren kann mit Auswahlregeln für die eintretende und austretende Basisvariable konfiguriert werden. Eine solche Auswahlregel wird *Pivotregel* genannt. Eine Funktion der Pivotregel kann es zum Beispiel sein, die Terminierung zu garantieren. Wir zeigen im Folgenden, wie das erreicht werden kann.

Nach Satz 5.12 terminiert das Simplex-Verfahren, sofern keine Ecke von (P) degeneriert ist. Das können wir durch Modifikation von (P) mittels der *Perturbationsmethode* erreichen. Dabei addieren wir eine kleine Störung  $\varepsilon \in \mathbb{R}^m$  mit  $\varepsilon_1 \gg \varepsilon_2 \gg \dots \gg \varepsilon_m > 0$  auf die rechte Seite  $\mathbf{b}$ . Jede Basislösung ist dann von der Form  $\bar{\mathbf{x}}_B = A_{\cdot B}^{-1}(\mathbf{b} + \varepsilon)$ ,  $\bar{\mathbf{x}}_{N_0} = \mathbf{0}$ . Da  $A_{\cdot B}^{-1}$  keine Nullzeile enthalten kann, enthält jede Zeile dieses Produkts einen Summanden, der ein Vielfaches mindestens eines Eintrags von  $\varepsilon$  ist und sich somit mit keinem anderen Summanden verrechnen kann. Damit ist  $\bar{\mathbf{x}}_B > \mathbf{0}$  und somit  $\text{supp}(\bar{\mathbf{x}}) = B$ , also ist  $\bar{\mathbf{x}}$  nicht degeneriert.

Natürlich wollen wir vermeiden mit sehr kleinen numerischen Störungen zu arbeiten (siehe *Einführung in die Numerik*). Wir können stattdessen wie folgt mit symbolischen Störungen rechnen. Dazu beobachten wir, dass die Störungen nur auf die Wahl der verlassenden Basisvariable einen Einfluss haben. Hier müssen wir statt  $\bar{x}_\ell/w_\ell = (A_{\cdot B}^{-1}\mathbf{b})_\ell/w_\ell$  nun das Verhältnis  $(A_{\cdot B}^{-1}(\mathbf{b} + \varepsilon))_\ell/w_\ell$  betrachten. Für alle Elemente von  $\arg \min\{\bar{x}_\ell/w_\ell : \ell \in B, w_\ell > 0\}$  müssen wir also die Zeile  $(A_{\cdot B}^{-1}\varepsilon)_\ell/w_\ell$  vergleichen. Im perturbierten System wählen wir die austretende Variable so  $\arg \min\{\bar{x}_\ell/w_\ell : \ell \in B, w_\ell > 0\}$ , dass die Zeile  $(A_{\cdot B}^{-1}\varepsilon)_\ell/w_\ell$  lexikographisch minimal ist (also im ersten ungleichen Eintrag kleiner als jede andere). Beachte, dass die lexikographisch minimale Zeile wegen Regularität von  $A_{\cdot B}^{-1}$  eindeutig ist. Wählen wir in `SIMPLEX` also diese Variable als austretende Basisvariable, landen wir bei der gleichen optimalen Basis wie für das perturbierte System. Diese Pivotregel wird die *lexikografische Regel* genannt.

**Satz 5.13.** Unter Verwendung der lexikographischen Regel terminiert `SIMPLEX`.

*Beweis.* Folgt aus der Äquivalenz der lexikographischen Regel und der Perturbationsmethode zusammen mit Satz 5.12.  $\square$

Eine weitere Pivotregel, die die Terminierung des Simplex-Algorithmus garantiert, ist die *Bland Regel*, die die Indizes  $j$  und  $i$  jeweils kleinstmöglich wählt.

**Satz 5.14.** Unter Verwendung der Bland Regel terminiert `SIMPLEX`.

*Beweis.* Terminiert `SIMPLEX` nicht innerhalb von  $\binom{n}{m}$  Iterationen, dann tritt dieselbe Basis  $B$  in zwei Iterationen  $k$  und  $\ell$  mit  $k < \ell$  auf. Da sich der Zielfunktionswert in jeder Iteration um genau  $\gamma\bar{z}_j \leq 0$  verbessert und in Iterationen  $k$  und  $\ell$  identisch ist, folgt  $\gamma = 0$  aus  $\bar{z}_j < 0$  für alle Iterationen  $k, k+1, \dots, \ell-1$ .

Jede Variable  $x_i$ , die im Laufe der Iterationen  $k, k+1, \dots, \ell-1$  in die Basis eintritt, muss sie innerhalb dieser Iterationen auch verlassen (haben) und umgekehrt. Seien  $F \subseteq \{1, \dots, n\}$  die Indizes dieser Variablen. Wegen  $\gamma = 0$  ändert sich  $\bar{\mathbf{x}}$  während der Iterationen  $k, k+1, \dots, \ell-1$  nicht und es gilt insbesondere  $\bar{\mathbf{x}}_F = \mathbf{0}$ .

Betrachte eine Iteration in der  $t := \max\{i \in F\}$  in die aktuelle Basis  $C$  eintritt. Seien  $N_C := \{1, \dots, n\} \setminus C$  und  $\bar{\mathbf{z}} := \mathbf{c} - A^T A_C^{-T} \mathbf{c}_C$ . Dann sind  $\bar{z}_{N_C}$  die reduzierten Kosten in dieser Iteration

und  $\bar{z}_C = \mathbf{0}$ . Damit gilt  $\bar{z}_t < 0$  und  $\bar{z}_r \geq 0$  für alle  $r \in F \setminus \{t\}$ , da sonst nach der Bland Regel nicht  $j = t$  zum Eintritt in  $C$  ausgewählt worden wäre.

Betrachte nun eine Iteration in der  $t$  die aktuelle Basis  $D$  verlässt und sei  $i = s$  der eintretende Index in dieser Iteration. Dann ist  $s \in F$  und  $s < t$ . Da  $\bar{x}_F = \mathbf{0}$ , kann jedes  $r \in F \cap D$  mit  $w_r > 0$  als austretende Variable ausgewählt werden. Da die Bland Regel  $i = t$  wählt, muss also  $w_t > 0$  und  $w_r \leq 0$  für alle  $r \in F \cap D \setminus \{t\}$  gelten. Wir definieren nun den Vektor  $\mathbf{d} \in \mathbb{R}^n$  über  $\mathbf{d}_D := -\mathbf{w}_D$ ,  $d_s := 1$  und  $\mathbf{d}_{N_D \setminus \{s\}} := \mathbf{0}$  für  $N_D := \{1, \dots, n\} \setminus D$ . Es gilt damit also  $d_t < 0$  und  $d_r \geq 0$  für alle  $r \in F \setminus \{t\}$ .

Wir zeigen, dass  $\bar{\mathbf{z}}^\top \mathbf{d} > 0$ : Es ist  $\bar{z}_t < 0$  und  $d_t < 0$ , also  $\bar{z}_t \cdot d_t > 0$ . Für  $r \in F \setminus \{t\}$  gilt  $\bar{z}_r \geq 0$  und  $d_r \geq 0$ . Für  $r \notin F$  gilt entweder  $r \in C \cap D$  und damit  $\bar{z}_r = 0$  oder  $r \notin C \cup D$  und damit  $d_r = 0$ .

Weiterhin gilt

$$A\mathbf{d} = \mathbf{A}_{\cdot s} - A_{\cdot D}\mathbf{w}_D = \mathbf{A}_{\cdot s} - A_{\cdot D}A_{\cdot D}^{-1}\mathbf{A}_{\cdot s} = \mathbf{0}. \quad (5.3)$$

Damit ist also

$$0 < \bar{\mathbf{z}}^\top \mathbf{d} = (\mathbf{c}^\top - \mathbf{c}_C^\top A_{\cdot C}^{-1}A)\mathbf{d} \stackrel{(5.3)}{=} \mathbf{c}^\top \mathbf{d} = c_s - \mathbf{c}_D^\top \mathbf{w}_D = c_s - \mathbf{c}_D^\top A_{\cdot D}^{-1}\mathbf{A}_{\cdot s}.$$

Das ist ein Widerspruch zur Wahl von  $s$  als eintretenden Index, da die rechte Seite die reduzierten Kosten von  $\bar{x}_s$  in Iteration  $\ell$  beschreibt, welche negativ sein müssen.  $\square$

Es wurden eine Vielzahl von Pivotregeln für das Simplex-Verfahren untersucht. Im Folgenden findet sich eine Übersicht über die wichtigsten Regeln. Außer der Bland Regel können alle Regeln mit der lexikographischen Regel kombiniert werden, so dass Terminierung sichergestellt ist.

**lexikographische Regel** Wähle  $i \in \arg \min\{\bar{x}_\ell/w_\ell : \ell \in B, w_\ell > 0\}$  so, dass  $(A_{\cdot B}^{-1})_{\cdot i}/w_i$  lexikographisch minimal ist.

**Bland Regel** Wähle  $j = \min\{\ell \in N_0 : \bar{z}_\ell < 0\}$  und  $i = \min\{\ell \in \arg \min\{\frac{\bar{x}_\ell}{w_\ell} : \ell \in B, w_\ell > 0\}\}$ .

**Dantzig Regel** Wähle  $j \in \arg \min\{\bar{z}_\ell : \ell \in N_0\}$  mit negativsten reduzierten Kosten.

**größte Verbesserung** Wähle  $j$  so, dass die resultierende Verbesserung  $|\gamma\bar{z}_j|$  der Zielfunktion maximal ist.

**steilster Anstieg** Wähle  $j \in \arg \max\left\{\left(\frac{c_B}{c_\ell}\right)^\top \left(A_{\cdot B}^{-1}A_{\cdot \ell}\right) \left\|\left(A_{\cdot B}^{-1}A_{\cdot \ell}\right)\right\|^{-1} : \ell \in N_0, \bar{z}_\ell < 0\right\}$ .

**zufällige Kante** Wähle  $j \in \{\ell \in N_0 : \bar{z}_\ell < 0\}$  gleichverteilt zufällig.

**zufällige Facette** Schränke das Problem auf eine zufällige Facette ein, die die aktuelle Ecke enthält und wähle  $j$  rekursiv, sofern es nicht schon eindeutig bestimmt ist.

**Cunningham Regel** Sei  $j'$  der Wert der vorigen Iteration. Wähle  $j = \min\{\ell \geq j' : \ell \in \mathbb{N}, (\ell \bmod n) + 1 \in N_0, \bar{z}_{(\ell \bmod n) + 1} < 0\} \bmod n + 1$  als die nächste Variable in zyklischer Reihenfolge mit negativen reduzierten Kosten.

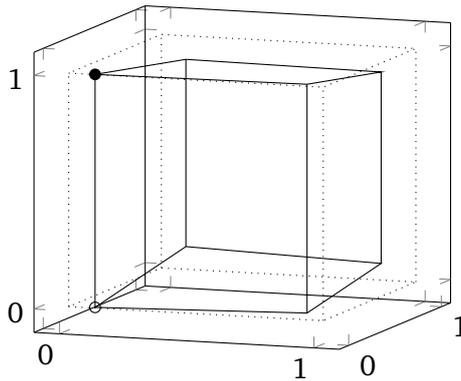
**Zadeh Regel** Wähle ein  $j \in \{\ell \in N_0 : \bar{z}_\ell < 0\}$ , das bislang am seltensten gewählt wurde.

---

## 5.4 Laufzeit des Simplex-Verfahrens

---

Die Endlichkeit des Simplex-Verfahrens ist also für alle Pivotregeln der vorigen Sektion, gegebenenfalls durch Kombination mit der lexikographischen Regel, gewährleistet. Es stellt sich



**Abbildung 5.4:** Der Klee-Minty Würfel in  $\mathbb{R}^3$  für  $\varepsilon = 10^{-1}$ .

die Frage, welchen Einfluss die Wahl der Pivotregel auf die Laufzeit des resultierenden Algorithmus hat. Insbesondere erhoffen wir uns eine effiziente, also polynomielle (siehe ADM) Worst-Case-Laufzeit. Das folgende Beispiel war das erste exponentielle Worst-Case-Beispiel für die ursprüngliche Fassung des Simplex-Algorithmus mit der Dantzig bzw. der Bland Regel. Die Idee hinter der Konstruktion ist es eine exponentielle Instanz für den Simplex-Algorithmus rekursiv aus zwei exponentiellen Instanzen kleinerer Dimension zu kombinieren.

**Definition 5.15.** Der *Klee-Minty-Würfel*  $\mathcal{K}(n) := \text{conv}(\mathcal{X}_n)$  von Dimension  $n \in \mathbb{N}$  ist rekursiv definiert über  $\mathcal{X}_1 := \{\varepsilon, 1 - \varepsilon\}$  für  $0 < \varepsilon \ll 1$  und

$$\mathcal{X}_{n \geq 2} := \left\{ \begin{pmatrix} x \\ \varepsilon x_{n-1} \end{pmatrix} : x \in \mathcal{X}_{n-1} \right\} \cup \left\{ \begin{pmatrix} x \\ 1 - \varepsilon x_{n-1} \end{pmatrix} : x \in \mathcal{X}_{n-1} \right\}.$$

**Beobachtung 5.16** (Übung). Es gilt  $\mathcal{K}(n) = \{x \in \mathbb{R}^n : \varepsilon x_{\ell-1} \leq x_\ell \leq 1 - \varepsilon x_{\ell-1} \forall \ell \in \{1, \dots, n\}\}$  mit  $x_0 := 1$  (siehe Abb. 5.4).

Der Klee-Minty-Würfel kann als eine Deformation des Hyperwürfels gesehen werden. Für die Bland Regel ist anschaulich klar, dass der Simplex-Algorithmus mit Startpunkt beim verschobenen Ursprung und Zielfunktion  $x_n$  zuerst die Variablen  $x_1, \dots, x_{n-1}$  optimiert, und dann von  $x_n \approx 0$  zu  $x_n \approx 1$  wechselt. Dort angekommen, optimiert er eine Spiegelung der Instanz auf den ersten  $n - 1$  Variablen, was im umgekehrten Verhalten verglichen mit der ersten Hälfte der Ausführung resultiert (siehe Abb. 5.4).

**Proposition 5.17.** Wird die eintretende Basisvariable nach der Bland Regel gewählt, besucht `SIMPLEX`, angefangen bei der Ecke  $x = (\varepsilon, \varepsilon^2, \varepsilon^3, \dots)$ , alle  $2^n$  Ecken der zulässigen Menge des LPs

$$\max\{x_n : x \in \mathcal{K}(n)\}.$$

*Beweis.* Wir beweisen die Aussage per Induktion über  $n$ , zusammen damit, dass die Optimallösung  $(\varepsilon, \varepsilon^2, \dots, \varepsilon^{n-1}, 1 - \varepsilon^n)$  ist. Beides ist klar für  $n = 1$ , also betrachten wir  $n \geq 2$ . In Standardform können die Nebenbedingungen des LPs (nach Beobachtung 5.16) als

$$\begin{aligned} x_\ell - x_{n+2\ell-1} - \varepsilon x_{\ell-1} &= 0, \\ x_\ell + x_{n+2\ell} + \varepsilon x_{\ell-1} &= 1, \end{aligned}$$

mit Schlupfvariablen  $x_{n+1}, \dots, x_{3n} \geq 0$  und  $x_0 \equiv 1$  geschrieben werden. Die eindeutige Basis zur Startecke ist dann  $B = \{1, \dots, n\} \cup \{n+2, n+4, \dots, 2n\}$ . Da die Bland Regel Variablen kleinerer Indizes präferiert, bleibt SIMPLEX solange wie möglich in Ecken  $\bar{x} \in \left\{ \binom{x}{\varepsilon x_{n-1}} : x \in \mathcal{X}_{n-1} \right\}$ . Für diese gilt  $\bar{x}_n = \varepsilon \bar{x}_{n-1}$ , also löst SIMPLEX zunächst das Teilproblem  $\arg \max \{x_{n-1} : x_{\{1, \dots, n-1\}} \in \mathcal{K}(n-1)\}$ . Per Induktion besucht SIMPLEX alle  $2^{n-1}$  Ecken, bis der Algorithmus die Optimallösung  $\bar{x}_{\{1, \dots, n-1\}} = (\varepsilon, \varepsilon^2, \dots, \varepsilon^{n-2}, 1 - \varepsilon^{n-1})$  des Teilproblems erreicht. Entsprechend ist dann  $\bar{x}_n = \varepsilon \bar{x}_{n-1} = \varepsilon - \varepsilon^n$ . Als nächstes wechselt SIMPLEX die Ecke, indem er  $\bar{x}_n = 1 - \varepsilon + \varepsilon^n$  setzt. Nun löst SIMPLEX das Teilproblem

$$\arg \max \{1 - \varepsilon x_{n-1} : x \in \text{conv}(\mathcal{X}_{n-1})\} = \arg \min \{x_{n-1} : x_{\{1, \dots, n-1\}} \in \mathcal{K}(n-1)\},$$

angefangen von der Ecke  $\bar{x}_{\{1, \dots, n-1\}}$ , die genau die Optimallösung des zuerst gelösten Teilproblems ist. Also sucht SIMPLEX die Ecken von  $\mathcal{K}(n-1)$  in umgekehrter Reihenfolge ab, wobei per Induktion wieder alle Ecken besucht werden. Insgesamt werden also alle  $2^n$  Ecken abgesucht. Die finale Lösung ist gegeben durch  $\bar{x}_{\{1, \dots, n-1\}} = (\varepsilon, \varepsilon^2, \dots, \varepsilon^{n-1})$  und damit  $\bar{x}_n = 1 - \varepsilon^n$ .  $\square$

Mittels ähnlicher Konstruktionen über Deformationen des Hyperwürfels kann man zeigen, dass die Dantzig Regel sowie die Regel der größten Verbesserung und des steilsten Anstiegs jeweils zu einer exponentiellen Laufzeit führen. Für die anderen Pivotregeln der vorigen Sektion sind ebenfalls (erheblich aufwändigere) exponentielle Laufzeitschranken bekannt, die teilweise erst kürzlich gezeigt werden konnten. Die einzige Ausnahme ist die Regel der zufälligen Facette, für die die erwartete Laufzeit mit  $\Theta(2^{\sqrt{n}})$  (für  $n > m$ ) aber immer noch „subexponentiell“ ist. Die Frage nach der Existenz einer polynomiellen Pivotregel wird oft als die *wichtigste offene Frage der diskreten Optimierung* bezeichnet.

**Problem 5.18.** Existiert eine Pivotregel für die SIMPLEX nur  $n^{\mathcal{O}(1)}$  Iterationen braucht?

Insbesondere würde eine polynomielle Pivotregel einen polynomiellen „kombinatorischen“ (nicht-numerischen) Algorithmus zur Lösung von LPs liefern. Ein solcher Algorithmus wäre sehr wahrscheinlich *stark polynomiell*, hätte also eine Laufzeit, die durch ein Polynom in  $n \cdot m$  beschränkt werden kann, unabhängig von der Darstellungsgröße der Zahlen im Input. Das wäre die Lösung eines der 18 *Probleme für das 21ste Jahrhundert*, die von Stephen Smale 1998 als Aktualisierung der Hilbert-Probleme (1900) vorgeschlagen wurden.

**Problem 5.19** (Smale-Problem #9). Können LPs in streng polynomieller Zeit  $(nm)^{\mathcal{O}(1)}$  gelöst werden?

Weiterhin würde die Existenz einer polynomiellen Pivotregel sofort einen Beweis für die folgende zentrale Vermutung der Polyedertheorie liefern.

**Vermutung 5.20** (polynomielle Hirsch-Vermutung). *Paare von Ecken eines Polyeders  $\mathcal{P}$  sind durch einen Pfad aus  $\dim(\mathcal{P})^{\mathcal{O}(1)}$  Kanten verbunden.*

---

## 5.5 Finden einer Startbasis

---

Bislang haben wir angenommen, dass eine primal zulässige Basis von (P) mit zugehöriger Basislösung vorliegt. Zur vollständigen Beschreibung des Simplex-Verfahrens müssen wir noch erklären, wie man an die anfängliche Basis kommt. Dazu betrachten wir das folgende LP in Standardform:

$$(PI) \quad \begin{array}{ll} \min & \mathbf{1}^\top \mathbf{y} \\ \text{s.t.} & \mathbf{A}\mathbf{x} + \mathbf{y} = \mathbf{b} \\ & \mathbf{x}, \mathbf{y} \geq \mathbf{0}. \end{array}$$

Im Vergleich zu (P) haben wir *künstliche Variablen*  $\mathbf{y} \in \mathbb{R}^m$  eingeführt und die Zielfunktion ersetzt. Der Lösungsraum von (PI) ist  $\mathcal{P}^=(D, \mathbf{b})$ , wobei  $D = \begin{pmatrix} A & \mathbb{1}_{m \times m} \end{pmatrix}$  vollen Zeilenrang  $\text{rank}(D) = m$  hat. Wir nehmen ohne Beschränkung der Allgemeinheit an, dass  $\mathbf{b} \geq \mathbf{0}$ , was wir durch Negierung von Nebenbedingungen von (P) erreichen können.

**Proposition 5.21.** Ist  $\mathbf{b} \geq \mathbf{0}$ , so gilt für (PI):

- (i)  $B = \{n+1, \dots, n+m\}$  ist eine primal zulässige Basis mit Basislösung  $\bar{\mathbf{u}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{b} \end{pmatrix}$ .
- (ii) Es existiert eine optimale Basislösung  $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix}$ .
- (iii) Ist  $\mathbf{y}^* \neq \mathbf{0}$ , dann ist (P) unzulässig. Ist  $\mathbf{y}^* = \mathbf{0}$  und  $\text{rank}(A) = m$ , dann ist  $\mathbf{x}^*$  primal zulässige Basislösung von (P).

*Beweis.*

(i): Aus  $D_{\cdot B} = \mathbb{1} = D_{\cdot \bar{B}}^{-1}$  und  $\mathbf{b} \geq \mathbf{0}$  folgt  $D_{\cdot B}$  regulär mit  $\bar{\mathbf{u}}_B = D_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}$  und  $\bar{\mathbf{u}}_N = \mathbf{0}$ , also ist  $B$  primal zulässige Basis mit zugehöriger Basislösung  $\bar{\mathbf{u}}$ .

(ii): Aus  $\mathbf{y} \geq \mathbf{0}$  folgt, dass die Zielfunktion beschränkt ist durch  $\mathbf{1}^\top \mathbf{y} \geq 0$ . Aus dem ersten Teil wissen wir, dass eine zulässige Lösung existiert. Nach Satz 4.17 folgt, dass eine endliche Optimallösung existiert. Aus der Korrektheit (Satz 5.10) und Endlichkeit (Satz 5.13 bzw. 5.14) des Simplex-Verfahrens mit geeigneter Pivotregel, sowie der Existenz einer zulässigen Startbasis folgt, dass eine optimale Basislösung existiert.

(iii): Ist  $\mathbf{y}^* \neq \mathbf{0}$ , dann ist, wegen  $\mathbf{y} \geq \mathbf{0}$ , der Optimalwert von (PI)  $\mathbf{1}^\top \mathbf{y}^* > 0$ . Damit kann (P) keine zulässige Lösung  $\mathbf{x} \in \mathcal{P}^=(A, \mathbf{b})$  haben, denn sonst wäre  $\begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix} \in \mathcal{P}^=(D, \mathbf{b})$  zulässig für (PI) mit kleinerem Zielfunktionswert als die Optimallösung  $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix}$ .

Ist  $\mathbf{y}^* = \mathbf{0}$ , dann ist  $\mathbf{A}\mathbf{x}^* = \mathbf{A}\mathbf{x}^* + \mathbf{y}^* = \mathbf{b}$ , also ist  $\mathbf{x}^* \in \mathcal{P}^=(A, \mathbf{b})$  zulässig für (P). Sei  $\bar{B}$  eine zugehörige Basis zur Basislösung  $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{0} \end{pmatrix}$  von (PI). Setze

$$\bar{B}_x = \{1, \dots, n\} \cap \bar{B}, \quad \bar{N}_x = \{1, \dots, n\} \setminus \bar{B}.$$

Da  $\bar{B}$  eine Basis für (PI) ist, gilt  $\mathbf{x}_{\bar{B}_x}^* \geq \mathbf{0}$ ,  $\mathbf{x}_{\bar{N}_x}^* = \mathbf{0}$  und  $A_{\cdot \bar{B}_x}$  hat linear unabhängige Spalten. Ist nun  $|\bar{B}_x| = m$  (bzw.  $\bar{B} \setminus \{1, \dots, n\} = \emptyset$ ), dann ist  $\bar{B}_x$  bereits eine primal zulässige Basis für (P). Andernfalls können wir, da  $\text{rank}(A) = m$  die Matrix  $A_{\cdot \bar{B}_x}$  durch  $m - |\bar{B}_x|$  linear unabhängige Spalten aus  $A_{\cdot \bar{N}_x}$  zu einer regulären  $(m \times m)$ -Matrix ergänzen. Für die entsprechenden Spaltenindizes  $B \supset \bar{B}_x$  gilt wegen  $\mathbf{x}_{\bar{N}_x}^* = \mathbf{0}$ , dass

$$A_{\cdot B} \mathbf{x}_B^* = A_{\cdot \bar{B}_x} \mathbf{x}_{\bar{B}_x}^* = \mathbf{b}$$

und somit  $A_{\cdot B}^{-1} \mathbf{b} = \mathbf{x}_B^* \geq \mathbf{0}$ . Damit ist  $B$  eine zulässige Basis von (P) mit zugehöriger Basislösung  $\mathbf{x}^*$ .  $\square$

Proposition 5.21 gibt eine Anleitung, wie wir eine anfängliche primal zulässige Basislösung von (P) zum Starten von SIMPLEX bestimmen können, indem wir zunächst SIMPLEX für (PI) ausführen. Sei  $\bar{B}$  die dabei bestimmte Basis von (PI). Analog zum Beweis von Proposition 5.21 müssen wir  $\{1, \dots, n\} \cap \bar{B}$  zu einer Basis von (P) ergänzen. Wir brauchen also zusätzliche Spalten aus  $A_{\bar{N}_x}$ , die linear unabhängig zu den Spalten von  $A_{\bar{B}_x}$  sind. Eine Spalte  $A_j$  mit  $j \in \bar{N}_x$  ist genau dann linear abhängig mit den Spalten von  $A_{\bar{B}_x}$ , wenn für die eindeutige Lösung  $w_{\bar{B}}$  von  $(A \ \mathbb{1})_{\cdot, \bar{B}} w_{\bar{B}} = A_j$  gilt, dass  $w_{\bar{B} \setminus \bar{B}_x} = \mathbf{0}$ . Andernfalls existiert  $i \in \bar{B} \setminus \bar{B}_x$  mit  $w_i \neq 0$  und es ist

$$(A \ \mathbb{1})_{\cdot, i} = \frac{1}{w_i} [A_{\cdot, j} - (A \ \mathbb{1})_{\cdot, \bar{B} \setminus \{i\}} w_{\bar{B} \setminus \{i\}}],$$

also können wir  $i$  in  $\bar{B}$  durch  $j$  ersetzen und die Invariante erhalten, dass  $\bar{B}$  eine Basis von (PI) ist. Ist  $\text{rank}(A) = m$ , können wir Spalten  $A_j$  für die Basisergänzung von  $\bar{B}_x$  finden, bis schließlich  $\bar{B} \subseteq \{1, \dots, n\}$  und wir somit eine zulässige Basis von (P) haben. Dieses Verfahren zur Bestimmung einer zulässigen Startbasis für SIMPLEX wird als Phase I des Simplex-Verfahrens bezeichnet.

---

**Algorithmus :** PHASEI( $A, b$ )

---

**input:**  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}_{\geq 0}^m$

**output:** primal zulässige Basis  $B$  für (P) mit Basislösung  $\bar{x}$ ,  
 $\infty$  falls (P) unzulässig,  $\emptyset$  falls  $\text{rank}(A) < m$

---

$((\frac{\bar{x}}{y}), \cdot, \bar{B}) \leftarrow \text{SIMPLEX}((A \ \mathbb{1}), b, (\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}), \{n+1, \dots, n+m\}, (\begin{smallmatrix} 0 \\ b \end{smallmatrix}))$

**if**  $\bar{y} \neq \mathbf{0}$ :

$\perp$  **return**  $\infty$

**while**  $\bar{B} \not\subseteq \{1, \dots, n\}$ :

$\bar{D} \leftarrow (A \ \mathbb{1})_{\cdot, \bar{B}}^{-1}$

**if**  $\exists j \in \{1, \dots, n\} \setminus \bar{B}, i \in \bar{B} \setminus \{1, \dots, n\} : (\bar{D}A_{\cdot, j})_i \neq 0$ :

$\bar{B} \leftarrow \bar{B} \cup \{j\} \setminus \{i\}$

**else**

$\perp$  **return**  $\emptyset$

**return**  $(\bar{B}, \bar{x})$

---

**Satz 5.22.** PHASEI ist korrekt.

*Beweis.* Korrektheit folgt unmittelbar aus Proposition 5.21 zusammen mit unserer Herleitung der Basisergänzung.  $\square$

---

## 5.6 Das duale Simplex-Verfahren

---

Das duale Programm (D) kann durch Einführung einer Schlupfvariablen  $z \in \mathbb{R}^n$  (Beobachtung 3.6) in (negierte) Standardform gebracht werden:

$$(DS) \quad \begin{aligned} \max \quad & b^\top y^+ - b^\top y^- \\ \text{s.t.} \quad & A^\top y^+ - A^\top y^- + z = c, \\ & y^+, y^-, z \geq 0. \end{aligned}$$

Nach Proposition 5.7 und wegen starker Dualität, können wir eine Optimallösung von (P) bestimmen, indem wir SIMPLEX auf (DS) anwenden. Da  $\mathbf{y}^+$  und  $\mathbf{y}^-$  genau komplementär auftreten, kann, bis auf Zwischenschritte, immer jeweils entweder  $y_i^+$  oder  $y_i^-$  in der Basis gehalten werden.

**Beobachtung 5.23** (Übung). Während der Ausführung von SIMPLEX für (DS) können nie sowohl  $y_i^+$  als auch  $y_i^-$  für  $i \in \{1, \dots, m\}$  Basisvariable sein. Ist keine von beiden Basisvariable, so kann eine der beiden Variablen als eintretende Variable gewählt werden.

Diese Beobachtung erlaubt es uns bei der Ausführung von SIMPLEX für (DS) effektiv nur die Variablen  $\mathbf{z}$  als eintretende bzw. austretende Basisvariablen zu betrachten. Dazu vereinfachen wir (DS) zu dem folgenden LP mit *freien Variablen*  $\mathbf{y}$ :

$$(DS') \quad \begin{aligned} \max \quad & \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} \quad & A^\top \mathbf{y} + \mathbf{z} = \mathbf{c}, \\ & \mathbf{z} \geq \mathbf{0}. \end{aligned}$$

Wir passen die Definition einer Basis nun so an, dass alle freien Variablen in der Basis gehalten werden.

**Definition 5.24.** Habe  $A \in \mathbb{R}^{m \times n}$  vollen Zeilenrang. Eine Menge  $H \subseteq \{1, \dots, m+n\}$  heißt *Basis* von (DS') wenn  $\{1, \dots, m\} \subseteq H$ ,  $|H| = n$  und  $(A^\top \quad \mathbb{1})_{.H}$  regulär ist. Die zugehörige *Basislösung*  $\bar{\mathbf{u}} \in \mathbb{R}^{m+n}$  ist gegeben durch  $\bar{\mathbf{u}}_H = (A^\top \quad \mathbb{1})_{.H}^{-1} \mathbf{c}$  und  $\bar{\mathbf{u}}_{M_0} = \mathbf{0}$ , wobei  $M_0 := \{1, \dots, m+n\} \setminus H$  die *Nichtbasis* bezeichnet. Die Basis  $H$  ist *zulässig*, wenn  $\bar{\mathbf{u}}_{\{m+1, \dots, m+n\}} \geq \mathbf{0}$ .

**Proposition 5.25.** Ist  $H$  Basis von (DS'), so ist  $B = \{i-m : i \in \{1, \dots, m+n\} \setminus H\}$  Basis von (P) mit Nichtbasis  $N_0 = \{i-m : i \in \{m+1, \dots, m+n\} \cap H\}$ . Dann ist  $H$  zulässig für (DS') genau dann, wenn  $B$  dual zulässig für (P) ist.

*Beweis.* Ist  $H$  Basis von (DS'), dann ist  $B \subseteq \{1, \dots, n\}$  mit  $|B| = (m+n) - |H| = m$ . Außerdem ist dann  $(A^\top \quad \mathbb{1})_{.H}$  und damit  $(\mathbb{1}_{N_0}^A)$  regulär, also hat  $(\mathbb{1}_{N_0}^A)_{.B} = \begin{pmatrix} A_{.B} \\ \mathbf{0} \end{pmatrix}$  vollen Spaltenrang. Wegen  $\text{rank}(A) = m$  und  $|B| = m$  ist  $A_{.B}$  regulär. Also ist  $B$  Basis von (P).

Sei  $\bar{\mathbf{u}} = \begin{pmatrix} \bar{\mathbf{y}} \\ \bar{\mathbf{z}} \end{pmatrix} \in \mathbb{R}^{m+n}$  die zu  $H$  gehörige Basislösung. Wegen  $\bar{\mathbf{u}}_{M_0} = \mathbf{0}$  gilt  $\bar{\mathbf{z}}_B = \mathbf{0}$  und damit

$$\mathbf{c}_B = \left[ (A^\top \quad \mathbb{1})_{.H} \bar{\mathbf{u}}_H \right]_B = A_{.B}^\top \bar{\mathbf{y}} + [\mathbb{1}_{N_0} \bar{\mathbf{z}}_{N_0}]_B = A_{.B}^\top \bar{\mathbf{y}}.$$

Analog ist

$$\mathbf{c}_{N_0} = A_{.N_0}^\top \bar{\mathbf{y}} + \mathbb{1}_{N_0} \bar{\mathbf{z}}_{N_0} = A_{.N_0}^\top A_{.B}^{-\top} \mathbf{c}_B + \bar{\mathbf{z}}_{N_0}. \quad (5.4)$$

Per Definition und wegen  $\bar{\mathbf{u}}_{M_0} = \mathbf{0}$  ist  $H$  zulässig genau dann, wenn  $\mathbf{z}_{N_0} \geq \mathbf{0}$ . Nach (5.4) ist das äquivalent dazu, dass  $B$  dual zulässig ist.  $\square$

Somit ist jede zulässige Basis von (DS') eindeutig durch eine dual zulässige Basis von (P) bestimmt. Die Idee des dualen Simplex-Verfahrens ist es nun direkt mit dual zulässigen Basen von (P) zu arbeiten. Wir formen nun die einzelnen Schritte von SIMPLEX entsprechend um. Dabei ist zu beachten, dass unsere Einschränkung auf zulässige Lösungen von (DS') eine Einschränkung der Pivotregel mit sich bringt.

Im ersten Schritt bestimmt SIMPLEX die Lösung von  $\begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}_H = (A^\top \quad \mathbb{1})_{.H}^\top \bar{\mathbf{x}}$ , also  $\bar{\mathbf{x}}$  mit

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} A \\ \mathbb{1}_{N_0} \end{pmatrix} \bar{\mathbf{x}} = \begin{pmatrix} A_{.B} \bar{\mathbf{x}}_B + A_{.N_0} \bar{\mathbf{x}}_{N_0} \\ \bar{\mathbf{x}}_{N_0} \end{pmatrix}.$$

Effektiv bestimmt SIMPLEX also die Basislösung  $\bar{\mathbf{x}}$  von  $B$  mit  $\bar{\mathbf{x}}_{N_0} = \mathbf{0}$  und  $\bar{\mathbf{x}}_B = A_{.B}^{-1} \mathbf{b}$ .

Als nächstes bestimmt SIMPLEX die reduzierten Kosten. Die Nichtbasisvariablen sind  $\mathbf{u}_{M_0} = \mathbf{z}_B$ , also sind die reduzierten Kosten gegeben durch

$$\bar{\mathbf{r}}_{M_0} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}_{M_0} - (A^\top \quad \mathbb{1})_{.M_0}^\top \bar{\mathbf{x}} = \mathbf{0} - \mathbb{1}_{.B} \bar{\mathbf{x}} = -\bar{\mathbf{x}}_B.$$

Wir betrachten ein Maximierungsproblem, also terminiert der Algorithmus wenn  $\bar{\mathbf{r}}_{M_0} \leq \mathbf{0}$ , beziehungsweise wenn  $\bar{\mathbf{x}}_B \geq \mathbf{0}$ . Das ist gerechtfertigt, denn dann ist  $B$  auch primal zulässig. Nach Proposition 5.7 haben wir dann eine Optimallösung  $\bar{\mathbf{x}}$  von (P) gefunden. Sonst wählt der Algorithmus eine eintretende Basisvariable  $u_{j'}$  mit  $j' \in M_0$  und  $r_{j'} > 0$ , also eine Variable  $z_i$  mit  $i \in B$  und  $\bar{x}_i < 0$ . Wir wählen hier also eine Variable  $z_i$ , die aus  $B$  austritt.

In der Forward Transformation bestimmt SIMPLEX die Lösung von  $(A^\top \quad \mathbb{1})_{.H} (\mathbf{a}_{N_0}^w) = (A^\top \quad \mathbb{1})_{.j'}$ , also  $\mathbf{w}$  und  $\mathbf{a}_{N_0}$  mit  $A^\top \mathbf{w} + \mathbb{1}_{.N_0} \mathbf{a}_{N_0} = \mathbf{e}_i$ , wobei  $\mathbf{e}_i \in \mathbb{R}^n$  den  $i$ -ten Einheitsvektor bezeichnet. Teilen wir dieses System in Zeilen aus  $B$  und aus  $N_0$ , erhalten wir

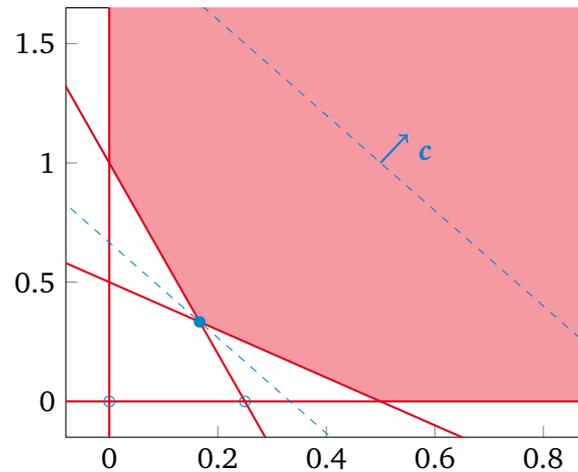
$$\begin{aligned} A_{.B}^\top \mathbf{w} &= (\mathbf{e}_i)_B, \\ A_{.N_0}^\top \mathbf{w} + \mathbf{a}_{N_0} &= \mathbf{0}. \end{aligned}$$

Also können wir  $\mathbf{w} = (A_{.B}^{-\top})_{.i}$  und  $\mathbf{a}_{N_0} = -A_{.N_0}^\top \mathbf{w}$  setzen.

Nach Beobachtung 5.23 müssen wir nur die Variablen  $\mathbf{z}$  als austretende Variablen in Betracht ziehen. Für den Ratio Test prüfen wir also lediglich ob  $(\mathbf{a}_{N_0}^w)_{H \setminus \{1, \dots, m\}} = \mathbf{a}_{N_0} \leq \mathbf{0}$ . Ist dies der Fall, dann ist (D) unbeschränkt und damit (P) unzulässig (nach Satz 4.17). Sonst wählen wir

$$j \in \arg \min \left\{ \frac{\bar{z}_\ell}{\alpha_\ell} : \ell \in N_0, \alpha_\ell > 0 \right\}$$

als austretende Variable für  $H$ , beziehungsweise als eintretende Variable für  $B$ .



**Abbildung 5.5:** Grafische Darstellung des LPs (LP) aus Beispiel 5.26 mit Zwischenlösungen von DUALSIMPLEX.

---

**Algorithmus :** DUALSIMPLEX( $A, \mathbf{b}, \mathbf{c}, B, N, \bar{\mathbf{z}}_N, \bar{\mathbf{y}}$ )

---

**input:**  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $\mathbf{c} \in \mathbb{R}^n$ , dual zulässige Basis  $B$ ,  $N_0 = \{1, \dots, n\} \setminus B$   
 Basislösung  $\bar{\mathbf{y}} = A_{\cdot B}^{-\top} \mathbf{c}_B$  und  $\bar{\mathbf{z}}_{N_0} = \mathbf{c}_{N_0} - A_{\cdot N_0}^{\top} \bar{\mathbf{y}}$

**output:** Optimallösungen  $\bar{\mathbf{x}}$  für (P),  $\bar{\mathbf{y}}$  für (D) und  $\bar{\mathbf{y}}, \bar{\mathbf{z}}$  für (DS'),  
 oder  $\infty$  falls (D) unbeschränkt bzw. (P) unzulässig

---

**repeat**

/\* Backward Transformation \*/

$\bar{\mathbf{x}}_B \leftarrow A_B^{-1} \mathbf{b}$  (Lösung von  $A_B \bar{\mathbf{x}}_B = \mathbf{b}$ )

/\* Pricing \*/

**if**  $\bar{\mathbf{x}}_B \geq \mathbf{0}$ :

$\bar{\mathbf{x}}_{N_0} \leftarrow \mathbf{0}, \bar{\mathbf{z}}_B \leftarrow \mathbf{0}$

**return**  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}, B)$

$i \in \{\ell \in B : \bar{x}_\ell < 0\}$  (Pivotschritt)

/\* Forward Transformation \*/

$\mathbf{w} \leftarrow (A_{\cdot B}^{-\top})_{\cdot i}, \alpha_{N_0} \leftarrow -A_{\cdot N_0}^{\top} \mathbf{w}$

/\* Ratio Test \*/

**if**  $\alpha_{N_0} \leq \mathbf{0}$ :

**return**  $\infty$

$j \in \arg \min \{ \frac{\bar{z}_\ell}{\alpha_\ell} : \ell \in N_0, \alpha_\ell > 0 \}$

$\gamma \leftarrow \frac{\bar{z}_j}{\alpha_j}$  (Schrittweite)

/\* Update \*/

$\bar{\mathbf{y}} \leftarrow \bar{\mathbf{y}} - \gamma \mathbf{w}, \bar{\mathbf{z}}_{N_0} \leftarrow \bar{\mathbf{z}}_{N_0} - \gamma \alpha_{N_0}$

$B \leftarrow B \cup \{j\} \setminus \{i\}, N_0 \leftarrow N_0 \cup \{i\} \setminus \{j\}$

$\bar{z}_i \leftarrow \gamma$

---

**Beispiel 5.26.** Wir betrachten das folgende lineare Programm (siehe Abb. 5.5):

$$\begin{aligned}
 \text{(LP)} \quad & \min \quad 2x_1 + x_2 \\
 & \text{s.t.} \quad -x_1 - x_2 \leq -\frac{1}{2}, \\
 & \quad \quad -4x_1 - x_2 \leq -1, \\
 & \quad \quad \mathbf{x} \geq \mathbf{0}.
 \end{aligned}$$

Durch Einführen von Schlupfvariablen (nach Bemerkung (3.6)) können wir das LP in Standardform überführen und erhalten

$$\begin{aligned}
 \text{(LP')} \quad & \min \quad (2 \ 1 \ 0 \ 0) \mathbf{x} \\
 & \text{s.t.} \quad \begin{pmatrix} -1 & -1 & 1 & 0 \\ -4 & -1 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} -\frac{1}{2} \\ -1 \end{pmatrix}, \\
 & \quad \quad \mathbf{x} \geq \mathbf{0}.
 \end{aligned}$$

Eine zulässige Basis ist  $B = \{3, 4\}$ ,  $N_0 = \{1, 2\}$  mit Basismatrix  $A_{.B} = \mathbb{I}$  und zugehöriger dualer Basislösung gegeben durch  $\bar{\mathbf{y}} = A_{.B}^{-\top} \mathbf{c}_B = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$  und

$$\bar{\mathbf{z}}_{N_0} = \mathbf{c}_{N_0} - A_{.N_0}^{-\top} \bar{\mathbf{y}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} -1 & -4 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \geq \mathbf{0},$$

insbesondere ist  $B$  dual zulässig. Wir fassen die Schritte des Simplexalgorithmus tabellarisch zusammen.

step	$\bar{\mathbf{x}}_B$	$i$	$\mathbf{w}$	$\mathbf{a}_{N_0}$	$j$	$\gamma$	$B$	$\bar{\mathbf{y}}$	$\bar{\mathbf{z}}_{N_0}$	$A_{.B}$	$A_{.B}^{-1}$
							$\{3, 4\}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
1	$\begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$	4	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 1 \end{pmatrix}$	1	$\frac{1}{2}$	$\{1, 3\}$	$\begin{pmatrix} 0 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ -4 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & -1/4 \\ 1 & -1/4 \end{pmatrix}$
2	$\begin{pmatrix} 1/4 \\ -1/4 \end{pmatrix}$	3	$\begin{pmatrix} -1 \\ -1/4 \end{pmatrix}$	$\begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix}$	2	$\frac{2}{3}$	$\{1, 2\}$	$\begin{pmatrix} -2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ -4 & -1 \end{pmatrix}$	$\begin{pmatrix} 1/3 & -1/3 \\ -4/3 & 1/3 \end{pmatrix}$
3	$\begin{pmatrix} 1/6 \\ 1/3 \end{pmatrix}$										

In Iteration 3 terminiert der Algorithmus im Pricing-Schritt und gibt  $\bar{\mathbf{x}} = (\frac{1}{6}, \frac{1}{3}, 0, 0)^\top$  und  $B = \{1, 2\}$  zurück, was der Optimallösung  $\frac{1}{6} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$  von (LP) mit Wert  $\frac{2}{3}$  entspricht.  $\triangle$

## 5.7 Sensitivitätsanalyse

Ein wichtiger Vorteil des Simplex-Verfahrens liegt darin, dass wir nachträglich auf gewisse Änderungen von (P) reagieren können, indem wir nachoptimieren, anstatt das veränderte Programm von Grund auf neu zu lösen. Nehmen wir dazu an, wir haben mit Hilfe des Simplex-Verfahrens eine optimale Basis  $B$  und die zugehörige optimale Basislösung  $\bar{\mathbf{x}}$  von (P) bestimmt. Wir behandeln im Folgenden Änderungen aller Parameter von (P).

---

## Änderung der Zielfunktion und der Nichtbasisspalten

---

Ändert sich die Zielfunktion von  $\mathbf{c}$  zu  $\tilde{\mathbf{c}}$  und die Nichtbasismatrix  $A_{N_0}$  zu  $\tilde{A}_{N_0}$ , dann bleibt  $B$  primal zulässig, da  $\bar{\mathbf{x}}_B = A_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}$  unverändert gilt. Also kann SIMPLEX für (P) mit Anfangsbasis  $B$  und Basislösung  $\bar{\mathbf{x}}_B$  gestartet werden. Gilt für die reduzierten Kosten

$$\tilde{\mathbf{z}}_{N_0} = \tilde{\mathbf{c}}_{N_0} - \tilde{A}_{N_0}^T A_{\cdot B}^{-T} \tilde{\mathbf{c}}_B \geq \mathbf{0},$$

so bleibt  $\bar{\mathbf{x}}$  insbesondere Optimallösung – mit verändertem Zielfunktionswert, falls  $\tilde{\mathbf{c}}_B^T \bar{\mathbf{x}}_B \neq \mathbf{c}_B^T \bar{\mathbf{x}}_B$ . Wir zeigen noch, dass sich der Optimalwert stetig mit der Zielfunktion ändert.

**Beobachtung 5.27.** Es ist

$$p(\tilde{\mathbf{c}}) := \inf\{\tilde{\mathbf{c}}^T \mathbf{x} : \mathbf{x} \in \mathcal{P}^=(A, \mathbf{b})\}$$

im Inneren von  $\mathcal{C} := \{\tilde{\mathbf{c}} \in \mathbb{R}^n : p(\tilde{\mathbf{c}}) > -\infty\}$  stetig.

*Beweis.* Wir behaupten,  $\mathcal{C}$  ist konvex. Nach dem starken Dualitätssatz (Satz 4.17) ist  $\mathcal{C} = \{\tilde{\mathbf{c}} \in \mathbb{R}^n : \mathcal{P}(A^T, \tilde{\mathbf{c}}) \neq \emptyset\}$ . Angenommen es existieren  $\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2 \in \mathcal{C}$  und  $\tilde{\mathbf{c}} = \lambda \tilde{\mathbf{c}}_1 + (1 - \lambda) \tilde{\mathbf{c}}_2 \notin \mathcal{C}$  mit  $\lambda \in (0, 1)$ . Nach dem Farkas-Lemma (Korollar 4.14) existiert  $\mathbf{y} \geq \mathbf{0}$  mit  $A\mathbf{y} = \mathbf{0}$  und  $\tilde{\mathbf{c}}^T \mathbf{y} < 0$ . Sei  $\bar{\mathbf{x}} \in \mathcal{P}^=(A, \mathbf{b})$  und  $(\bar{\mathbf{x}}_i)_{i \in \mathbb{N}}$  mit  $\bar{\mathbf{x}}_i = \bar{\mathbf{x}} + i\mathbf{y}$ . Wegen  $A\mathbf{y} = \mathbf{0}$  und  $\mathbf{y} \geq \mathbf{0}$  ist  $\bar{\mathbf{x}}_i \in \mathcal{P}^=(A, \mathbf{b})$  und damit

$$-\infty = \lim_{i \rightarrow \infty} \tilde{\mathbf{c}}^T \bar{\mathbf{x}}_i = \lim_{i \rightarrow \infty} \lambda \tilde{\mathbf{c}}_1^T \bar{\mathbf{x}}_i + \lim_{i \rightarrow \infty} (1 - \lambda) \tilde{\mathbf{c}}_2^T \bar{\mathbf{x}}_i \geq \lambda p(\tilde{\mathbf{c}}_1) + (1 - \lambda) p(\tilde{\mathbf{c}}_2).$$

Das ist ein Widerspruch zu  $\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2 \in \mathcal{C}$ . Also ist  $\mathcal{C}$  konvex.

Über  $\mathcal{C}$  ist die Funktion  $p(\tilde{\mathbf{c}})$  das punktweise Infimum der linearen Funktionen  $g_{\bar{\mathbf{x}}}(\tilde{\mathbf{c}}) = \tilde{\mathbf{c}}^T \bar{\mathbf{x}}$  mit  $\bar{\mathbf{x}} \in \mathcal{P}^=(A, \mathbf{b})$ . Analog zu Proposition 2.32 folgt, dass  $-p(\tilde{\mathbf{c}}) = \sup_{\bar{\mathbf{x}}} (-g_{\bar{\mathbf{x}}}(\tilde{\mathbf{c}}))$  über  $\mathcal{C}$  konvex ist. Nach Satz 2.35 ist also  $-p(\tilde{\mathbf{c}})$  und damit  $p(\tilde{\mathbf{c}})$  stetig auf dem Inneren der konvexen Menge  $\mathcal{C}$ .  $\square$

---

## Änderung der rechten Seite

---

Ändert sich die rechte Seite von  $\mathbf{b}$  zu  $\tilde{\mathbf{b}}$ , dann bleibt  $B$  dual zulässig, da  $\bar{\mathbf{z}}_{N_0} = \mathbf{c}_{N_0} - A_{N_0}^T A_{\cdot B}^{-T} \mathbf{c}_B \geq \mathbf{0}$  unverändert gilt. Also kann DUALSIMPLEX für (P) mit Anfangsbasis  $B$  und dualer Basislösung  $\bar{\mathbf{y}} = A_{\cdot B}^{-T} \mathbf{c}_B$  und  $\bar{\mathbf{z}}_{N_0} = \mathbf{c}_{N_0} - A_{N_0}^T \bar{\mathbf{y}}$  gestartet werden. Gilt für die primale Basislösung

$$\tilde{\mathbf{x}}_B = A_{\cdot B}^{-1} \tilde{\mathbf{b}} \geq \mathbf{0},$$

so bleibt  $B$  insbesondere optimal – gegebenenfalls mit veränderter primaler Basislösung und verändertem Zielfunktionswert. Wir zeigen noch, dass sich der Optimalwert stetig mit der rechten Seite ändert.

**Beobachtung 5.28.** Es ist

$$\nu(\tilde{\mathbf{b}}) := \inf\{\mathbf{c}^T \mathbf{x} : \mathbf{x} \in \mathcal{P}^=(A, \tilde{\mathbf{b}})\}$$

im Inneren von  $\mathcal{B} := \{\tilde{\mathbf{b}} \in \mathbb{R}^m : \nu(\tilde{\mathbf{b}}) \in \mathbb{R}\}$  stetig.

*Beweis.* Nach dem starken Dualitätssatz (Satz 4.17) ist

$$\mathcal{B} = \{\bar{\mathbf{b}} \in \mathbb{R}^m : \mathcal{P}(A^\top, \mathbf{c}) \neq \emptyset\} \cap \{\bar{\mathbf{b}} \in \mathbb{R}^m : \mathcal{P}^\circ(A, \bar{\mathbf{b}}) \neq \emptyset\}.$$

Die erste Menge ist  $\emptyset$  oder  $\mathbb{R}^m$ , also konvex, und die zweite Menge ist konvex wie im Beweis von Beobachtung 5.27. Also ist  $\mathcal{B}$  konvex (Proposition 2.5).

Sei

$$d(\tilde{\mathbf{b}}) := \sup\{\tilde{\mathbf{b}}^\top \mathbf{y} : \mathbf{y} \in \mathcal{P}(A^\top, \mathbf{c})\}$$

die Optimalwertfunktion von (D). Nach dem starken Dualitätssatz (Satz 4.17) ist  $\nu(\tilde{\mathbf{b}}) = d(\tilde{\mathbf{b}})$ . Über  $\mathcal{B}$  ist die Funktion  $d(\tilde{\mathbf{b}})$  das Punktweise Supremum der linearen Funktionen  $h_{\tilde{\mathbf{y}}}(\tilde{\mathbf{b}}) = \tilde{\mathbf{b}}^\top \tilde{\mathbf{y}}$  mit  $\tilde{\mathbf{y}} \in \mathcal{P}(A^\top, \mathbf{c})$ . Aus Proposition 2.32 folgt, dass  $d(\tilde{\mathbf{b}})$  über  $\mathcal{B}$  konvex ist. Nach Satz 2.35 ist also  $\nu(\tilde{\mathbf{b}}) = d(\tilde{\mathbf{b}})$  stetig auf dem Inneren der konvexen Menge  $\mathcal{B}$ .  $\square$

---

### Hinzufügen einer neuen Variablen

---

Fügen wir eine neue Variable  $x_{n+1} \geq 0$  hinzu, dann ändert sich  $A$  zu  $\tilde{A} = \begin{pmatrix} A & \mathbf{a} \end{pmatrix}$  mit  $\mathbf{a} \in \mathbb{R}^m$  und  $\mathbf{c}$  zu  $\tilde{\mathbf{c}} = \begin{pmatrix} \mathbf{c} \\ \gamma \end{pmatrix}$  mit  $\gamma \in \mathbb{R}$ . Dabei bleibt  $B$  eine primal zulässige Basis, da  $\tilde{\mathbf{x}}_B = \tilde{A}_{\cdot B}^{-1} \tilde{\mathbf{b}} = A_{\cdot B}^{-1} \mathbf{b} \geq \mathbf{0}$  unverändert gilt. Also kann SIMPLEX mit Anfangsbasis  $B$  und Basislösung  $\begin{pmatrix} \tilde{\mathbf{x}} \\ 0 \end{pmatrix}$  gestartet werden. Gilt für die reduzierten Kosten von  $x_{n+1}$ , dass

$$\tilde{z}_{n+1} = \gamma - \mathbf{a}^\top A_{\cdot B}^{-\top} \mathbf{c}_B \geq 0,$$

so ist  $\begin{pmatrix} \tilde{\mathbf{x}} \\ 0 \end{pmatrix}$  insbesondere Optimallösung mit demselben Zielfunktionswert. Ansonsten wird die neue Variable  $x_{n+1}$  in der ersten Iteration von SIMPLEX in die Basis aufgenommen.

---

### Hinzufügen einer Nebenbedingung

---

Nehmen wir an, wir fügen eine neue Nebenbedingung der Form

$$\mathbf{a}^\top \mathbf{x} + y = \beta$$

mit  $\mathbf{a} \in \mathbb{R}^n$ ,  $\beta \in \mathbb{R}$  und künstlicher Variable  $y \geq 0$  zu (P) hinzu. Seien

$$\tilde{A} := \begin{pmatrix} A & \mathbf{0} \\ \mathbf{a}^\top & 1 \end{pmatrix} \in \mathbb{R}^{(m+1) \times (n+1)}, \quad \tilde{\mathbf{b}} := \begin{pmatrix} \mathbf{b} \\ \beta \end{pmatrix} \in \mathbb{R}^{m+1}, \quad \tilde{\mathbf{c}} = \begin{pmatrix} \mathbf{c} \\ 0 \end{pmatrix} \in \mathbb{R}^{n+1}$$

die entsprechende neue Nebenbedingungsmatrix, rechte Seite und Zielfunktion. Die Spalte  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  ist linear unabhängig zu den Spalten in  $\tilde{A}_{\cdot B}$ , also können wir die Basis zu  $\tilde{B} := B \cup \{n+1\}$  erweitern.

Für die zugehörige Basismatrix gilt dann

$$\tilde{A}_{\cdot \tilde{B}} = \begin{pmatrix} A_{\cdot B} & \mathbf{0} \\ \mathbf{a}_B^\top & 1 \end{pmatrix} \quad \text{und} \quad \tilde{A}_{\cdot \tilde{B}}^{-1} = \begin{pmatrix} A_{\cdot B}^{-1} & \mathbf{0} \\ -\mathbf{a}_B^\top A_{\cdot B}^{-1} & 1 \end{pmatrix}.$$

---

Damit ergeben sich die reduzierten Kosten zu

$$\tilde{\mathbf{z}}_{N_0} = \mathbf{c}_{N_0} - \tilde{A}_{\cdot N_0}^\top \tilde{A}_{\cdot \tilde{B}}^{-\top} \mathbf{c}_{\tilde{B}} = \mathbf{c}_{N_0} - (A_{\cdot N_0}^\top \quad \mathbf{a}_{N_0}) \begin{pmatrix} A_{\cdot B}^{-\top} & -A_{\cdot B}^{-\top} \mathbf{a}_B \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{c}_B \\ 0 \end{pmatrix} = \mathbf{c}_{N_0} - A_{\cdot N_0}^\top A_{\cdot B}^{-\top} \mathbf{c}_B = \tilde{\mathbf{z}}_{N_0} \geq \mathbf{0}.$$

Also ist  $\tilde{B}$  dual zulässig und wir können DUALSIMPLEX starten.

Erfüllt die bisherige Optimallösung bereits  $\mathbf{a}^\top \bar{\mathbf{x}} \leq \beta$ , so können wir

$$\bar{\mathbf{y}} := \beta - \mathbf{a}^\top \bar{\mathbf{x}} \geq 0$$

setzen und erhalten somit  $\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix} \geq \mathbf{0}$ . Wegen

$$\tilde{A}_{\cdot \tilde{B}} \begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix}_{\tilde{B}} = \begin{pmatrix} A_{\cdot B} \bar{\mathbf{x}}_B \\ \mathbf{a}_B^\top \bar{\mathbf{x}}_B + \bar{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \beta \end{pmatrix}$$

ist  $\tilde{B}$  primal zulässige Basis mit zugehöriger Basislösung  $\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix}$ . Nach Proposition 5.7 ist also  $\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix}$  bereits optimale Lösung mit dem gleichen Zielfunktionswert.

---

---

## 6 Die Ellipsoidmethode

Obwohl das Simplex-Verfahren in vielen praktischen Anwendungen zügig läuft, haben wir in Sektion 5.4 besprochen, dass es im worst-case für alle untersuchten Pivotregeln eine exponentielle (oder zumindest super-polynomielle) Anzahl Iterationen benötigt. Ob es eine Pivotregel gibt, die eine polynomielle Anzahl Schritte garantiert und ob es überhaupt einen polynomiellen „kombinatorischen“ Algorithmus zur Lösung von LPs gibt, ist die zentrale offene Frage in der linearen Optimierung. In diesem Kapitel betrachten wir ein „algebraisches“ Verfahren zur Lösung von LPs in polynomieller Zeit. Dabei bezieht sich „algebraisch“, darauf, dass Operationen wie Division und das Ziehen von Wurzeln benötigt werden und dass das Verhalten des Verfahrens von den konkreten Zahlenwerten in der Beschreibung des LPs abhängt.

Wir verwenden das Einheitskostenmodell als unser Berechnungsmodell und erinnern zunächst an die formale Definition (siehe *ADM*).

**Definition 6.1** (*ADM*). Ein *algorithmisches Problem* ist ein Tupel  $(\mathcal{I}, (\mathcal{S}_I)_{I \in \mathcal{I}})$  aus einer Menge von *Instanzen/Inputs*  $\mathcal{I}$  zusammen mit einer Familie von *Lösungsmengen*  $(\mathcal{S}_I)_{I \in \mathcal{I}}$ .

Ein Algorithmus  $A$  löst das Problem, falls er für jede Instanz  $I \in \mathcal{I}$  nach einer endlichen Anzahl Schritten terminiert und eine Lösung  $A(I) \in \mathcal{S}_I$  ausgibt.

Die (*worst-case*) *Laufzeit* von  $A$  ist gegeben durch

$$f(\kappa) := \max_{I \in \mathcal{I}, \langle I \rangle \leq \kappa} \{\text{Anzahl benötigter elementarer Operationen von } A \text{ zur Berechnung von } A(I)\},$$

wobei die *Kodierungslänge*  $\langle I \rangle$  von  $I$  die Anzahl Bits in der Beschreibung von  $I$  ist.

Der Algorithmus hat eine *polynomielle Laufzeit*, falls  $f(\kappa) = \kappa^{\mathcal{O}(1)}$ .

Ganz ähnlich zur polynomiellen Reduzierbarkeit von Entscheidungsproblemen (siehe *ADM*), können wir Komplexitätsklassen für algorithmische Probleme definieren, in Bezug auf ihre Lösbarkeit mit polynomieller Laufzeit.

**Definition 6.2.** Zwei algorithmische Probleme sind *polynomiell äquivalent*, wenn beide oder keins mit polynomieller Laufzeit lösbar sind.

Ob eine polynomielle Laufzeit möglich ist, hängt wesentlich von der Beschreibung der Instanzen ab. Da diese Beschreibung endlich sein muss, können wir uns im Folgenden auf Instanzen beschränken, die als LPs der Form

$$\text{(LP)} \quad \begin{array}{ll} \max & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{array}$$

mit rationalen Daten  $\mathbf{A} \in \mathbb{Q}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{Q}^m$ ,  $\mathbf{c} \in \mathbb{Q}^n$  gegeben sind. Wir bezeichnen mit  $K \in \mathbb{N}$  die größte natürliche Zahl, die als Zähler oder Nenner in der Beschreibung von (LP) auftritt. Wir nehmen dabei an, dass die Beschreibung von Instanzen die Binärkodierung aller Zähler und Nenner enthält. Damit genügt es die Laufzeit in den Parametern  $n$ ,  $m$  und  $K$  zu beschränken.

**Beobachtung 6.3.** Ein Algorithmus zur Lösung von LPs in der Form (LP) hat polynomielle Laufzeit, genau dann, wenn seine Laufzeit in  $(nm \log_2(K))^{O(1)}$  liegt.

*Beweis.* Die Kodierungslänge  $\kappa$  von (LP) ist offensichtlich mindestens  $\log_2(K)$ , da selbst unter Vernachlässigung von Vorzeichen und Trennzeichen so viele Bits benötigt werden, um alleine die Zahl  $K$  zu kodieren. Andererseits ist die Kodierungslänge mindestens  $nm$ , selbst wenn jeder Eintrag von  $A$  mit einem Bit kodiert werden kann. Also ist  $\kappa \geq (nm)^{1/2}(\log_2(K))^{1/2}$ . Andererseits ist  $\kappa = \mathcal{O}(nm \log(K))$ , selbst wenn alle Zahlen in der Beschreibung von (LP) von der Größenordnung  $K$  sind. Somit ist die Laufzeit in  $\kappa^{O(1)}$  genau dann, wenn sie in  $(nm \log_2(K))^{O(1)}$  liegt.  $\square$

Hat ein Algorithmus sogar Laufzeit  $(nm)^{O(1)}$ , sprechen wir von einer *streng polynomiellen* Laufzeit. Ob ein solcher Algorithmus zur Lösung von LPs existiert ist unbekannt (siehe Problem 5.19).

## 6.1 Reduktion auf volldimensionale Zulässigkeitsprobleme

Kennen wir geeignete Schranken an den optimalen Zielfunktionswert, lässt sich das Problem, den Optimalwert von (LP) zu finden, mittels binärer Suche auf das Entscheidungsproblem zurückführen, ob  $\mathcal{P}(A, \mathbf{b}) \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} \geq \gamma\} \neq \emptyset$  für gegebenes  $\gamma \in \mathbb{R}$  (siehe ADM). Wir wollen diese Erkenntnis auf das Finden von Optimallösungen und das Feststellen von Unbeschränktheit, also auf das folgende Suchproblem, erweitern:

### Optimierungsproblem (Suchproblem)

**Input:**  $A \in \mathbb{Q}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{Q}^m$ ,  $\mathbf{c} \in \mathbb{Q}^n$

**Problem:** finde  $\mathbf{x}^* \in \arg \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}(A, \mathbf{b})\}$ , oder  
entscheide, dass  $\mathcal{P}(A, \mathbf{b}) = \emptyset$ , oder

finde  $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$  mit  $\mathbf{x} + \lambda \mathbf{z} \in \mathcal{P}(A, \mathbf{b})$  für alle  $\lambda \geq 0$  und  $\mathbf{c}^\top \mathbf{z} > 0$ .

Wir führen das zugehörige Zulässigkeitsproblem ein, das wir als das Entscheidungsproblem ob  $\mathcal{P}(A, \mathbf{b}) \neq \emptyset$  sehen können, bei dem zusätzlich ein Zertifikat für den Fall  $\mathcal{P}(A, \mathbf{b}) \neq \emptyset$  verlangt ist.

### Zulässigkeitsproblem

**Input:**  $A \in \mathbb{Q}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{Q}^m$

**Problem:** finde  $\bar{\mathbf{x}} \in \mathcal{P}(A, \mathbf{b})$ , oder  
entscheide, dass  $\mathcal{P}(A, \mathbf{b}) = \emptyset$ .

Um die polynomielle Lösbarkeit von LPs zu zeigen dürfen wir uns auf Zulässigkeitsprobleme einschränken.

**Satz 6.4.** Das Optimierungsproblem und das Zulässigkeitsproblem sind polynomiell äquivalent.

*Beweis.* Jeder Algorithmus für das Optimierungsproblem löst per Definition auch das Zulässigkeitsproblem. Andersherum entscheidet ein Algorithmus für das Zulässigkeitsproblem bereits,

ob die zulässige Menge des Optimierungsproblems leer ist. Nehmen wir also an, das Optimierungsproblem (LP) sei zulässig.

Existiert eine Optimallösung, dann hat nach starker Dualität (Satz 4.17) auch das duale LP

$$(DP) \quad \begin{array}{ll} \min & \mathbf{b}^\top \mathbf{y} \\ \text{s.t.} & A^\top \mathbf{y} = \mathbf{c}, \\ & \mathbf{y} \geq \mathbf{0}, \end{array}$$

eine Optimallösung mit gleichem Zielfunktionswert. Damit genügt es wegen schwacher Dualität (Satz 4.5) eine zulässige Lösung von

$$\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & A^\top \\ \mathbf{0} & -A^\top \\ \mathbf{0} & -\mathbb{1} \\ -\mathbf{c}^\top & \mathbf{b}^\top \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \leq \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \\ -\mathbf{c} \\ -\mathbf{0} \\ 0 \end{pmatrix}$$

zu bestimmen. Beachte, dass die Kodierungslänge dieses Systems polynomiell in der Kodierungslänge von (LP) beschränkt ist.

Ist das kombinierte System unzulässig, muss, wegen Zulässigkeit von (LP), das System (DP) unzulässig sein. Nach dem Farkas-Lemma (Satz 4.13) angewendet auf (DP) existiert (durch Negieren und Reskalieren)  $\mathbf{z} \in \mathbb{R}^n$  mit

$$A\mathbf{z} \leq \mathbf{0} \quad \text{und} \quad \mathbf{c}^\top \mathbf{z} = 1.$$

Beachte, dass die Kodierungslänge dieses Systems wieder polynomiell in der Kodierungslänge von (LP) ist. Wir können also eine zulässige Lösung  $\bar{\mathbf{z}}$  dieses Systems bestimmen sowie eine zulässige Lösung von (LP). Damit ist  $\bar{\mathbf{x}} + \lambda \bar{\mathbf{z}} \in \mathcal{P}(A, \mathbf{b})$  für alle  $\lambda > 0$  und  $\mathbf{c}^\top \bar{\mathbf{z}} > 0$ . Also können wir das Optimierungsproblem auch im unbeschränkten Fall in polynomieller Zeit lösen.  $\square$

Wir werden einen Algorithmus beschreiben, mit dem, unter gewissen Voraussetzungen, das Zulässigkeitsproblem für volldimensionale Polyeder mit Laufzeit  $\mathcal{O}((nK)^{\mathcal{O}(1)})$  gelöst werden kann. Insbesondere kann in diesem Fall eine polynomielle Laufzeit selbst für exponentielle  $m$  erzielt werden. Dafür ist die Konstruktion im Beweis von Satz 6.4 allerdings unbrauchbar, da sie die Anzahl Dimensionen auf  $n + m$  erhöht, wobei im Allgemeinen auch eine anfängliche Volldimensionalität verloren geht.

Wir führen daher das Optimierungsproblem nochmal aufwändiger mittels binärer Suche auf das Zulässigkeitsproblem zurück. Zur Initialisierung der binären Suche benötigen wir eine obere Schranke an den Betrag des optimalen Zielfunktionswerts. Da der optimale Zielfunktionswert im Allgemeinen fraktional sein kann, benötigen wir außerdem ein Abbruchkriterium an die maximale numerische Auflösung der Suche und ein Rundungsverfahren, um von den finalen Schranken auf den exakten Zielfunktionswert  $\gamma^*$  zu schließen. Wir erhalten dann eine optimale Lösung als Lösung des Zulässigkeitsproblems mit der zusätzlichen Nebenbedingung  $\mathbf{c}^\top \mathbf{x} \geq \gamma^*$ .

Wir beginnen damit, den maximalen Betrag des optimalen Zielfunktionswerts abzuschätzen. Dazu beschränken wir zunächst Nenner und Zähler aller Ecken von  $\mathcal{P}(A, \mathbf{b})$ . Wir verwenden folgende Schranke an die Determinante einer Matrix.

**Lemma 6.5.** Sei  $B \in \mathbb{R}^{n \times n}$  mit  $|B_{ij}| \leq K$ . Dann ist  $|\det(B)| \leq (nK)^n$ .

*Beweis.* Nach der Leibniz-Formel (siehe *Lineare Algebra II*) gilt

$$\det(B) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n B_{i,\sigma(i)},$$

wobei  $S_n$  die symmetrische Gruppe aller Permutationen von  $\{1, \dots, n\}$  bezeichnet und  $\operatorname{sgn}(\sigma) \in \{-1, 1\}$ . Damit ist

$$|\det(B)| \leq \sum_{\sigma \in S_n} \prod_{i=1}^n |B_{i,\sigma(i)}| \leq n! \cdot K^n \leq (nK)^n. \quad \square$$

Damit können wir Ecken beschränken.

**Proposition 6.6.** Für jede Ecke  $\hat{\mathbf{x}}$  von  $\mathcal{P}(A, \mathbf{b})$  existieren  $r_0, \dots, r_n \in \mathbb{Z}$  mit  $|r_i| \leq (nK)^{3n^2}$  für alle  $i \in \{0, \dots, n\}$  und  $\hat{x}_i = r_i/r_0$  für alle  $i \in \{1, \dots, n\}$ .

*Beweis.* Nach Satz 3.23 existiert eine Menge von Zeilenindizes  $I \subseteq \operatorname{eq}(\hat{\mathbf{x}})$  mit  $A_I \cdot \hat{\mathbf{x}} = \mathbf{b}_I$  wobei  $|I| = n$  und  $A_I$  regulär ist. Nach der Cramerschen Regel (siehe *Lineare Algebra II*) ist

$$\hat{x}_i = \frac{\det A_{[i]}}{\det A_I} \quad (6.1)$$

für  $i \in I$ , wobei  $A_{[i]}$  die Matrix bezeichnet, die durch Ersetzen der  $i$ -ten Spalte von  $A_I$  durch  $\mathbf{b}_I$  entsteht. Da  $A_I$  eine Teilmatrix von  $A$  ist, sind alle Einträge von  $A_{[i]}$  und  $A_I$  durch  $K$  beschränkt. Also folgt mit Lemma 6.5, dass

$$|\det A_{[i]}| \leq (nK)^n \quad \text{und} \quad |\det A_I| \leq (nK)^n.$$

Sei  $q \in \mathbb{N}$  das Produkt der Nenner der Einträge von  $A_I$  und  $\mathbf{b}_I$ . Dann ist  $q \cdot |\det A_I| \in \mathbb{N}$ ,  $q \cdot |\det A_{[i]}| \in \mathbb{N}$  und  $q \leq K^{n^2+n} \leq K^{2n^2}$ .

Mit  $r_i := q \det(A_{[i]}) \in \mathbb{Z}$  und  $r_0 := q \det(A_I) \in \mathbb{Z}$  können wir  $\hat{x}_i = r_i/r_0$  schreiben, wobei

$$\begin{aligned} |r_i| &= q \cdot |\det A_{[i]}| \leq (nK)^{3n^2}, \\ |r_0| &= q \cdot |\det A_I| \leq (nK)^{3n^2}. \end{aligned} \quad \square$$

Aus Proposition 6.6 folgt, dass wir uns auf das Lösen von Optimierungsproblemen über Polytopen beschränken können. Im unbeschränkten Fall können wir einen Strahl  $\mathbf{x} + \lambda \mathbf{z} \in \mathcal{P}(A, \mathbf{b})$  mit Hilfe des Farkas-Lemmas (analog zum Beweis von Satz 6.4) finden. Beachte, dass in allen Fällen der zusätzliche Aufwand in  $(nK)^{\mathcal{O}(1)}$  liegt.

**Satz 6.7.** Das LP

$$\max\{\mathbf{c}^\top \mathbf{x} : A\mathbf{x} \leq \mathbf{b}, -U \cdot \mathbf{1} \leq \mathbf{x} \leq U \cdot \mathbf{1}\}$$

ist genau dann unzulässig für  $U = (nK)^{3n^2}$ , wenn (LP) unzulässig ist. Ansonsten ist (LP) beschränkt mit Optimallösung  $\mathbf{x}^*$  genau dann, wenn  $\mathbf{x}^*$  Optimallösung für  $U = (nK)^{3n^2}$  und  $U = (nK)^{3n^2} + 1$  ist.

*Beweis.* Wir können  $\mathbb{R}^n$  in Polyeder der Form  $\{\mathbf{x} \in \mathbb{R}^n : \tilde{\mathbb{I}}\mathbf{x} \geq \mathbf{0}\}$  partitionieren, wobei  $\tilde{\mathbb{I}}$  Einträge  $\{-1, 1\}$  auf der Diagonalen hat und sonst alle Einträge 0 sind. Also existiert ein solches  $\tilde{\mathbb{I}}$ , so dass

$$\max\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\} = \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \tilde{\mathbb{I}}\mathbf{x} \geq \mathbf{0}\}.$$

Da das Teilsystem  $-U \cdot \mathbf{1} \leq \mathbf{x} \leq U \cdot \mathbf{1}$  invariant unter der Variablentransformationen  $\mathbf{z} = \tilde{\mathbb{I}}\mathbf{x}$  ist, können wir  $\tilde{\mathbb{I}} = \mathbb{I}$  annehmen. Damit dürfen wir ohne Beschränkung der Allgemeinheit die Aussage für das System

$$\max\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{0} \leq \mathbf{x} \leq U \cdot \mathbf{1}\} \quad (6.2)$$

betrachten.

Ist (6.2) für  $U = (nK)^{3n^2}$  zulässig, dann natürlich auch für  $U = \infty$ . Nehmen wir andersherum an, es sei für  $U = \infty$  zulässig. Dann ist  $\arg \min\{\mathbf{1}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$  ein Polytop und besitzt somit eine Ecke  $\hat{\mathbf{x}}$ , die auch Ecke der zulässigen Menge von (6.2) für  $U = \infty$  ist. Da die Dimension von (6.2) weiterhin  $n$  ist und alle Zähler und Nenner weiterhin durch  $K$  beschränkt sind, folgt mit Proposition 6.6, dass  $\hat{\mathbf{x}}$  und damit (6.2) auch für  $U = (nK)^{3n^2}$  zulässig ist.

Ist (6.2) zulässig und beschränkt für  $U = \infty$ , dann folgt aus der starken Dualität (Satz 4.17), dass eine Optimallösung existiert. Die Existenz einer Startecke  $\hat{\mathbf{x}}$  impliziert zusammen mit der Korrektheit und Endlichkeit des Simplexverfahrens (Sätze 5.10 und 5.12), die Existenz einer optimalen Ecke  $\mathbf{x}^*$ . Diese muss wiederum nach Proposition 6.6 auch zulässig und somit optimal für  $U = (nK)^{3n^2}$  und  $U = (nK)^{3n^2} + 1$  sein. Also ist jede Optimallösung für  $U = (nK)^{3n^2}$  und  $U = (nK)^{3n^2} + 1$  auch Optimallösung für  $U = \infty$ .

Nimmt (6.2) andersherum für  $U = (nK)^{3n^2}$  und  $U = (nK)^{3n^2} + 1$  den gleichen Optimalwert an, dann ist für jede Optimallösung  $\mathbf{x}^*$  zu  $U = (nK)^{3n^2}$  im System mit  $U = (nK)^{3n^2} + 1$  keine Ungleichung aus  $\mathbf{x} \leq ((nK)^{3n^2} + 1) \cdot \mathbf{1}$  aktiv. Diese Ungleichungen können somit weggelassen werden, ohne die Optimalität von  $\mathbf{x}^*$  zu beeinträchtigen. Insbesondere ist (6.2) für  $U = \infty$  zulässig und beschränkt.  $\square$

Damit können wir sowohl den maximal möglichen Wert, als auch die maximale Auflösung des optimalen Zielfunktionswerts beschränken.

**Proposition 6.8.** Ist  $\mathcal{P}(\mathbf{A}, \mathbf{b}) \neq \emptyset$  ein Polytop, so ist

$$\max\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\} \in \left\{ \frac{p}{q} : p \in \mathbb{Z}, |p| \leq (nK)^{4n^2}, q \in \mathbb{N}, q \leq (nK)^{4n^2} \right\}.$$

*Beweis.* Da  $\mathcal{P} \neq \emptyset$  ein Polytop ist, wird das Optimum an einer Ecke angenommen (Korollar 2.47 und Satz 3.23). Es genügt also, die Werte  $\mathbf{c}^\top \hat{\mathbf{x}}$  für Ecken  $\hat{\mathbf{x}}$  von  $\mathcal{P}$  abzuschätzen. Seien  $r_0, r_1, \dots, r_n \in \mathbb{Z}$  mit  $\hat{x}_i = r_i/r_0$  aus Proposition 6.6 und sei die Zielfunktion beschrieben ist über  $c_i = s_i/t_i$  mit  $s_1, \dots, s_n \in \mathbb{Z}$  und  $t_1, \dots, t_n \in \mathbb{N}$  für alle  $i \in \{1, \dots, n\}$ . Dann ist

$$\mathbf{c}^\top \hat{\mathbf{x}} = \frac{1}{r_0 t} \sum_{i=1}^n s_i r_i \bar{t}_i$$

mit  $t := \prod_{i=1}^n t_i$  und  $\bar{t}_i := t/t_i$ . Nach Proposition 6.6 ist also  $\mathbf{c}^\top \mathbf{x} = p/q$  mit  $p \in \mathbb{Z}$ ,  $q \in \mathbb{N}$  und

$$|p| = \left| \sum_{i=1}^n s_i r_i \bar{t}_i \right| \leq \sum_{i=1}^n |s_i r_i \bar{t}_i| \leq \sum_{i=1}^n K \cdot (nK)^{3n^2} \cdot K^{n-1} \leq (nK)^{4n^2},$$

$$q = r_0 t \leq (nK)^{3n^2} \cdot K^n \leq (nK)^{4n^2}. \quad \square$$

Wir können nun die binäre Suche konkretisieren. Proposition 6.8 liefert sowohl einerseits anfängliche Schranken an den Suchbereich, als auch eine untere Schranke wie klein wir den Suchbereich werden lassen müssen. Wir erhalten den folgenden Algorithmus, wobei wir das Ergebnis geschickt runden (siehe Übung).

---

**Algorithmus :** BINARYSEARCH( $A, \mathbf{b}, \mathbf{c}$ )

---

**input:**  $A \in \mathbb{Q}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{Q}^m$ ,  $\mathbf{c} \in \mathbb{Q}^n$ ,  
so dass  $\mathcal{P}(A, \mathbf{b}) \neq \emptyset$  ein Polytop ist

**output:** Optimalwert  $\gamma^*$  von (LP)

---

$Q \leftarrow (nK)^{4n^2}$

$\ell \leftarrow -Q$

$u \leftarrow Q$

**while**  $u - \ell > \frac{1}{4Q^2}$ :

$m \leftarrow \frac{\ell + u}{2}$

**if**  $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}, \mathbf{c}^\top \mathbf{x} \geq m\} = \emptyset$ :

$u \leftarrow m$

**else**

$\ell \leftarrow m$

finde  $p \in \mathbb{Z}$  und  $q \in \{1, \dots, 2Q - 1\}$  mit  $|\ell - \frac{p}{q}| \leq \frac{1}{2Qq}$  (runden)

**return**  $\frac{p}{q}$

---

Eine Eigenschaft der binären Suche ist, dass sie mit einer polynomielle Anzahl Iterationen auskommt, die nicht von der Anzahl Nebenbedingungen  $m$  abhängt.

**Satz 6.9.** BINARYSEARCH ist korrekt und terminiert in  $\mathcal{O}(n^2 \log(nK))$  Iterationen.

*Beweis.* Der Algorithmus erhält die Invariante, dass  $\gamma^* \in [\ell, u]$ : Dies gilt am Anfang nach Proposition 6.8 und wird per Definition von BINARYSEARCH bei der Aktualisierung von  $\ell$  und  $u$  sichergestellt.

Weiterhin gilt anfangs  $u - \ell = 2Q$  und bei Terminierung  $u - \ell \geq \frac{1}{8Q^2}$ . Da  $u - \ell$  in jeder Iteration genau halbiert wird, ist die Anzahl an Iterationen gegeben durch

$$\lceil \log_2(16Q^3) \rceil \leq 3 \log_2 Q + 5 = 12n^2 \log_2(nK) + 5 = \mathcal{O}(n^2 \log(nK)).$$

Es bleibt zu zeigen, dass  $\gamma^* = p/q$ . Nach Proposition 6.8 ist  $\gamma^* = s/t$  mit  $s \in \mathbb{Z}$  und  $t \in \mathbb{N}$ , wobei  $t \leq Q$ . Nach der Invariante gilt am Ende

$$s/t - \ell = \gamma^* - \ell \leq u - \ell \leq \frac{1}{4Q^2} < \frac{1}{2Qq}.$$

Nach Dreiecksungleichung ist

$$\frac{|sq - pt|}{qt} = \left| \frac{s}{t} - \frac{p}{q} \right| \leq \left| \frac{s}{t} - \ell \right| + \left| \ell - \frac{p}{q} \right| < \frac{1}{Qq} \leq \frac{1}{tq},$$

also  $|sq - pt| < 1$ . Aus der Ganzzahligkeit von  $p, t, s, q$  folgt  $pt = sq$ , also  $p/q = s/t = \gamma^*$ . Damit gibt BINARYSEARCH das korrekte Ergebnis zurück.  $\square$

Insbesondere liefert ein Algorithmus für das Entscheidungsproblem mit polynomieller Laufzeit in  $n$  und  $\log_2(K)$  einen ebensolchen Algorithmus für das Optimierungsproblem.

**Korollar 6.10.** Gegeben ein Algorithmus, der das Zulässigkeitsproblem mit Laufzeit  $(n \log_2(K))^{\mathcal{O}(1)}$  löst, kann das Optimierungsproblem ebenfalls mit Laufzeit  $(n \log_2(K))^{\mathcal{O}(1)}$  gelöst werden.

*Beweis.* Ein Algorithmus für das Zulässigkeitsproblem entscheidet per Definition bereits, ob die zulässige Menge des Optimierungsproblems leer ist. Weiterhin können wir uns nach Satz 6.7 für die Lösung des Optimierungsproblems auf Polytope beschränken. Dabei wächst die Kodierungslänge um

$$2n \log_2((nK)^{3n^2}) = 6n^3 \log_2(nK) = (n \log_2(K))^{\mathcal{O}(1)},$$

also polynomiell. Für Polytope können wir den Optimalwert mittels der binären Suche bestimmen. Dabei nimmt die Kodierungslänge von  $u$  und  $\ell$  in Summe in jeder Iteration maximal um eine additive Konstante zu, also ist die Kodierungslänge des Polytops  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}, \mathbf{c}^\top \mathbf{x} \geq m\}$  in jeder Iteration linear in der Kodierungslänge von (LP) beschränkt. Nach Satz 6.9 liefert somit ein polynomieller Algorithmus für das Zulässigkeitsproblem in polynomieller Zeit den Optimalwert. Zuletzt können wir einen solchen Algorithmus, per Definition des Zulässigkeitsproblems, verwenden, um eine optimale Lösung als ein beliebiges Element  $\mathbf{x}^* \in \{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}, \mathbf{c}^\top \mathbf{x} \geq \gamma^*\}$  zu bestimmen.  $\square$

Zum Abschluss zeigen wir noch, dass wir beim Zulässigkeitsproblem annehmen dürfen  $\mathcal{P}(A, \mathbf{b})$  sei volldimensional. Dazu zeigen wir, dass eine kleine Störung von (LP) dessen Zulässigkeit nicht beeinträchtigt. Beachte, dass die Störung zwar nur von  $n$  und  $K$  abhängt, aber Laufzeit der Umrechnung der Lösung durchaus von  $m$  abhängen kann.

**Satz 6.11.** Aus jeder Lösung von

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b} + \varepsilon \mathbf{1}\} \tag{6.3}$$

mit  $\varepsilon := (n+2)^{-1}(nK)^{-3n-1}$  lässt sich in polynomieller Zeit eine Lösung von

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} \leq \mathbf{b}\} \tag{6.4}$$

konstruieren, und umgekehrt.

*Beweis (für Interessierte).* Jede Lösung von (6.4) ist offensichtlich bereits eine Lösung von (6.3). Nehmen wir nun an wir haben eine Lösung  $\bar{\mathbf{y}}$  von (6.3) gegeben.

Wir dürfen annehmen, dass  $A$  vollen Spaltenrang hat, denn linear abhängige Spalten gehören zu überflüssigen Variablen, da keine Vorzeichenbeschränkungen vorliegen. Wie können linear abhängige Spalten in polynomieller Zeit mittels der Gauß-Elimination (siehe *Lineare Algebra I*) entfernen und den Wert der entsprechenden Komponente von  $\mathbf{y}$  auf die anderen Komponenten umlegen.

Sei  $I := \{i \in \{1, \dots, m\} : A_{i \cdot} \bar{\mathbf{y}} > b_i\}$ . Falls  $A_{I \cdot}$  nicht  $n$  unabhängige Zeilen enthält, existiert eine linear unabhängige Zeile  $A_j$  mit  $j \notin I$ , da  $A$  vollen Spaltenrang hat. Damit hat das System

$$\{\mathbf{z} \in \mathbb{R}^n : \begin{pmatrix} A_{I \cdot} \\ A_j \cdot \end{pmatrix} \mathbf{z} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}\}$$

eine Lösung  $\bar{\mathbf{z}}$ . Wir können also (in polynomieller Zeit)  $\lambda > 0$  finden, so dass  $A(\bar{\mathbf{y}} + \lambda \bar{\mathbf{z}}) \leq \mathbf{b} + \varepsilon \mathbf{1}$  und  $A_{j' \cdot}(\bar{\mathbf{y}} + \lambda \bar{\mathbf{z}}) > b_{j'}$  für ein  $j' \notin I$ . Wir können also die Lösung  $\bar{\mathbf{y}} + \lambda \bar{\mathbf{z}}$  von (6.3) betrachten und damit die Menge  $I$  vergrößern. Da wir diesen Vorgang nötigenfalls bis zu  $n$  mal wiederholen können, dürfen wir im Folgenden annehmen, dass  $I' \subseteq I$  existiert, so dass  $|I'| = n$  und  $A_{I' \cdot}$  regulär.

Dann sind alle Zeilen  $A_j$  mit  $j \in \{1, \dots, m\}$  linear abhängig von den Zeilen in  $A_{I' \cdot}$ , es existieren also  $\beta_{ji}$  mit

$$A_j \cdot = \sum_{i \in I'} \beta_{ji} A_{i \cdot} \quad (6.5)$$

Nach der Cramerschen Regel (siehe *Lineare Algebra II*) ist  $\beta_{ij} = \frac{D_{ji}}{|D|}$ , wobei  $D_{ji}$  (bis auf sein Vorzeichen) und  $D = \det(A_{I' \cdot})$  Determinanten von Matrizen mit Einträgen aus  $A$  sind. Nach Lemma 6.5 ist

$$|D_{ij}| \leq (nK)^n, |D| \leq (nK)^n. \quad (6.6)$$

Sei nun  $\bar{\mathbf{x}}$  die Lösung von  $A_{I' \cdot} \mathbf{x} = \mathbf{b}_{I'}$ . Nach der Cramerschen Regel sind die Komponenten von  $\bar{\mathbf{x}}$  von der Form  $\bar{x}_i = \frac{D_i}{|D|}$ , wobei  $D_i$  die Determinante einer Matrix mit Einträgen aus  $A$  und  $\mathbf{b}$  ist. Weiterhin gilt für alle  $j \in \{1, \dots, m\}$ , dass

$$\begin{aligned} |D|(A_j \cdot \bar{\mathbf{x}} - b_j) &\stackrel{(6.5)}{=} \sum_{i \in I'} D_{ji} A_{i \cdot} \bar{\mathbf{x}} - |D| b_j \\ &= \sum_{i \in I'} D_{ji} b_i - |D| b_j + |D|(A_j \cdot \bar{\mathbf{y}} - A_j \cdot \bar{\mathbf{y}}) \\ &\stackrel{(6.5)}{=} |D|(A_j \cdot \bar{\mathbf{y}} - b_j) - \sum_{i \in I'} D_{ij} (A_{i \cdot} \bar{\mathbf{y}} - b_i). \end{aligned}$$

Wegen  $A_j \cdot \bar{\mathbf{y}} - b_j \leq \varepsilon$  und  $|A_{i \cdot} \bar{\mathbf{y}} - b_i| \leq \varepsilon$  ist also

$$|D|(A_j \cdot \bar{\mathbf{x}} - b_j) \leq \varepsilon(|D| + \sum_{i \in I'} |D_{ij}|) \stackrel{(6.6)}{\leq} \varepsilon(n+1)(nK)^n < (nK)^{-2n-1}.$$

Der Nenner der Einträge von  $\bar{\mathbf{x}}$  ist ein Vielfaches der rationalen Zahl  $|D|$ , lässt sich also auf den Nenner von  $D_{ij}$  kürzen. Dieser ist analog zu Lemma 6.5 maximal  $(nK)^n$ . Also ist die kleinstmögliche positive linke Seite andererseits mindestens  $(nK)^{-n-(n+1)}$ , wobei  $K^{n+1}$  eine obere Schranke

an den Hauptnenner der Einträge von  $A_j$  und  $b_j$  ist. Es folgt, dass  $A_j \bar{x} = b_j$ . Da dies für alle  $j \in \{1, \dots, m\}$  gilt, ist  $\bar{x}$  eine zulässige Lösung von (6.4). Da  $\bar{x}$  die Lösung eines linearen Gleichungssystems mit Einträgen aus  $A$  und  $\mathbf{b}$  ist, können wir es mit Hilfe der Gauß-Elimination in polynomieller Zeit bestimmen.  $\square$

Ist  $\mathcal{P}(A, \mathbf{b})$  also nicht volldimensional, können wir stattdessen das Zulässigkeitsproblem für  $\mathcal{P}(A, \mathbf{b} + \varepsilon \mathbf{1})$  betrachten: Die Kodierungslänge vergrößert sich polynomiell um maximal  $m \log_2(\varepsilon^{-1})$ ,  $K$  vergrößert sich polynomiell um maximal  $\log_2(\varepsilon^{-1})$ , und  $n$  bleibt unberührt. Ist  $\mathcal{P}(A, \mathbf{b} + \varepsilon \mathbf{1}) = \emptyset$ , so ist natürlich auch  $\mathcal{P}(A, \mathbf{b}) = \emptyset$ . Ansonsten können wir aus  $\bar{y} \in \mathcal{P}(A, \mathbf{b} + \varepsilon \mathbf{1})$  nach Satz 6.11 in polynomieller Zeit ein  $\bar{x} \in \mathcal{P}(A, \mathbf{b})$  konstruieren. Es bleibt noch festzustellen, dass  $\mathcal{P}(A, \mathbf{b} + \varepsilon \mathbf{1})$  volldimensional oder leer ist.

**Beobachtung 6.12.** Ist  $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\} \neq \emptyset$ , dann ist  $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b} + \varepsilon \mathbf{1}\}$  mit  $\varepsilon > 0$  volldimensional.

*Beweis.* Es ist  $\mathcal{P}^\circ := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} < \mathbf{b} + \varepsilon \mathbf{1}\}$  der Schnitt von endlich vielen offenen Mengen, also offen.

Außerdem ist

$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\} \subseteq \mathcal{P}^\circ \subseteq \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b} + \varepsilon \mathbf{1}\}.$$

Aus  $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\} \neq \emptyset$  folgt also  $\mathcal{P}^\circ \neq \emptyset$ . Dann enthält  $\mathcal{P}^\circ$  und damit  $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b} + \varepsilon \mathbf{1}\}$  eine Umgebung und ist somit volldimensional.  $\square$

## 6.2 Die Ellipsoidmethode

Wir wollen nun die binäre Suche auf mehrdimensionale Suchprobleme verallgemeinern. Dabei verwenden wir die folgende namensgebende Verallgemeinerung von Intervallen als unsere Suchräume.

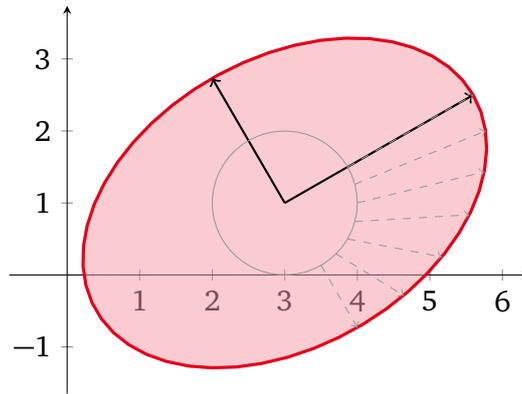
**Definition 6.13.** Ein *Ellipsoid* mit Mittelpunkt  $\mathbf{a} \in \mathbb{R}^n$  ist eine Menge

$$\mathcal{E}(A, \mathbf{a}) := \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} - \mathbf{a})^\top A^{-1}(\mathbf{x} - \mathbf{a}) \leq 1\},$$

mit  $A \in \mathbb{R}^{n \times n}$  (und damit  $A^{-1}$ ) symmetrisch und positiv definit.

Wir definieren dabei das Ellipsoid  $\mathcal{E}(A, \mathbf{a})$  über die Inverse der Matrix  $A$ , da sich zentrale Eigenschaften von  $A$  ableiten lassen. So entsprechen die Eigenvektoren von  $A$  beispielsweise den Symmetrieachsen des Ellipsoids und die Eigenwerte von  $A$  entsprechen den Quadraten der Halbachsen. Weiterhin beschreibt die Quadratwurzel von  $A$  die Deformation von  $\mathcal{E}(A, \mathbf{a})$  bezüglich der Einheitskugel.

**Beobachtung 6.14.** Es ist  $\mathcal{E}(A, \mathbf{a}) = \{A^{1/2}\mathbf{u} + \mathbf{a} : \mathbf{u} \in \mathcal{B}_1(\mathbf{0})\}$  eine affine Transformation der Einheitskugel  $\mathcal{B}_1(\mathbf{0}) = \mathcal{E}(\mathbb{1}, \mathbf{0}) = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_2 \leq 1\}$ , wobei aus  $A$  symmetrisch, positiv definit folgt, dass  $A^{1/2}$  eindeutig bestimmt ist über  $A^{1/2}A^{1/2} = A$ .



**Abbildung 6.1:** Zweidimensionales Ellipsoid aus Beispiel 6.15. Die Transformation von der Einheitskugel aus Beobachtung 6.14 ist angedeutet.

**Beispiel 6.15.** Betrachte das zweidimensionale Ellipsoid  $\mathcal{E}(A, \mathbf{a})$  mit  $\mathbf{a} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ , Radien  $r = 3$ ,  $q = 2$  und Hauptachsen  $r \cdot \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3\sqrt{3} \\ 3 \end{pmatrix}$  und  $q \cdot \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix} = \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}$ , wobei  $\alpha = \pi/6$  (siehe Abb. 6.1). Die Matrix  $A$  lässt sich berechnen als

$$A = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} r^2 & 0 \\ 0 & q^2 \end{pmatrix} \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 31 & 5\sqrt{3} \\ 5\sqrt{3} & 21 \end{pmatrix},$$

$$A^{-1} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} r^{-2} & 0 \\ 0 & q^{-2} \end{pmatrix} \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} = \frac{1}{48} \begin{pmatrix} 7 & -\frac{5}{\sqrt{3}} \\ -\frac{5}{\sqrt{3}} & \frac{31}{3} \end{pmatrix},$$

$$A^{1/2} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & q \end{pmatrix} \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 11 & \sqrt{3} \\ \sqrt{3} & 9 \end{pmatrix}. \quad \triangle$$

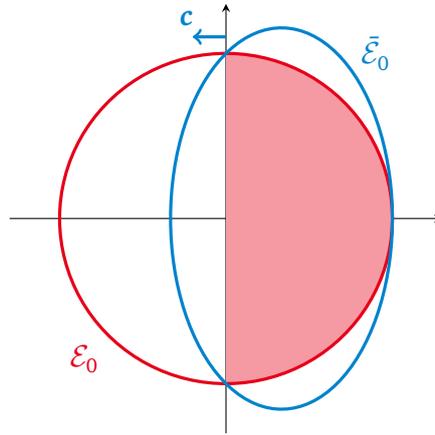
Um wie bei der binären Suche verfahren zu können, benötigen wir Ellipsoid, das unseren anfänglichen Suchraum beschränkt, also insbesondere  $\mathcal{P}(A, \mathbf{b})$  enthält. Wir erinnern daran, dass wir uns nach Satz 6.7 auf Polytope beschränken können.

**Proposition 6.16.** Ist  $\mathcal{P} = \mathcal{P}(A, \mathbf{b})$  ein Polytop, so ist  $\mathcal{P} \subseteq \mathcal{B}_R(\mathbf{0}) = \mathcal{E}(R^2 \mathbb{1}, \mathbf{0})$  mit  $R = (nK)^{4n^2}$ .

*Beweis.* Nach Satz 2.7 genügt es zu zeigen, dass alle Ecken  $\hat{\mathbf{x}}$  von  $\mathcal{P}(A, \mathbf{b})$  in  $\mathcal{B}_R(\mathbf{0})$  enthalten sind. Nach der Dreiecksungleichung und Proposition 6.6 gilt

$$\|\hat{\mathbf{x}}\|_2 \leq \|\hat{\mathbf{x}}\|_1 \leq n(nK)^{3n^2} \leq (nK)^{4n^2}. \quad \square$$

Während wir bei der eindimensionalen binären Suche Intervalle jeweils exakt in zwei Hälften teilen können, ist es nicht klar, wie wir einen Suchraum in der Form eines Ellipsoids verfeinern können. Insbesondere können wir ein Ellipsoid nicht in zwei nichttriviale Ellipsoide partitionieren. Wollen wir keinen Teil unseres Suchraums verlieren, müssen wir stattdessen mit einer Überdeckung durch zwei Ellipsoide arbeiten. Um dennoch das logarithmische Verhalten der binären Suche zu bewahren, müssen wir die Größe dieser Ellipsoide beschränken. Diese werden mittels des Volumens  $\text{vol}(\mathcal{E}) := \int_{\mathbf{x} \in \mathcal{E}} 1 \, d\mathbf{x}$  messen. Das folgende Lemma erlaubt es uns wegen Beobachtung 6.14 das Volumen eines Ellipsoids über das Volumen der Einheitskugel zu berechnen.



**Abbildung 6.2:** Das Ellipsoid  $\mathcal{E}_0 = \mathcal{E}(\mathbb{1}, \mathbf{0}) = \mathcal{B}_1(\mathbf{0})$  zusammen mit dem überdeckenden Ellipsoid  $\bar{\mathcal{E}}_0 \supseteq \mathcal{E}_0 \cap \mathcal{H}$  mit  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^2 : -x_1 \leq 0\}$  im zweidimensionalen Fall des Beweises von Satz 6.18.

**Lemma 6.17.** Sei  $f(\mathbf{x}) = D\mathbf{x} + \mathbf{d}$  eine affine Abbildung und  $\mathcal{M} \subseteq \mathbb{R}^n$ . Dann gilt

$$\text{vol}(f(\mathcal{M})) = |\det(D)| \cdot \text{vol}(\mathcal{M}).$$

*Beweis.* Mit der Variablentransformation  $\mathbf{y} = f(\mathbf{x})$  gilt nach dem Transformationssatz (siehe *Analysis II*)

$$\text{vol}(f(\mathcal{M})) = \int_{\mathbf{y} \in f(\mathcal{M})} 1 \, d\mathbf{y} = \int_{\mathbf{x} \in \mathcal{M}} |\det(J_f(\mathbf{x}))| \, d\mathbf{x},$$

wobei  $J_f(\mathbf{x}) = (\partial_j f_i(\mathbf{x}))_{ij} = D$  die Jacobimatrix von  $f$  bezeichnet.  $\square$

Der folgende Satz ist das zentrale Resultat zur Funktionsweise der Ellipsoidmethode. Er besagt, dass wir den Suchraum durch zwei Ellipsoide überdecken können, die jeweils komplementäre Hälften des Suchraums enthalten und dennoch in jeder Dimension um einen Faktor kleiner sind. Die konstruierten Ellipsoide sind tatsächlich kleinstmöglich und werden *Löwner-John-Ellipsoide* genannt.

**Satz 6.18.** Sei  $\mathcal{E} = \mathcal{E}(A, \mathbf{a})$  ein Ellipsoid in Raumdimension  $n \geq 2$ , sei  $\mathbf{c} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$  und  $\mathcal{H} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^\top \mathbf{x} \leq \mathbf{c}^\top \mathbf{a}\}$ . Setze

$$\mathbf{d} := \frac{1}{\sqrt{\mathbf{c}^\top A \mathbf{c}}} A \mathbf{c}, \quad \bar{\mathbf{a}} := \mathbf{a} - \frac{1}{n+1} \mathbf{d}, \quad \bar{A} := \frac{n^2}{n^2-1} \left( A - \frac{2}{n+1} \mathbf{d} \mathbf{d}^\top \right).$$

Dann ist  $\bar{\mathcal{E}} = \mathcal{E}(\bar{A}, \bar{\mathbf{a}})$  ein Ellipsoid mit  $\bar{\mathcal{E}} \supseteq \mathcal{E} \cap \mathcal{H}$  und  $\text{vol}(\bar{\mathcal{E}}) < e^{-\frac{1}{2n}} \cdot \text{vol}(\mathcal{E})$ .

*Beweis (für Interessierte).* Wir betrachten zunächst den Fall  $A = A_0 := \mathbb{1}$ ,  $\mathbf{a} = \mathbf{a}_0 := \mathbf{0}$  und  $\mathbf{c} = \mathbf{c}_0 := \begin{pmatrix} -1 \\ \mathbf{0} \end{pmatrix}$  (siehe Abb. 6.2). Sei  $\mathcal{E}_0 := \mathcal{E}(A_0, \mathbf{a}_0) = \mathcal{B}_1(\mathbf{0})$ ,  $\mathcal{H}_0 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}_0^\top \mathbf{x} \geq \mathbf{c}_0^\top \mathbf{a}_0\} = \{\mathbf{x} \in \mathbb{R}^n : x_1 \geq 0\}$  und  $\bar{\mathcal{E}}_0 := \mathcal{E}(\bar{A}_0, \bar{\mathbf{a}}_0)$  mit

$$\mathbf{d}_0 = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}, \quad \bar{\mathbf{a}}_0 = \begin{pmatrix} (n+1)^{-1} \\ \mathbf{0} \end{pmatrix}, \quad \bar{A}_0 = \frac{n^2}{n^2-1} \begin{pmatrix} \frac{n-1}{n+1} & \mathbf{0}^\top \\ \mathbf{0} & \mathbb{1} \end{pmatrix}.$$

Es ist  $\bar{A}_0$  symmetrisch und positiv definit, also ist  $\bar{\mathcal{E}}_0$  ein Ellipsoid. Außerdem ist  $\bar{A}_0^{-1}$  eine Diagonalmatrix und insbesondere gilt für  $\mathbf{x} \in \mathbb{R}^n$ , dass

$$\begin{aligned} (\mathbf{x} - \bar{\mathbf{a}}_0)^\top \bar{A}_0^{-1} (\mathbf{x} - \bar{\mathbf{a}}_0) &= \sum_{i=1}^n (\bar{A}_0^{-1})_{ii} (x_i - (\bar{\mathbf{a}}_0)_i)^2 \\ &= \left(x_1 - \frac{1}{n+1}\right)^2 \frac{(n+1)^2}{n^2} + \frac{n^2-1}{n^2} \sum_{i=2}^n x_i^2 \\ &= \frac{(n+1)^2}{n^2} \left(x_1^2 - \frac{2x_1}{n+1} + \frac{1}{(n+1)^2}\right) + \frac{n^2-1}{n^2} \sum_{i=2}^n x_i^2 \\ &= \frac{n^2-1}{n^2} \sum_{i=1}^n x_i^2 + \frac{2(n+1)}{n^2} x_1(x_1-1) + \frac{1}{n^2}. \end{aligned}$$

Sei nun  $\mathbf{x} \in \mathcal{E}_0 \cap \mathcal{H}_0$ . Dann ist  $0 \leq x_1 \leq 1$ , also  $x_1(x_1-1) \leq 0$ . Aus  $\mathbf{x} \in \mathcal{E}_0 = \mathcal{B}_1(\mathbf{0})$  folgt  $\sum_{i=1}^n x_i^2 \leq 1$ . Damit ergibt sich

$$(\mathbf{x} - \bar{\mathbf{a}}_0)^\top \bar{A}_0^{-1} (\mathbf{x} - \bar{\mathbf{a}}_0) \leq \frac{n^2-1}{n^2} + \frac{1}{n^2} = 1.$$

Also ist  $\mathbf{x} \in \bar{\mathcal{E}}_0$  und somit  $\bar{\mathcal{E}}_0 \supseteq \mathcal{E}_0 \cap \mathcal{H}_0$ .

Nach Lemma 6.17 gilt außerdem

$$\begin{aligned} \frac{\text{vol}(\bar{\mathcal{E}}_0)}{\text{vol}(\mathcal{E}_0)} &= \frac{\sqrt{\det(\bar{A}_0)} \cdot \text{vol}(\mathcal{B}_1)}{\sqrt{\det(A_0)} \cdot \text{vol}(\mathcal{B}_1)} \\ &= \sqrt{\det(A')} \\ &= \left( \frac{n^{2n}}{(n+1)^n (n-1)^n} \cdot \frac{n-1}{n+1} \right)^{1/2} \\ &= \frac{n}{n+1} \left( \frac{n^2}{(n+1)(n-1)} \right)^{(n-1)/2} \\ &= \left( 1 - \frac{1}{n+1} \right) \left( 1 + \frac{1}{(n+1)(n-1)} \right)^{(n-1)/2}. \end{aligned}$$

Mit  $1 + \alpha < e^\alpha$  für alle  $\alpha \neq 0$  (wegen  $n \geq 2$ ) folgt

$$\frac{\text{vol}(\bar{\mathcal{E}}_0)}{\text{vol}(\mathcal{E}_0)} < e^{-\frac{1}{n+1}} \left( e^{\frac{1}{(n+1)(n-1)}} \right)^{\frac{n-1}{2}} = e^{-\frac{1}{2(n+1)}}.$$

Betrachte nun den allgemeinen Fall. Nach Beobachtung 6.14 ist  $S(\mathcal{E}) = \mathcal{E}_0$  mit  $S(\mathbf{x}) = A^{-1/2}(\mathbf{x} - \mathbf{a})$ . Sei  $R$  ein Rotationsmatrix, die den Raum um den Winkel  $\pi$  um die Achse  $\frac{A^{1/2}\mathbf{c}}{\|A^{1/2}\mathbf{c}\|} + \begin{pmatrix} -1 \\ \mathbf{0} \end{pmatrix}$  dreht. Dann ist  $R$  orthonormal und daher  $R^\top R = \mathbb{1}$  und es gilt

$$RA^{1/2}\mathbf{c} = \|A^{1/2}\mathbf{c}\| \begin{pmatrix} -1 \\ \mathbf{0} \end{pmatrix}.$$

Wir definieren die affine Abbildung

$$T(\mathbf{x}) := RS(\mathbf{x}) = RA^{-1/2}(\mathbf{x} - \mathbf{a}).$$

Da die Einheitskugel  $\mathcal{E}_0$  invariant unter Rotation ist, gilt

$$T(\mathcal{E}) = RS(\mathcal{E}) = \mathcal{E}_0.$$

Es ist  $\mathbf{x} \in \mathcal{H}$  genau dann, wenn

$$0 \leq \mathbf{c}^\top (\mathbf{x} - \mathbf{a}) = \mathbf{c}^\top A^{1/2} R^\top R A^{-1/2} (\mathbf{x} - \mathbf{a}) = (R A^{1/2} \mathbf{c})^\top T(\mathbf{x}) = \|A^{1/2} \mathbf{c}\| \begin{pmatrix} -1 \\ \mathbf{0} \end{pmatrix}^\top T(\mathbf{x}),$$

also genau dann wenn  $\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}^\top T(\mathbf{x}) \leq 0$ , beziehungsweise wenn  $T(\mathbf{x}) \in \mathcal{H}_0$ . Es folgt dass  $T(\mathcal{H}) = \mathcal{H}_0$ .

Man kann zeigen (Übung), dass  $\bar{\mathcal{E}}$  ein Ellipsoid ist. Schließlich gilt (ohne Beweis) auch  $T(\bar{\mathcal{E}}) = \bar{\mathcal{E}}_0$ .

Also ist

$$T(\mathcal{E}) \cap T(\mathcal{H}) = \mathcal{E}_0 \cap \mathcal{H}_0 \subseteq \bar{\mathcal{E}}_0 = T(\bar{\mathcal{E}}).$$

Mit Injektivität von  $R$  und  $A$  (positiv Definit) und damit von  $T$ , folgt  $\bar{\mathcal{E}} \supseteq \mathcal{E} \cap \mathcal{H}$ . □

Proposition 6.16 liefert einen anfänglichen Suchraum und Satz 6.18 beschreibt, wie wir den Suchraum iterativ verfeinern können. Es fehlt nun nur noch ein Abbruchkriterium für die Suche. Der folgende Satz macht klar, warum wir uns auf volldimensionale Polyeder einschränken müssen, da andernfalls  $\text{vol}(\mathcal{P}(A, \mathbf{b})) = 0$ .

**Proposition 6.19.** Ist  $\mathcal{P}(A, \mathbf{b}) \neq \emptyset$  ein volldimensionales Polytop in Raumdimension  $n \geq 2$ , dann gilt

$$\text{vol}(\mathcal{P}(A, \mathbf{b})) > (nK)^{-5n^3}.$$

*Beweis.* Nach Satz 2.21 und Satz 3.23 ist  $\mathcal{P}(A, \mathbf{b})$  die konvexe Hülle seiner Ecken. Da  $\mathcal{P}(A, \mathbf{b})$  volldimensional ist, enthält es  $n + 1$  affin unabhängige Ecken  $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n$  und damit deren konvexe Hülle  $\mathcal{Q} = \text{conv}(\{\hat{\mathbf{x}}_0, \dots, \hat{\mathbf{x}}_n\})$ . Es gilt (siehe Übung)

$$\text{vol}(\mathcal{Q}) = \frac{1}{n!} |\det(\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_0 \quad \dots \quad \hat{\mathbf{x}}_n - \hat{\mathbf{x}}_0)| = \frac{1}{n!} \left| \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ \hat{\mathbf{x}}_0 & \hat{\mathbf{x}}_1 & \dots & \hat{\mathbf{x}}_n \end{pmatrix} \right| > 0.$$

Sei  $q$  das Produkt aller (positiven) Hauptnenner von  $\hat{\mathbf{x}}_0, \dots, \hat{\mathbf{x}}_n$ . Nach Proposition 6.6 ist  $q \leq (nK)^{3n^2(n+1)}$ . Außerdem ist  $qn! \text{vol}(\mathcal{Q}) \geq 1$ , also mit  $n \geq 2$ ,

$$\text{vol}(\mathcal{P}(A, \mathbf{b})) \geq \text{vol}(\mathcal{Q}) \geq \frac{1}{qn!} \geq (nK)^{-3n^2(n+1)} n^{-n} > (nK)^{-5n^3}. \quad \square$$

Wir haben damit alle Komponenten zur Formulierung der Ellipsoidmethode zusammengetragen. Wir testen in jedem Schritt, ob der Mittelpunkt des aktuellen Ellipsoids bereits ein zulässiger Punkt ist. Andernfalls erhalten wir eine trennende Hyperebene durch den Mittelpunkt über die linke Seite einer verletzten Zeile des Ungleichungssystems. Insbesondere ist  $\mathcal{P}(A, \mathbf{b})$  in einer Hälfte des Ellipsoids enthalten. Wir verfeinern unseren Suchraum zu dem überdeckenden Ellipsoids dieser Hälfte des Ellipsoids nach Satz 6.18.

---

**Algorithmus :** ELLIPSOIDMETHOD( $A, \mathbf{b}$ )

---

**input:**  $A \in \mathbb{Q}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{Q}^m$

**output:** Punkt  $\bar{\mathbf{x}} \in \mathcal{P}(A, \mathbf{b})$ , oder  $\emptyset$  falls  $\text{vol}(\mathcal{P}(A, \mathbf{b})) = 0$

---

$A^{(0)} \leftarrow (nK)^{8n^2} \cdot \mathbb{I}$ ,  $\mathbf{a}^{(0)} \leftarrow \mathbf{0}$

$\mathcal{E}_0 \leftarrow \mathcal{E}(A^{(0)}, \mathbf{a}^{(0)})$ ,  $k \leftarrow 0$

**while**  $k < \lceil \ln(nK) \rceil \cdot 18n^4 + \lceil \ln(2) \rceil \cdot 2n^2$ :

**if**  $\exists i \in \{1, \dots, m\}$  with  $\mathbf{A}_i \cdot \mathbf{a}^{(k)} > b_i$ :

$\mathbf{d} \leftarrow (\mathbf{A}_i \mathbf{A}^{(k)} \mathbf{A}_i^\top)^{-1/2} \cdot \mathbf{A}^{(k)} \mathbf{A}_i^\top$

$\mathbf{A}^{(k+1)} \leftarrow \frac{n^2}{n^2-1} (\mathbf{A}^{(k)} - \frac{2}{n+1} \mathbf{d} \mathbf{d}^\top)$

$\mathbf{a}^{(k+1)} \leftarrow \mathbf{a}^{(k)} - \frac{1}{n+1} \mathbf{d}$

$\mathcal{E}_{k+1} \leftarrow \mathcal{E}(\mathbf{A}^{(k+1)}, \mathbf{a}^{(k+1)})$

**else**

**return**  $\mathbf{a}^{(k)}$

$k \leftarrow k + 1$

**return**  $\emptyset$

---

**Beispiel 6.20.** Betrachte das Polytop  $\mathcal{P}$ , das durch die zulässige Menge von

$$\begin{aligned} -x_1 - x_2 &\leq -2, \\ 3x_1 &\leq 4, \\ -2x_1 + 2x_2 &\leq 3 \end{aligned}$$

gegeben ist. Abb. 6.3 zeigt die Ausführung der Ellipsoidmethode beginnend mit  $\mathcal{E}_0 = \mathcal{B}_7(\mathbf{0}) \supset \mathcal{P}$ . △

An diesem Beispiel sehen wir insbesondere, dass jedes Ellipsoid durch Stauchung in äußere Normalenrichtung des Halbraums  $\mathcal{H}$ , der  $\mathcal{P}$  enthält, entsteht und durch Streckung in den orthogonalen Richtungen. Diesen Effekt illustriert das nächste Beispiel nochmal.

**Beispiel 6.21.** Betrachte das Polytop  $\mathcal{P}$ , das durch die zulässige Menge von

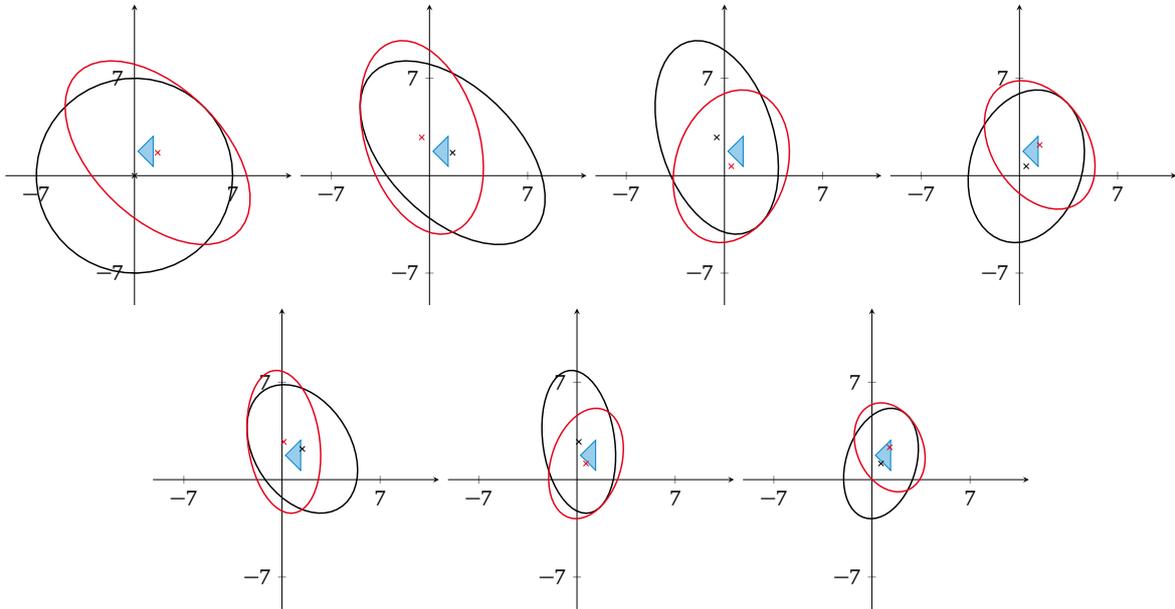
$$\begin{aligned} -\frac{1}{5} &\leq x_1 \leq \frac{1}{5}, \\ \frac{17}{20} &\leq x_2 \leq \frac{18}{20} \end{aligned}$$

gegeben ist (siehe Übung). Abb. 6.4 zeigt die Ausführung der Ellipsoidmethode beginnend mit der Kugel  $\mathcal{B}_1(\mathbf{0}) \supset \mathcal{P}$ . △

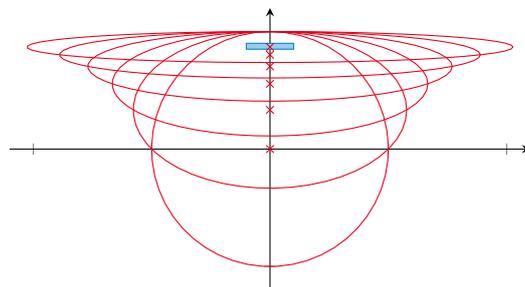
Um die Korrektheit der Ellipsoidmethode zu zeigen, müssen wir im Wesentlichen noch nachweisen, dass die Abbruchbedingung richtig gewählt ist. Wir erhalten insbesondere eine polynomielle Schranke an die benötigte Anzahl Iterationen, die wiederum nur von  $n$  und  $\log(K)$  abhängt.

**Satz 6.22.** ELLIPSOIDMETHOD ist korrekt und benötigt  $\mathcal{O}(n^4 \log(nK))$  Iterationen.

---



**Abbildung 6.3:** Iterationen der Ellipsoidmethode für Beispiel 6.20.



**Abbildung 6.4:** Iterationen der Ellipsoidmethode für Beispiel 6.21.

*Beweis.* Offensichtlich benötigt ELLIPSOIDMETHOD  $U = \lceil \ln(nK) \rceil \cdot 18n^4 + \lceil \ln(2) \rceil 2n^2 = \mathcal{O}(n^4 \log(nK))$  Iterationen. Gibt sie einen Vektor  $\mathbf{a}^{(k)}$  zurück, so gilt  $\mathbf{a}^{(k)} \in \mathcal{P}(A, \mathbf{b})$ . Ist  $\mathcal{P}(A, \mathbf{b})$  nicht volldimensional, dann ist  $\text{vol}(\mathcal{P}(A, \mathbf{b})) = 0$ , also ist das Ergebnis von ELLIPSOIDMETHOD in diesem Fall korrekt. Sei nun  $\mathcal{P}(A, \mathbf{b})$  volldimensional.

Wir zeigen, dass der Algorithmus die Invariante erhält, dass  $\mathcal{E}_k$  ein Ellipsoid ist, mit  $\mathcal{P}(A, \mathbf{b}) \subseteq \mathcal{E}_k$  und  $\text{vol}(\mathcal{E}_k) \leq e^{-i/2n} \text{vol}(\mathcal{E}_0)$ . Die Inklusion gilt anfangs nach Proposition 6.16 und die anderen Eigenschaften sind für  $\mathcal{E}_0$  trivial. Bei der Berechnung von  $\mathcal{E}_{k+1}$  verletzt  $\mathbf{a}^{(k)}$  die Ungleichung  $A_i \cdot \mathbf{x} \leq b_i$ . Also gilt mit der Invariante in Iteration  $k$ , dass  $\mathcal{P}(A, \mathbf{b}) \subseteq \mathcal{H} \cap \mathcal{E}_k$  mit  $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n : A_i \cdot \mathbf{x} \leq A_i \cdot \mathbf{a}^{(k)}\}$ . Nach Satz 6.18 gilt somit, dass  $\mathcal{E}_{k+1}$  ein Ellipsoid ist mit  $\mathcal{E}_{k+1} \supseteq \mathcal{H} \cap \mathcal{E}_k \supseteq \mathcal{P}(A, \mathbf{b})$  und  $\text{vol}(\mathcal{E}_{k+1}) \leq e^{-1/2n} \text{vol}(\mathcal{E}_k) \leq e^{-(k+1)/2n} \text{vol}(\mathcal{E}_0)$ , wobei wir die Invariante für Iteration  $k$  verwendet haben.

Nehmen wir nun an, ELLIPSOIDMETHOD terminiert ohne einen Vektor zurück zu geben, also nach Iteration  $U$ . Für  $\mathcal{E}_0$  gilt nach Beobachtung 6.14 und Lemma 6.17

$$\text{vol}(\mathcal{E}_0) = \left| \det((nK)^{4n^2} \mathbb{1}) \right| \cdot \text{vol}(\mathcal{B}_1(\mathbf{0})) \leq (nK)^{4n^3} \cdot 2^n.$$

Mit der Invariante gilt also  $\mathcal{P}(A, \mathbf{b}) \subseteq \mathcal{E}_k$  und somit

$$\text{vol}(\mathcal{P}(A, \mathbf{b})) \leq \text{vol}(\mathcal{E}_k) \leq e^{-U/2n} (nK)^{4n^3} 2^n \leq (nK)^{-9n^3} (nK)^{4n^3} \leq (nK)^{-5n^3}.$$

Nach Proposition 6.19 folgt mit Volldimensionalität, dass  $\mathcal{P}(A, \mathbf{b}) = \emptyset$ , also ist das Ergebnis von ELLIPSOIDMETHOD korrekt.  $\square$

Um die Laufzeit der Ellipsoidmethode abzuschätzen, müssen wir den Aufwand in jeder Iteration untersuchen. Zunächst einmal muss die trennende Hyperebene gefunden werden. Dies erfolgt in linearer Zeit, da es genügt alle Zeilen des Ungleichungssystems zu prüfen. Allerdings hängt diese lineare Laufzeit auch von  $m$  ab, im Gegensatz zu allen bisherigen Schranken. Im Allgemeinen können wir natürlich auch nicht hoffen das Zulässigkeitsproblem zu lösen, ohne zumindest alle Einträge von  $A$  und  $\mathbf{b}$  anzuschauen. Daher ist eine lineare Abhängigkeit der Laufzeit von  $m$  nicht vermeidbar.

Wesentlich problematischer ist es, die Laufzeit für die Verfeinerung des Suchraums zu beschränken. Einerseits müssen hier Wurzeln gezogen werden, was nur mit endlicher Genauigkeit passieren kann. Durch die entsprechende Fehlerfortpflanzung (siehe *Einführung in die Numerik*), geht die Korrektheit des Verfahrens verloren. Insbesondere kann es passieren, dass  $A$  nicht mehr positiv definit ist, oder das Ellipsoid nicht mehr  $\mathcal{E}(A, \mathbf{b})$  enthält. Beide Probleme lassen sich durch eine leichte Skalierung der Ellipsoide vermeiden. Außerdem treten bei der Ausführung der Ellipsoidmethode Zahlen auf, die durch Produkte von polynomiell vielen Termen entstehen und deren Kodierungslänge nicht polynomiell beschränkt ist. Dieses Problem lässt sich durch sorgfältige Wahl der Berechnungsgenauigkeit ebenfalls vermeiden. Insgesamt konnte so die polynomielle Lösbarkeit der Linearen Optimierung gezeigt werden.

**Satz 6.23** (Khachiyan, 1979). LPs sind in polynomieller Zeit lösbar.

Dieses Resultat liefert allerdings keinen stark polynomiellen Algorithmus, da die Laufzeit von  $\log K$  abhängt (siehe Problem 5.19). Außerdem ist es wichtig zu erwähnen, dass die Ellipsoidmethode zwar aus theoretischer Sicht, im Sinne der worst-case-Analyse, dem Simplex-Verfahren überlegen ist, es sich in der Anwendung aber typischerweise genau entgegengesetzt darstellt: Das Simplex-Verfahren ist oft sehr effizient, während die hohe (polynomielle) Anzahl an aufwändigen Iterationen der Ellipsoidmethode in der Regel zu schlechter Performanz führt.

---

### 6.3 Separieren und Optimieren

---

Fast alle Resultate in den vorangegangenen Abschnitten kommen ohne eine Abhängigkeit von der Anzahl  $m$  an Nebenbedingungen aus. Insbesondere ist der einzige Schritt der Ellipsoidmethode mit einer konkreten Abhängigkeit von den Nebenbedingungen der Separierungsschritt, in dem die linke Seite einer Nebenbedingung als trennende Hyperebene für den aktuellen Mittelpunkt verwendet wird. An dieser Stelle würde uns eine beliebige trennende Hyperebene ausreichen. Wir brauchen also lediglich eine Lösung des folgenden Problems.

#### Separierungsproblem

**Input:** Polyeder  $\mathcal{P}$ ,  $\mathbf{y} \in \mathbb{Q}^n$

**Problem:** entscheide, dass  $\mathbf{y} \in \mathcal{P}$ , oder  
finde  $\mathbf{c} \in \mathbb{Q}^n$  mit  $\mathbf{c}^\top \mathbf{y} > \max\{\mathbf{c}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ .

Können wir das Separierungsproblem effizient lösen, so erhalten wir also eine effiziente Lösung für das Zulässigkeitsproblem. Das gilt selbst dann, wenn die Nebenbedingungen in einer kompakten Form formuliert sind aber  $m$  sehr groß ist.

**Satz 6.24.** Existiert ein Algorithmus mit Laufzeit  $f(\kappa)$  für das Separierungsproblem über ein volldimensionales Polytop  $\mathcal{P}$ , dann existiert ein Algorithmus mit Laufzeit  $(n \log_2 K)^{\mathcal{O}(1)} \cdot f(\kappa)$  für das Zulässigkeitsproblem über  $\mathcal{P}$ .

*Beweis.* Der einzige Schritt in der Ellipsoidmethode, dessen Laufzeit von  $m$  abhängt, ist das Finden einer Zeile von  $A$ , die den Mittelpunkt des aktuellen Ellipsoids von  $\mathcal{P}$  trennt. Da diese trennende Hyperebene genau eine Zeile von  $A$  ist, ist für die Funktion der Ellipsoidmethode unerheblich. Wir können diesen Schritt also durch Anwendung eines Algorithmus für das Separierungsproblem über  $\mathcal{P}$  ersetzen. Die Aussage von Satz 6.22 bleibt damit erhalten. Wie erwähnt lassen sich die Rechnungen in jeder Iteration in polynomieller Zeit ausführen (ohne Beweis). Da diese Rechnungen im Raum  $\mathbb{Q}^n$  stattfinden und in keiner Weise von  $m$  abhängen, folgt die Aussage.  $\square$

Die Umkehrung dieser Aussage können wir nicht ohne eine Modifikation des Polytops zeigen.

**Definition 6.25.** Das *Polare* einer Menge  $\mathcal{M} \subseteq \mathbb{R}^n$  ist

$$\mathcal{M}^\circ := \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^\top \mathbf{x} \leq 1 \forall \mathbf{x} \in \mathcal{M}\}.$$

Es gilt die folgende Komplementarität.

**Proposition 6.26** (Übung). Sei  $\mathcal{P}$  ein Polytop mit  $\mathbf{0} \in \mathcal{P} \setminus \partial \mathcal{P}$ . Dann ist  $\mathcal{P}^\circ$  ein Polytop und es gilt  $(\mathcal{P}^\circ)^\circ = \mathcal{P}$ .

Damit können wir die Umkehrung von Satz 6.24 formulieren.

**Satz 6.27.** Existiert ein Algorithmus mit Laufzeit  $f(\kappa)$  für das Optimierungsproblem über ein Polytop  $\mathcal{P}$  (mit beliebigen Zielfunktionen), dann existiert ein Algorithmus mit Laufzeit  $\mathcal{O}(n \log_2 K) \cdot f(\kappa)$  für das Separierungsproblem über  $\mathcal{P}^\circ$ .

*Beweis.* Wir können das Separierungsproblem  $\mathcal{P}^\circ$  und einen gegebenen Punkt  $\mathbf{y} \in \mathbb{Q}^n$  wie folgt lösen: Wir bestimmen zunächst  $\mathbf{c} \in \arg \max\{\mathbf{y}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$  mit Hilfe des Optimierungsalgorithmus. Ist  $\mathbf{y}^\top \mathbf{c} \leq 1$ , dann ist offensichtlich  $\mathbf{y} \in \mathcal{P}^\circ$ . Sonst ist  $\mathbf{c}$  der gesuchte Normalvektor einer Hyperebene, die  $\mathbf{y}$  strikt von  $\mathcal{P}^\circ$  trennt.  $\square$

Können wir das Optimierungsproblem also über eine Klasse von Polytopen effizient lösen, die unter polarer Transformationen abgeschlossen ist, so erhalten wir einen effizienten Algorithmus für das Separierungsproblem über diese Klasse. Schwächer formuliert erhalten wir zusammen mit Satz 6.24 die folgende Aussage.

**Satz 6.28.** Existiert ein Algorithmus mit Laufzeit  $f(\kappa)$  für das Separierungsproblem oder das Optimierungsproblem über volldimensionale Polytope, dann existiert ein Algorithmus mit Laufzeit  $(n \log_2 K)^{O(1)} \cdot f(\kappa)$  für das jeweils andere Problem über volldimensionale Polytope.

*Beweis.* Aus einem Algorithmus für das Separierungsproblem können wir über Satz 6.24 und Korollar 6.10 einen Algorithmus für das Optimierungsproblem erhalten.

Wollen wir nun mit Hilfe eines Algorithmus für das Optimierungsproblem über  $\mathcal{P}$ , das Separierungsproblem für  $\mathbf{y} \in \mathbb{Q}^n$  über  $\mathcal{P}$  lösen, so können wir zunächst  $\mathcal{P}$  so transformieren, dass  $\mathbf{0} \in \mathcal{P} \setminus \partial \mathcal{P}$ : Dazu bestimmen wir Punkte  $\mathbf{x}_1 = \arg \max\{\mathbf{1}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$  und  $\mathbf{x}_2 \in \arg \max\{-\mathbf{1}^\top \mathbf{x} : \mathbf{x} \in \mathcal{P}\}$ . Da  $\mathcal{P}$  volldimensional ist, gilt  $\mathbf{x}_1 \neq \mathbf{x}_2$ . Wir können nun iterativ affin unabhängige Punkte bestimmen, indem wir in die beiden Richtungen  $\mathbf{c}$  und  $-\mathbf{c}$  optimieren, wobei  $\mathbf{c}$  orthogonal zu den bisherigen Punkten ist. Wegen Volldimensionalität von  $\mathcal{P}$  erhalten wir nach  $n$  Iterationen genau  $n$  affin unabhängige Punkte. Der Durchschnitt dieser Punkte liegt im Inneren von  $\mathcal{P}$  und wir können ihn durch geeignete Koordinatentransformation auf  $\mathbf{0}$  verschieben. Dabei nutzen wir implizit, dass alle bisherigen Resultate translationsinvariant sind. Insgesamt können wir also annehmen, dass  $\mathbf{0} \in \mathcal{P} \setminus \partial \mathcal{P}$ .

Nach Satz 6.27 liefert ein Algorithmus für das Optimierungsproblem über  $\mathcal{P}^\circ$  einen Separierungsalgorithmus für  $(\mathcal{P}^\circ)^\circ$ . Wegen Proposition 6.26 und  $\mathbf{0} \in \mathcal{P} \setminus \partial \mathcal{P}$  ist  $(\mathcal{P}^\circ)^\circ = \mathcal{P}$ , also können wir das Separierungsproblem lösen.  $\square$

Insbesondere können wir die polynomielle Äquivalenz auf das Separierungsproblem erweitern. Die *polynomielle Äquivalenz von Optimierung und Separierung* ist ein zentrales Resultat der kombinatorischen Optimierung. Der Einfachheit halber beschränken wir uns auf Polytope.

**Korollar 6.29.** Das Optimierungsproblem, das Zulässigkeitsproblem, und das Separierungsproblem sind über Polytopen polynomiell äquivalent.

*Beweis.* Wir verwenden Satz 6.11, um auf den volldimensionalen Fall zu reduzieren. Das fügt einen polynomiellen Faktor (auch in  $m$ ) zur Laufzeit hinzu. Nun folgt die Aussage aus Satz 6.4 und Satz 6.28.  $\square$

Wir geben zuletzt ein Beispiel in dem eine effiziente Lösung möglich ist, obwohl die Anzahl  $m$  an Nebenbedingungen exponentiell in  $n$  ist. Weitere Beispiele werden in den Vorlesungen *Diskrete Optimierung* und *Optimierung in Transport und Verkehr* behandelt.

**Beispiel 6.30.** Beim *Traveling-Salesperson-Problem (TSP)* siehe ADM) ist ein ungerichteter Graph  $G = (V, E)$  mit Kantenkosten  $\mathbf{c} \in \mathbb{R}^E$  gegeben. Gesucht ist einen Kreis  $C \subseteq E$ , der alle Knoten

jeweils genau einmal besucht und dabei die Kosten  $\sum_{e \in C} c_e$  minimiert. Wir können dieses Problem als diskretes Optimierungsproblem formulieren, wobei  $x_e = 1$  genau dann, wenn  $e \in C$ :

$$\begin{aligned} \min \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } \sum_{e \in \delta(v)} x_e &= 2, & \forall v \in V, \\ \sum_{e \in \delta(S)} x_e &\geq 2, & \forall \emptyset \subset S \subset V, \\ \mathbf{x} &\in \{0, 1\}^E. \end{aligned}$$

Dieses Problem ist NP-schwer. Verzichten wir auf die Ganzzahligkeitsbedingungen, erhalten wir ein LP, das eine untere Schranke an die Kosten einer optimalen Tour bestimmt. Dieses LP hat allerdings exponentiell viele Nebenbedingungen, da es alle nichttrivialen Teilmengen von  $V$  betrachtet.

Dennoch können wir das Separierungsproblem für einen Punkt  $\bar{\mathbf{x}} \in \mathbb{R}^n$  effizient lösen. Die Nebenbedingungen  $\sum_{e \in \delta(v)} \bar{x}_e = 2$  und  $0 \leq \bar{x}_e \leq 1$  lassen sich offensichtlich effizient prüfen. Für die anderen Nebenbedingungen betrachten wir den gerichteten Graphen  $(V, E') := (V, \{(u, v), (v, u) : \{u, v\} \in E\})$  mit Kantenkapazitäten  $\mu((u, v)) := \bar{x}_{\{u, v\}}$  für alle  $(u, v) \in E'$ . Für jedes Paar  $s, t \in V$  bestimmen wir einen minimalen  $s$ - $t$ -Schnitt  $S_{s,t}$  im Flussnetzwerk  $((V, E'), \mu, s, t)$ . Dies ist in polynomieller Zeit in  $n$  möglich, indem wir ein Netzwerkflussproblem lösen (siehe ADM). Ist  $\mu(S_{s,t}) \geq 2$  für alle  $s, t \in V$ , dann ist  $\bar{\mathbf{x}}$  eine zulässige Lösung. Sonst liefert ein Schnitt mit  $\mu(S_{s,t}) < 2$  eine trennende Ungleichung.

Nach Satz 6.28 können wir somit das LP in polynomieller Zeit lösen, sofern es volldimensional ist, obwohl es exponentiell viele Nebenbedingungen besitzt. Das LP ist zum Beispiel volldimensional, wenn alle Knoten Grad mindestens 3 haben (warum?).  $\triangle$



---

# 7 Grundlagen der Nichtlinearen Optimierung

Die Nichtlineare Optimierung behandelt Optimierungsprobleme (NLPs) der Form

$$\begin{array}{ll} \text{(NLP)} & \min f(\mathbf{x}) \\ & \text{s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0}, \end{array}$$

mit  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  und  $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^k$ . Insbesondere wird im Allgemeinen nicht vorausgesetzt, dass die Funktionen linear oder konvex sind.

Wir wollen uns in dieser Vorlesung der Einfachheit halber auf lineare Nebenbedingungen beschränken, also auf Probleme der Form

$$\begin{array}{ll} \text{(NL)} & \min f(\mathbf{x}) \\ & \text{s.t. } A\mathbf{x} \leq \mathbf{b}, \end{array}$$

mit  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ . Der allgemeine Fall wird ausführlich in der Vorlesung *Nichtlineare Optimierung* behandelt.

Wir haben einige nützliche Eigenschaften von linearen Optimierungsproblemen kennen gelernt. Insbesondere treten nur globale Optima auf, es gibt endlich viele Randpunkte (Ecken), die als Optimallösungen in Frage kommen und es gilt starke Dualität. Diese Eigenschaften erlauben es letztendlich LPs in polynomialer Zeit zu lösen. Für nichtlineare Optimierungsprobleme gelten diese Eigenschaften im Allgemeinen nicht.

---

## 7.1 Optimalitätsbedingungen

---

Wir beginnen damit, Optimalitätsbedingungen für die Nichtlineare Optimierung abzuleiten. Dazu gehen wir analog zur linearen Optimierung so vor, dass wir zunächst notwendige Bedingungen in Bezug auf zulässige Richtungen formulieren und diese dann mit Hilfe des Farkas-Lemmas in eine handlichere Form überführen.

---

### 7.1.1 Tangential- und Linearisierungskegel

---

Wir erweitern zunächst Definition 4.9 auf NLPs mit linearen Nebenbedingungen. Im Allgemeinen ist eine kompliziertere Definition nötig. Dies wird in der Vorlesung *Nichtlineare Optimierung* genauer behandelt.

**Definition 7.1.** Ein Vektor  $\mathbf{s} \in \mathbb{R}^n$  heißt *zulässige Richtung* von (NL) in  $\bar{\mathbf{x}} \in \mathcal{P}(A, \mathbf{b})$ , wenn  $\lambda > 0$  existiert, so dass  $\bar{\mathbf{x}} + \lambda \mathbf{s} \in \mathcal{P}(A, \mathbf{b})$ . Der *Tangentialkegel*  $\mathcal{Z}(\bar{\mathbf{x}}) \subseteq \mathbb{R}^n$  in  $\bar{\mathbf{x}}$  ist der (topologische) Abschluss der Menge der zulässigen Richtungen in  $\bar{\mathbf{x}}$ .

Damit können wir eine Optimalitätsbedingung analog zu Satz 4.11 formulieren. Im Gegensatz zum linearen Fall ist diese Bedingung nur notwendig, aber nicht hinreichend. Zum Beispiel kann es Punkte  $\bar{\mathbf{x}}$  mit  $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$  geben, die keine lokalen Minimalpunkte sind.

**Satz 7.2.** Für jeden lokalen Minimalpunkt  $\bar{\mathbf{x}}$  von (NL) mit  $f \in C^1$  gilt

$$(\nabla f(\bar{\mathbf{x}}))^\top \mathbf{s} \geq 0 \quad \forall \mathbf{s} \in \mathcal{Z}(\bar{\mathbf{x}}).$$

*Beweis.* Sei  $\mathbf{s}$  zulässige Richtung im lokalen Minimalpunkt  $\bar{\mathbf{x}}$ . Wegen Zulässigkeit von  $\mathbf{s}$ , Konvexität von  $\mathcal{P}(A, \mathbf{b})$  und da  $\bar{\mathbf{x}}$  lokal minimal ist, existiert  $\bar{\lambda} > 0$  mit

$$\bar{\mathbf{x}} + \lambda \mathbf{s} \in \mathcal{P}(A, \mathbf{b}) \wedge f(\bar{\mathbf{x}}) \leq f(\bar{\mathbf{x}} + \lambda \mathbf{s}) \quad \forall \lambda \in [0, \bar{\lambda}].$$

Daraus folgt

$$(\nabla f(\bar{\mathbf{x}}))^\top \mathbf{s} = \lim_{\lambda \searrow 0} \frac{f(\bar{\mathbf{x}} + \lambda \mathbf{s}) - f(\bar{\mathbf{x}})}{\lambda} \geq 0.$$

Dies gilt für alle zulässigen Richtungen  $\mathbf{s}$  in  $\bar{\mathbf{x}}$ , also auch für den Abschluss  $\mathcal{Z}(\bar{\mathbf{x}})$  all dieser Richtungen.  $\square$

Wir wollen nun diese Optimalitätsbedingung in eine nützlichere Form bringen. Dazu erweitern wir zunächst die Definition von aktiven Nebenbedingungen (Definition 3.17) auf nichtlineare Ungleichungssysteme.

**Definition 7.3.** Sei  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  und  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$ . Dann ist die Menge der in  $\bar{\mathbf{x}} \in \mathcal{X}$  aktiven Nebenbedingungen gegeben durch

$$\text{eq}(\bar{\mathbf{x}}) := \{i \in \{1, \dots, m\} : g_i(\bar{\mathbf{x}}) = 0\}.$$

Damit können wir den Begriff des Linearisierungskegels einführen. Intuitiv ist der Linearisierungskegel in  $\bar{\mathbf{x}}$  der Kegel, der durch die in  $\bar{\mathbf{x}}$  linearisierten, aktiven Nebenbedingungen erzeugt wird. Wie zu erwarten, werden wir sehen, dass dieser Kegel bei linearen Nebenbedingungen genau mit dem Tangentialkegel übereinstimmt. Im Folgenden bezeichnet  $J_{\mathbf{g}}(\mathbf{x}) = (\partial_j g_i(\mathbf{x}))_{ij}$  die Jacobimatrix von  $\mathbf{g}$ .

**Definition 7.4.** Sei  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  mit  $\mathbf{g} \in C^1$  und sei  $\mathcal{X} = \mathcal{X}(\mathbf{g}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$ . Dann ist der Linearisierungskegel von  $\mathcal{X}$  in  $\bar{\mathbf{x}} \in \mathcal{X}$  zur Darstellung  $\mathcal{X}(\mathbf{g})$  gegeben durch

$$\mathcal{L}_{\mathcal{X}(\mathbf{g})}(\bar{\mathbf{x}}) := \{\mathbf{s} \in \mathbb{R}^n : J_{\mathbf{g}_{\text{eq}(\bar{\mathbf{x}})}}(\bar{\mathbf{x}}) \mathbf{s} \leq \mathbf{0}\}.$$

**Beispiel 7.5.** Sei  $\mathcal{X}(\mathbf{g}) = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$  mit

$$\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \mathbf{x}^\top \mathbf{x} - 1 \\ -x_1 \\ -x_2 \end{pmatrix}.$$

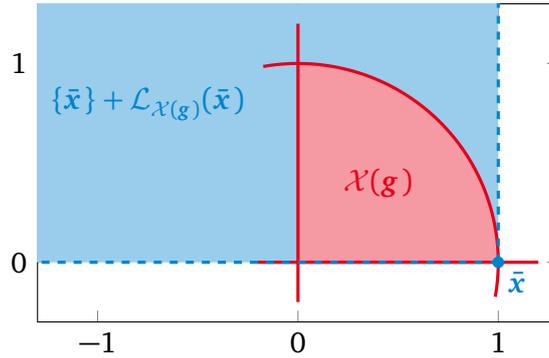


Abbildung 7.1: Linearisierungskegel aus Beispiel 7.5.

Es gilt  $\text{eq}(\bar{x}) = \{1, 3\}$  für  $\bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ , sowie  $\text{eq}(\mathbf{0}) = \{2, 3\}$ ,  $\text{eq}(\frac{1}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix}) = \{1\}$  und  $\text{eq}(\frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}) = \emptyset$ . Weiterhin ist (siehe Abb. 7.1)

$$\mathcal{L}_{\mathcal{X}(g)}(\bar{x}) = \{\mathbf{s} \in \mathbb{R}^2 : \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{s} \leq \mathbf{0}\} = \{\mathbf{s} \in \mathbb{R}^2 : s_1 \leq 0, s_2 \geq 0\}.$$

In diesem Beispiel ist offensichtlich  $\mathcal{Z}(\bar{x}) = \mathcal{L}_{\mathcal{X}(g)}(\bar{x})$ . Beachte dazu, dass  $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathcal{L}_{\mathcal{X}(g)}(\bar{x})$  keine zulässige Richtung ist, aber im Abschluss  $\mathcal{Z}(\bar{x})$  liegt.  $\triangle$

Wir zeigen nun, dass wir den Tangentialkegel für (NL) durch den handlicheren Linearisierungskegel ersetzen können.

**Proposition 7.6.** Für  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$  und  $\bar{x} \in \mathcal{P}(A, \mathbf{b})$  gilt

$$\mathcal{Z}(\bar{x}) = \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x}) = \{\mathbf{s} \in \mathbb{R}^n : A_{\text{eq}(\bar{x})} \cdot \mathbf{s} \leq \mathbf{0}\}.$$

*Beweis.* Mit  $\mathbf{g}(\mathbf{x}) := A\mathbf{x} - \mathbf{b}$  können wir

$$\mathcal{P}(A, \mathbf{b}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$$

schreiben. Per Definition ist der Linearisierungskegel zu dieser Darstellung

$$\begin{aligned} \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x}) &= \{\mathbf{s} \in \mathbb{R}^n : J_{\mathbf{g}_{\text{eq}(\bar{x})}}(\bar{x}) \mathbf{s} \leq \mathbf{0}\} \\ &= \{\mathbf{s} \in \mathbb{R}^n : A_{\text{eq}(\bar{x})} \cdot \mathbf{s} \leq \mathbf{0}\}. \end{aligned}$$

Sei  $\mathbf{s}$  eine zulässige Richtung in  $\bar{x} \in \mathcal{P}(A, \mathbf{b})$ . Dann existiert  $\lambda > 0$  mit

$$A_{\text{eq}(\bar{x})} \cdot (\bar{x} + \lambda \mathbf{s}) \leq \mathbf{b}_{\text{eq}(\bar{x})}.$$

Wegen  $A_{\text{eq}(\bar{x})} \cdot \bar{x} = \mathbf{b}_{\text{eq}(\bar{x})}$  folgt  $A_{\text{eq}(\bar{x})} \cdot \mathbf{s} \leq \mathbf{0}$ . Also ist  $\mathbf{s} \in \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x})$ . Da dies für alle zulässigen Richtungen  $\mathbf{s}$  gilt und  $\mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x})$  abgeschlossen ist, folgt  $\mathcal{Z}(\bar{x}) \subseteq \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x})$ .

Sei nun  $\mathbf{s} \in \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x})$ . Dann gilt

$$A_{\text{eq}(\bar{x})} \cdot (\bar{x} + \lambda \mathbf{s}) = \mathbf{b}_{\text{eq}(\bar{x})} + \lambda A_{\text{eq}(\bar{x})} \cdot \mathbf{s} \leq \mathbf{b}_{\text{eq}(\bar{x})}$$

für alle  $\lambda > 0$ . Wählen wir  $\lambda$  hinreichend klein, gilt weiterhin

$$A_i \cdot (\bar{x} + \lambda \mathbf{s}) < b_i \quad \forall i \in \{1, \dots, m\} \setminus \text{eq}(\bar{x}).$$

Damit ist  $\bar{x} + \lambda \mathbf{s} \in \mathcal{P}(A, \mathbf{b})$  und somit  $\mathbf{s} \in \mathcal{Z}(\bar{x})$ . Es folgt  $\mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x}) \subseteq \mathcal{Z}(\bar{x})$  und insgesamt  $\mathcal{Z}(\bar{x}) = \mathcal{L}_{\mathcal{P}(A, \mathbf{b})}(\bar{x})$ .  $\square$

Insbesondere können wir die Optimalitätsbedingung aus Satz 7.2 damit in eine explizitere Form bringen. Intuitiv darf  $\mathbf{g}_{\text{eq}(\bar{\mathbf{x}})}$  in eine zulässige Richtung nicht wachsen, damit keine Ungleichung verletzt wird.

**Korollar 7.7.** Für jeden lokalen Minimalpunkt  $\bar{\mathbf{x}}$  von (NL) mit  $f \in C^1$  gilt

$$(\nabla f(\bar{\mathbf{x}}))^\top \mathbf{s} \geq 0 \quad \forall \mathbf{s} \in \mathbb{R}^n : A_{\text{eq}(\bar{\mathbf{x}})} \cdot \mathbf{s} \leq \mathbf{0}.$$

*Beweis.* Folgt unmittelbar aus Satz 7.2 und Proposition 7.6. □

*Bemerkung 7.8.* Ist  $\mathbf{g} \in C^1$ , so gilt grundsätzlich  $\mathcal{Z}(\bar{\mathbf{x}}) \subset \mathcal{L}_{\mathcal{X}(\mathbf{g})}(\bar{\mathbf{x}})$  für  $\bar{\mathbf{x}} \in \mathcal{X}(\mathbf{g}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$ . Die Rückrichtung, die es uns erlaubt, den Tangentialkegel in Satz 7.2 durch den Linearisierungskegel zu ersetzen gilt nur unter zusätzlichen Bedingungen, die *Constraint Qualifications* genannt werden. Diese werden in der Vorlesung *Nichtlineare Optimierung* behandelt.

---

## 7.1.2 KKT-Bedingungen

---

Wie im linearen Fall (Satz 4.16), können wir die notwendige Bedingung aus Korollar 7.7 mittels des Farkas-Lemmas weiter vereinfachen.

**Satz 7.9** (KKT-Bedingungen für (NL)). Für jeden lokalen Minimalpunkt  $\bar{\mathbf{x}}$  von (NL) mit  $f \in C^1$  existiert  $\bar{\mathbf{u}} \in \mathbb{R}^m$  mit

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + A^\top \bar{\mathbf{u}} &= \mathbf{0}, && \text{(Multiplikatorregel)} \\ A\bar{\mathbf{x}} &\leq \mathbf{b}, && \text{(Zulässigkeit)} \\ \bar{\mathbf{u}} \geq \mathbf{0} \wedge (A_i \bar{\mathbf{x}} - b_i) \bar{u}_i &= 0 \quad \forall i \in \{1, \dots, m\}. && \text{(Komplementarität)} \end{aligned}$$

*Beweis.* Die Zulässigkeitsbedingung gilt offensichtlich wegen  $\bar{\mathbf{x}} \in \mathcal{P}(A, \mathbf{b})$ .

Nach Korollar 7.7 hat das System

$$A_{\text{eq}(\bar{\mathbf{x}})} \cdot \mathbf{s} \leq \mathbf{0}, \quad (\nabla f(\bar{\mathbf{x}}))^\top \mathbf{s} < 0$$

keine Lösung  $\mathbf{s} \in \mathbb{R}^n$ . Nach dem Farkas-Lemma (Satz 4.13) existiert also  $\mathbf{z} \geq \mathbf{0}$  mit

$$-A_{\text{eq}(\bar{\mathbf{x}})}^\top \mathbf{z} = \nabla f(\bar{\mathbf{x}}).$$

Dann erfüllt  $\bar{\mathbf{u}} \in \mathbb{R}^m$  gegeben durch  $\text{supp}(\bar{\mathbf{u}}) = \text{eq}(\bar{\mathbf{x}})$  und  $\bar{\mathbf{u}}_{\text{eq}(\bar{\mathbf{x}})} = \mathbf{z}$  die Multiplikatorregel sowie die Komplementaritätsbedingung. □

Geometrisch interpretiert, besagen die KKT-Bedingungen, dass der negative Gradient der Zielfunktion im Kegel liegt, der von den Gradienten der in  $\bar{\mathbf{x}}$  aktiven Nebenbedingungen aufgespannt wird.

Wir erweitern die KKT-Bedingungen für gemischte (Un)gleichungssysteme.

**Korollar 7.10.** Für jeden lokalen Minimalpunkt  $\bar{x}$  von

$$\min f(x) \quad \text{s.t.} \quad Ax \leq b, Cx = d$$

mit  $f \in C^1$  existieren dimensionsverträgliche  $\bar{u}, \bar{v}$  mit

$$\begin{aligned} \nabla f(\bar{x}) + A^\top \bar{u} + C^\top \bar{v} &= \mathbf{0}, & (\text{Multiplikatorregel}) \\ A\bar{x} \leq b \wedge C\bar{x} &= d, & (\text{Zulässigkeit}) \\ \bar{u} \geq \mathbf{0} \wedge (A_i \bar{x} - b_i) \bar{u}_i &= 0 \quad \forall i \in \{1, \dots, m\}. & (\text{Komplementarität}) \end{aligned}$$

*Beweis.* Zulässigkeit gilt offensichtlich, da  $\bar{x}$  Minimalpunkt ist. Wir können die Nebenbedingungen umformen zu  $Dx \leq \bar{d}$  mit

$$D = \begin{pmatrix} A \\ C \\ -C \end{pmatrix}, \quad \bar{d} = \begin{pmatrix} b \\ d \\ -d \end{pmatrix}.$$

Dann liefert Satz 7.9 Lagrange-Multiplikatoren  $\bar{u}, \bar{v}^+, \bar{v}^- \geq \mathbf{0}$  mit

$$\begin{aligned} \nabla f(\bar{x}) + A^\top \bar{u} + C^\top (\bar{v}^+ - \bar{v}^-) &= \mathbf{0} \\ (A_i \bar{x} - b_i) \bar{u}_i &= 0, & \forall i \\ (C_j \bar{x} - d_j) \bar{v}_j^+ &= 0, & \forall j \\ (d_\ell - C_\ell \bar{x}) \bar{v}_\ell^- &= 0, & \forall \ell \end{aligned}$$

wobei die letzten zwei Zeilen wegen  $C\bar{x} = d$  ohnehin gelten. Mit  $\bar{v} = \bar{v}^+ - \bar{v}^-$  folgt die Aussage.  $\square$

**Definition 7.11.** Erfüllt  $\bar{x} \in \mathbb{R}^n$  die KKT-Bedingungen aus Korollar 7.10 mit  $\bar{u}, \bar{v}$ , so heißt  $\bar{x}$  *KKT-Punkt* und die Komponenten von  $\bar{u}$  und  $\bar{v}$  heißen *Lagrange-Multiplikatoren*.

Im konvexen Fall sind die KKT-Bedingungen auch hinreichend.

**Satz 7.12** (KKT-Bedingungen für (NL), konvexer Fall). Sei  $f \in C^1$  konvex. Dann ist  $\bar{x} \in \mathbb{R}^n$  genau dann globaler Minimalpunkt von

$$\min f(x) \quad \text{s.t.} \quad Ax \leq b, Cx = d,$$

wenn  $\bar{x}$  KKT-Punkt ist.

*Beweis.* Die Notwendigkeit der KKT-Bedingungen haben wir bereits in Korollar 7.10 gezeigt. Erfülle nun  $\bar{x} \in \mathbb{R}^n$  die KKT-Bedingungen mit Lagrange-Multiplikatoren  $\bar{u}, \bar{v}$ . Nach Satz 2.36 gilt für alle  $z \in \{x \in \mathbb{R}^n : Ax \leq b, Cx = d\}$ , dass

$$\begin{aligned} f(z) - f(\bar{x}) &\geq (\nabla f(\bar{x}))^\top (z - \bar{x}) & (f \text{ konvex}) \\ &= -\bar{u}^\top A(z - \bar{x}) - \bar{v}^\top C(z - \bar{x}) & (\text{Multiplikatorregel}) \\ &= -\bar{u}^\top (Az - b) + \bar{u}^\top (A\bar{x} - b) & (\text{Zulässigkeit von } z) \\ &= -\bar{u}^\top (Az - b) & (\text{Komplementarität}) \\ &\geq 0. & (\text{Zulässigkeit von } z \text{ und } \bar{u} \geq \mathbf{0}) \end{aligned}$$

---

Es folgt  $f(\bar{x}) \leq f(z)$ , also ist  $\bar{x}$  globaler Minimalpunkt. □

*Bemerkung 7.13.* Die KKT-Bedingungen bleiben hinreichend für

$$\min f(x) \quad \text{s.t.} \quad g(x) \leq 0,$$

sofern  $f, g \in C^1$  und konvex (siehe *Nichtlineare Optimierung*).

---

## 7.2 Quadratische Probleme

---

Wir betrachten nun die wichtige Klasse der Quadratischen Optimierungsprobleme (QPs), also Optimierungsproblem der Form

$$\begin{aligned} \text{(QP)} \quad & \min \quad \frac{1}{2} \mathbf{x}^\top Q \mathbf{x} + \mathbf{c}^\top \mathbf{x} \\ & \text{s.t.} \quad A \mathbf{x} \leq \mathbf{b}, \\ & \quad \quad C \mathbf{x} = \mathbf{d}, \end{aligned}$$

mit  $Q \in \mathbb{R}^{n \times n}$  symmetrisch,  $A \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $C \in \mathbb{R}^{k \times n}$ ,  $\mathbf{d} \in \mathbb{R}^k$ .

Wir konzentrieren uns zunächst auf den Fall  $m = 0$ , also, dass nur Gleichungsrestriktionen vorliegen und verallgemeinern unser Vorgehen dann auf den allgemeinen Fall.

---

### 7.2.1 Gleichungsrestriktionen

---

Ist  $m = 0$ , so sind die Nebenbedingungen durch ein Gleichungssystem beschrieben, das sich direkt lösen lässt. Insbesondere lassen sich KKT-Punkte einfach bestimmen. Beachte, dass es entscheidend ist, dass  $\mathbf{x}$  nicht vorzeichenbeschränkt ist, im Gegensatz zu linearen Optimierungsproblemen in Standardform.

**Satz 7.14.** Für  $m = 0$  ist  $\bar{x} \in \mathbb{R}^n$  ein KKT-Punkt von (QP) mit Lagrange-Multiplikatoren  $\bar{v} \in \mathbb{R}^k$  genau dann, wenn

$$\begin{pmatrix} Q & C^\top \\ C & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{v} \end{pmatrix} = \begin{pmatrix} -\mathbf{c} \\ \mathbf{d} \end{pmatrix}.$$

*Beweis.* Für die Zielfunktion  $f(x) = \frac{1}{2} \mathbf{x}^\top Q \mathbf{x} + \mathbf{c}^\top \mathbf{x}$  gilt  $\nabla f(x) = Qx + c$ . Weiter entfallen wegen  $m = 0$  die Komplementaritätsbedingungen aus Korollar 7.10. Also ist  $\bar{x} \in \mathbb{R}^n$  genau dann ein KKT-Punkt von (QP) mit Lagrange-Multiplikatoren  $\bar{v} \in \mathbb{R}^k$ , wenn

$$\begin{aligned} Q\bar{x} + c + C^\top \bar{v} &= 0, \\ C\bar{x} &= d, \end{aligned}$$

was äquivalent zur Aussage ist. □

Ist die Zielfunktion konvex, haben wir damit bereits ein einfaches Lösungsverfahren identifiziert.

**Korollar 7.15.** Für  $m = 0$  und  $Q$  positiv semi-definit lässt sich (QP) auf ein lineares Gleichungssystem zurück führen.

*Beweis.* Für  $Q$  positiv semi-definit, ist die Zielfunktion von (QP) konvex (siehe Satz 2.37 bzw. Beispiel 2.39). Damit sind die KKT-Bedingungen nach Satz 7.12 notwendig und hinreichend für die Optimalität von  $\bar{x}$ . Also können wir ein solches Optimum nach Satz 7.14 durch Lösung eines Gleichungssystems bestimmen.  $\square$

Wir bringen Satz 7.14 in eine Form, die im Folgenden nützlich sein wird.

**Satz 7.16.** Sei  $m = 0$ ,  $f(x) := \frac{1}{2}x^T Qx + c^T x$  und  $\bar{x}_0 \in \mathbb{R}^n$  zulässig für (QP). Es ist  $\bar{x} \in \mathbb{R}^n$  genau dann ein KKT-Punkt von (QP) mit Lagrange-Multiplikatoren  $\bar{v} \in \mathbb{R}^k$ , wenn für  $\bar{x}_\delta = \bar{x} - \bar{x}_0$  gilt, dass

$$\begin{pmatrix} Q & c^T \\ c & 0 \end{pmatrix} \begin{pmatrix} \bar{x}_\delta \\ \bar{v} \end{pmatrix} = \begin{pmatrix} -\nabla f(\bar{x}_0) \\ \mathbf{0} \end{pmatrix}.$$

*Beweis.* Nach Satz 7.14 ist  $\bar{x}$  genau dann ein KKT-Punkt von (QP) mit Lagrange-Multiplikatoren  $\bar{v} \in \mathbb{R}^k$ , wenn

$$\begin{pmatrix} -c \\ d \end{pmatrix} = \begin{pmatrix} Q & c^T \\ c & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{v} \end{pmatrix} = \begin{pmatrix} Q & c^T \\ c & 0 \end{pmatrix} \begin{pmatrix} \bar{x}_0 + \bar{x}_\delta \\ \bar{v} \end{pmatrix},$$

also wenn

$$\begin{pmatrix} Q & c^T \\ c & 0 \end{pmatrix} \begin{pmatrix} \bar{x}_\delta \\ \bar{v} \end{pmatrix} = \begin{pmatrix} -c \\ d \end{pmatrix} - \begin{pmatrix} Q & c^T \\ c & 0 \end{pmatrix} \begin{pmatrix} \bar{x}_0 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -c - Q\bar{x}_0 \\ d - c\bar{x}_0 \end{pmatrix} = \begin{pmatrix} -\nabla f(\bar{x}_0) \\ \mathbf{0} \end{pmatrix},$$

wobei wir im letzten Schritt die Zulässigkeit von  $\bar{x}_0$  genutzt haben.  $\square$

---

## 7.2.2 Strategie der aktiven Menge

---

Wir stellen nun ein Lösungsverfahren für strikt konvexe Quadratische Programme der Form (QP) vor, also für den Fall, dass  $Q$  positiv definit ist (Korollar 2.38). Nach Satz 7.12 genügt es, den eindeutigen (Korollar 2.43) KKT-Punkt zu bestimmen. Wir haben in der vorigen Sektion gesehen, wie das für Gleichungsrestriktionen möglich ist. Wäre nun die Menge  $I \subseteq \{1, \dots, m\}$  der aktiven Ungleichungen in einer optimalen Lösung bekannt, so können wir (QP) auf das gleichungsrestringierte Teilproblem

$$\begin{aligned} \text{(QP')} \quad & \min f(x) := \frac{1}{2}x^T Qx + c^T x \\ & \text{s.t. } A_I x = b_I, \\ & \quad Cx = d \end{aligned}$$

reduzieren.

**Beobachtung 7.17.** Sei  $x^*$  der eindeutige KKT-Punkt von (QP). Dann ist  $x^*$  auch der eindeutige KKT-Punkt von (QP') für  $I = \{i \in \{1, \dots, m\} : A_i x^* = b_i\}$ .

*Beweis.* Die Eindeutigkeit der KKT-Punkte folgt aus Satz 7.12 und den Korollaren 2.38 und 2.43. Sei  $\bar{\mathbf{x}} \neq \mathbf{x}^*$  der KKT-Punkt von (QP'). Da  $\mathbf{x}^*$  zulässig für (QP') ist, folgt  $f(\bar{\mathbf{x}}) < f(\mathbf{x}^*)$ . Wegen  $A_I \bar{\mathbf{x}} = \mathbf{b}_I = A_I \mathbf{x}^*$  gilt  $A_I(\mathbf{x}^* + \lambda(\bar{\mathbf{x}} - \mathbf{x}^*)) = \mathbf{b}_I$  für alle  $\lambda \in [0, 1]$ . Sei  $\lambda \in [0, 1)$  maximal gewählt, so dass  $\mathbf{z} := \mathbf{x}^* + \lambda(\bar{\mathbf{x}} - \mathbf{x}^*) \in \mathcal{P}(A, \mathbf{b})$ . Dann ist  $\lambda > 0$  und es existiert  $j \notin I$  mit  $A_j \mathbf{z} = b_j$ . Das ist im Widerspruch zur Optimalität von  $\mathbf{x}^*$ , da, wegen Konvexität von  $f$ , gilt

$$f(\mathbf{z}) < \lambda f(\bar{\mathbf{x}}) + (1 - \lambda)f(\mathbf{x}^*) < f(\mathbf{x}^*). \quad \square$$

Nun kennen wir natürlich nicht die Menge  $I$ . Die Idee ist nun (QP') für wechselnde Kandidaten  $I$  zu lösen und dabei sicher zu stellen, dass der jeweils optimale Zielfunktionswert streng monoton fällt und dass wir dabei nur zulässige Lösungen  $\bar{\mathbf{x}}$  von (QP) betrachten. Da es endlich viele Möglichkeiten für die Menge  $I$  gibt muss ein solches Verfahren terminieren, wenn es jeweils gelingt (QP') zu lösen.

Die Strategie zur Lösung von (QP') ist die folgende: Wir lösen das KKT-System für (QP') aus Satz 7.16 für die aktuelle Lösung  $\bar{\mathbf{x}}$ . Wenn wir nicht bereits einen KKT-Punkt von (QP) erhalten, gehen wir so weit in Richtung  $\bar{\mathbf{x}}_\delta$  wie möglich ohne unzulässig zu werden. Ist  $\bar{\mathbf{x}}_\delta = \mathbf{0}$  oder erreichen wir  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta$ , so haben wir (QP') gelöst und müssen eine Zeile aus  $I$  streichen um weiter Fortschritt machen zu können. Ist  $\bar{\mathbf{x}} \neq \mathbf{0}$  und wir erreichen nicht ganz  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta$ , so wird immerhin eine neue Nebenbedingung aktiv, die wir zu  $I$  hinzufügen können.

Das resultierende *Verfahren der aktiven Menge* ist das Folgende. Beachte, dass wir auch mit  $I = \emptyset$  starten könnten, aber unsere Wahl die ersten, trivialen Aktivierungsschritte überspringt.

---

**Algorithmus : ACTIVESET(Q, c, A, b, C, d)**

---

**input:** zulässiges Programm (QP) mit  $Q$  positiv definit und  $\text{rank}(C) = k$

**output:** KKT-Punkt  $\bar{\mathbf{x}}$  mit Lagrange-Multiplikatoren  $\bar{\mathbf{u}}, \bar{\mathbf{v}}$

---

$\bar{\mathbf{x}} \leftarrow \text{PHASEI}\left(\begin{pmatrix} A \\ C \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right)$  (zulässig für (QP))

$I := \arg \max\{|I'| : \text{rank}\left(\begin{pmatrix} A_{I'} \\ C \end{pmatrix}\right) = |I'| + k \wedge A_{I'} \bar{\mathbf{x}} = b_{I'}\}$  (oder  $I \leftarrow \emptyset$ )

**repeat**

$\bar{\mathbf{u}} \leftarrow \mathbf{0}$

$\begin{pmatrix} \bar{\mathbf{x}}_\delta \\ \bar{\mathbf{u}}_I \\ \bar{\mathbf{v}} \end{pmatrix} \leftarrow \begin{pmatrix} Q & A_I^\top & C^\top \\ A_I & \mathbf{0} & \mathbf{0} \\ C & \mathbf{0} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} -\nabla f(\bar{\mathbf{x}}) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}$  (KKT-System für (QP'))

**if**  $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$  erfüllen KKT-Bedingungen für (QP):

**return**  $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$

**else if**  $\bar{\mathbf{x}}_\delta = \mathbf{0}$ :

$q := \{i \in I : \bar{u}_i < 0\}$

$I \leftarrow I \setminus \{q\}$  (Deaktivierung)

**else**

**if**  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta \in \mathcal{P}(A, \mathbf{b})$ :

$\bar{\mathbf{x}} \leftarrow \bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta$

**else**

$r := \arg \min \left\{ \frac{b_i - A_i \bar{\mathbf{x}}}{A_i \bar{\mathbf{x}}_\delta} : i \notin I \wedge A_i \bar{\mathbf{x}}_\delta > 0 \right\}$

$\bar{\mathbf{x}} \leftarrow \bar{\mathbf{x}} + \frac{b_r - A_r \bar{\mathbf{x}}}{A_r \bar{\mathbf{x}}_\delta} \bar{\mathbf{x}}_\delta$

$I \leftarrow I \cup \{r\}$  (Aktivierung)

---

Wir müssen zunächst nachweise, dass das Verfahren wohldefiniert ist. Dazu zeigen wir, dass die folgende Invariante erhalten wird.

**Lemma 7.18.** Während der Ausführung von `ACTIVESET` ist durchweg  $\bar{\mathbf{x}}$  zulässig für (QP) und (QP'). Weiter ist  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta$  ein KKT Punkt von (QP') mit Lagrange-Multiplikatoren  $\bar{\mathbf{u}}_I$  und  $\bar{\mathbf{v}}$ .

*Beweis.* Zulässigkeit von  $\bar{\mathbf{x}}$  wird anfangs explizit sichergestellt. Da sich  $\bar{\mathbf{x}}_\delta$  aus dem KKT-System für (QP') ergibt, gilt  $A_I \bar{\mathbf{x}}_\delta = \mathbf{0}$  und  $C \bar{\mathbf{x}}_\delta = \mathbf{0}$ . Unter Verwendung der bisherigen Zulässigkeit von  $\bar{\mathbf{x}}$ , müssen wir also nur sicherstellen, dass  $A_i \bar{\mathbf{x}} \leq b_i$  für alle  $i \notin I$  erhalten bleibt. Im Fall  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta \in \mathcal{P}(A, \mathbf{b})$  ist das offensichtlich gegeben. Bei der Aktivierung der Zeile  $A_r$ , gilt per Auswahlregel von  $r$  für alle  $i \notin I$  mit  $A_i \bar{\mathbf{x}}_\delta > 0$ , dass

$$A_i \cdot \left( \bar{\mathbf{x}} + \frac{b_r - A_r \bar{\mathbf{x}}}{A_r \bar{\mathbf{x}}_\delta} \bar{\mathbf{x}}_\delta \right) \leq A_i \bar{\mathbf{x}} + \frac{b_i - A_i \bar{\mathbf{x}}}{A_i \bar{\mathbf{x}}_\delta} A_i \bar{\mathbf{x}}_\delta = b_i.$$

Insbesondere gilt Gleichheit für die Zeile  $A_r$ . Für alle  $i \notin I$  mit  $A_i \bar{\mathbf{x}}_\delta \leq 0$  gilt wegen  $\frac{b_r - A_r \bar{\mathbf{x}}}{A_r \bar{\mathbf{x}}_\delta} \geq 0$ , dass

$$A_i \cdot \left( \bar{\mathbf{x}} + \frac{b_r - A_r \bar{\mathbf{x}}}{A_r \bar{\mathbf{x}}_\delta} \bar{\mathbf{x}}_\delta \right) \leq A_i \bar{\mathbf{x}} \leq b_i.$$

Also bleibt  $\bar{\mathbf{x}}$  zulässig für (QP).

Wegen  $A_r \bar{\mathbf{x}} = b_r$  nach Aktualisierung von  $\bar{\mathbf{x}}$  gilt jederzeit  $A_I \bar{\mathbf{x}} = \mathbf{b}_I$ . Mit Zulässigkeit von  $\bar{\mathbf{x}}$  für (QP) folgt Zulässigkeit für (QP'). Mit Satz 7.16 angewendet auf (QP') ergibt sich also, dass  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta$  ein KKT Punkt von (QP') mit Lagrange-Multiplikatoren  $\bar{\mathbf{u}}_I$  und  $\bar{\mathbf{v}}$ .  $\square$

Damit können wir folgern, dass die Auswahl von  $q$  und  $r$  jeweils wohldefiniert ist.

**Lemma 7.19.** Bei der Auswahl von  $q$  und  $r$  in `ACTIVESET` sind die entsprechenden Mengen nicht leer.

*Beweis.* Die Menge aus der  $q$  bei der Deaktivierung ausgewählt wird ist nicht leer: Sonst gilt  $\bar{\mathbf{u}}_I \geq \mathbf{0}$  und  $\bar{u}_i = 0$  für  $i \notin I$ . Nach Lemma 7.18 und wegen  $\bar{\mathbf{x}}_\delta = \mathbf{0}$  hätten also  $\bar{\mathbf{x}}$ ,  $\bar{\mathbf{u}}$ ,  $\bar{\mathbf{v}}$  die KKT-Bedingungen für (QP) erfüllen müssen.

Die Menge aus der  $r$  bei der Deaktivierung ausgewählt wird ist ebenfalls nicht leer: Sonst gilt  $A_i \bar{\mathbf{x}}_\delta \leq 0$ , also  $A_i (\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta) \leq b_i$  für alle  $i \notin I$ , da  $\bar{\mathbf{x}}$  zulässig für (QP) ist (Lemma 7.18). Nach Lemma 7.18 ist  $\bar{\mathbf{x}}$  auch zulässig für (QP'), also insbesondere  $A_I \bar{\mathbf{x}} = \mathbf{b}_I$ . Damit gilt wegen  $A_I \bar{\mathbf{x}}_\delta = \mathbf{0}$ , dass  $A_I (\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta) = \mathbf{b}_I$ . Insgesamt hätte also  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta \in \mathcal{P}(A, \mathbf{b})$  gelten müssen.  $\square$

Als letztes zeigen wir noch, dass das KKT-System für (QP') in jeder Iteration eindeutig lösbar ist.

**Lemma 7.20.** In jedem Schritt von `ACTIVESET` ist die Matrix  $K := \begin{pmatrix} Q & A_I^\top & C^\top \\ A_I & 0 & 0 \\ C & 0 & 0 \end{pmatrix}$  invertierbar.

*Beweis.* Der Algorithmus stellt anfangs explizit sicher, dass  $\begin{pmatrix} A_I \\ C \end{pmatrix}$  vollen Zeilenrang hat. Wir zeigen zunächst, dass dann das KKT-System für (QP') eindeutig lösbar ist, also dass die zugehörige Matrix  $K$  invertierbar ist, wenn  $\begin{pmatrix} A_I \\ C \end{pmatrix}$  vollen Zeilenrang hat.

Dazu genügt es zu zeigen, dass das homogene Gleichungssystem  $K \begin{pmatrix} x_\delta \\ u \\ v \end{pmatrix} = \mathbf{0}$  nur die triviale Lösung besitzt (siehe *Lineare Algebra 1*). Für jede Lösung gilt

$$\begin{aligned} Qx_\delta + A_I^\top u + C^\top v &= \mathbf{0}, \\ A_I x_\delta &= \mathbf{0}, \\ Cx_\delta &= \mathbf{0}. \end{aligned}$$

Es folgt

$$0 = x_\delta^\top Qx_\delta + x_\delta^\top A_I^\top u + x_\delta^\top C^\top v = x_\delta^\top Qx_\delta.$$

Wegen positiver Definitheit von  $Q$  ist also  $x_\delta = \mathbf{0}$ . Damit folgt weiter

$$\mathbf{0} = Qx_\delta + A_I^\top u + C^\top v = \begin{pmatrix} A_I^\top & C^\top \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Da die Zeilen von  $\begin{pmatrix} A_I \\ C \end{pmatrix}$ , bzw. die Spalten von  $\begin{pmatrix} A_I^\top & C^\top \end{pmatrix}$ , nach Annahme linear unabhängig sind, folgt  $u = \mathbf{0}$  und  $v = \mathbf{0}$ . Damit ist die einzige Lösung des homogenen Systems  $\mathbf{0}$ , also ist  $K$  invertierbar.

Es bleibt zu zeigen, dass  $\begin{pmatrix} A_I \\ C \end{pmatrix}$  stets vollen Zeilenrang behält. Offensichtlich bleiben die Zeilen von  $A_I$  und  $C$  bei der Deaktivierung einer Zeile linear unabhängig. Nehmen wir nun an,  $A_r$  sei bei der Aktivierung linear abhängig von den Zeilen in  $A_I$  und  $C$ . Dann existieren  $\alpha \in \mathbb{R}^l$  und  $\gamma \in \mathbb{R}^k$  mit

$$A_r = \alpha^\top A_I + \gamma^\top C.$$

Sei  $x_\delta$  Teil der Lösung des KKT-Systems für (QP') in der aktuellen Iteration. Dann ist  $A_I x_\delta = \mathbf{0}$  und  $Cx_\delta = \mathbf{0}$ . Es folgt

$$A_r x_\delta = \alpha^\top A_I x_\delta + \gamma^\top Cx_\delta = \mathbf{0},$$

was ein Widerspruch zur Wahl von  $r$  in ACTIVESET ist. □

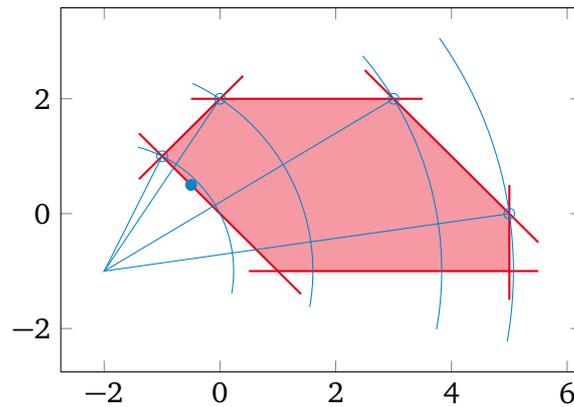
Aus Lemmas 7.18 bis 7.20 folgt, dass ACTIVESET wohldefiniert ist. Offensichtlich liefert das Verfahren ein korrektes Ergebnis, sofern es terminiert.

**Satz 7.21.** ACTIVESET ist wohldefiniert und korrekt.

Wir betrachten ein ausführliches Beispiel.

**Beispiel 7.22.** Wir betrachten das Problem

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{x}^\top \mathbf{x} + 2x_1 + x_2 \\ \text{s.t.} \quad & -x_1 - x_2 \leq 0, \\ & x_2 \leq 2, \\ & x_1 + x_2 \leq 5, \\ & -x_1 + x_2 \leq 2, \\ & x_1 \leq 5, \\ & -x_2 \leq 1, \end{aligned}$$



**Abbildung 7.2:** Illustration des QPs aus Beispiel 7.22. Die Zielfunktion und Zwischenlösungen sind blau dargestellt.

siehe Abb. 7.2. Wir haben also  $m = 6$ ,  $k = 0$ , sowie  $Q = \mathbb{I}$ ,  $\mathbf{c} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$  und  $\nabla f(\mathbf{x}) = Q\mathbf{x} + \mathbf{c} = \mathbf{x} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ .

Wir beginnen mit dem Startpunkt  $\bar{\mathbf{x}} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$  und  $I = \text{eq}(\bar{\mathbf{x}}) = \{3, 5\}$ . Beachte, dass  $A_I = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  bereits vollen Zeilenrang hat. Das KKT-System für (QP) hat die Lösung

$$\begin{pmatrix} \bar{\mathbf{x}}_\delta \\ \bar{\mathbf{u}}_I \end{pmatrix} = \begin{pmatrix} Q & A_I^\top \\ A_I & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \nabla f(\bar{\mathbf{x}}) \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} -7 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & 1 & -2 \end{pmatrix} \begin{pmatrix} -7 \\ -1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1 \\ -6 \end{pmatrix}.$$

Es erfüllen  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{u}} \not\geq \mathbf{0}$  nicht die KKT-Bedingungen und es gilt  $\bar{\mathbf{x}}_\delta = \mathbf{0}$ . Wir führen also einen Deaktivierungsschritt mit  $q = 5$  durch, wodurch die aktive Menge auf  $I = \{3\}$  gesetzt wird.

In der nächsten Iteration lösen wir das KKT-System

$$\begin{pmatrix} \bar{\mathbf{x}}_\delta \\ \bar{\mathbf{u}}_I \end{pmatrix} = \begin{pmatrix} Q & A_I^\top \\ A_I & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \nabla f(\bar{\mathbf{x}}) \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} -7 \\ -1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} -7 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 3 \\ -4 \end{pmatrix}.$$

Wieder erfüllen  $\bar{\mathbf{x}}$  und  $\bar{\mathbf{u}} \not\geq \mathbf{0}$  nicht die KKT-Bedingungen, aber diesmal ist  $\bar{\mathbf{x}}_\delta \neq \mathbf{0}$ . Der Vektor  $\bar{\mathbf{x}} + \bar{\mathbf{x}}_\delta = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$  ist nicht zulässig, wegen  $\bar{x}_2 > 2$ . Wir führen also einen Aktivierungsschritt aus.

Bei der Aktivierung ist  $\{i \notin I : \mathbf{A}_i \cdot \bar{\mathbf{x}}_\delta > 0\} = \{2, 4\}$ , also  $r = 2$  wegen

$$\frac{b_2 - \mathbf{A}_2 \cdot \bar{\mathbf{x}}}{\mathbf{A}_2 \cdot \bar{\mathbf{x}}_\delta} = \min \left\{ \frac{2 - (0 \ 1) \begin{pmatrix} 5 \\ 0 \end{pmatrix}}{(0 \ 1) \begin{pmatrix} -3 \\ 3 \end{pmatrix}}, \frac{2 - (-1 \ 1) \begin{pmatrix} 5 \\ 0 \end{pmatrix}}{(-1 \ 1) \begin{pmatrix} -3 \\ 3 \end{pmatrix}} \right\} = \min \left\{ \frac{2}{3}, \frac{7}{6} \right\} = \frac{2}{3}.$$

Wir setzen entsprechend  $I = \{2, 3\}$  und  $\bar{\mathbf{x}}$  auf  $\bar{\mathbf{x}} + \frac{2}{3} \bar{\mathbf{x}}_\delta = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ .

Die weiteren Iterationen sind in der folgenden Tabelle zusammengefasst. Wir weisen darauf hin, dass  $-\nabla f(\bar{\mathbf{x}}) = -\bar{\mathbf{x}} - \begin{pmatrix} 2 \\ 1 \end{pmatrix}$  und  $\bar{\mathbf{x}}_\delta$ , sowie die resultierende Aktualisierung von  $\bar{\mathbf{x}}$  und  $I$  jeweils zeichnerisch bestimmt werden können.

step	$\bar{x}^\top$	$f(\bar{x})$	$\nabla f(\bar{x})^\top$	$I$	$\bar{x}_\delta^\top$	$\bar{u}^\top$	Bemerkung
1	(5, 0)	22.5	(7, 1)	{3, 5}	(0, 0)	(0, 0, -1, 0, -6, 0)	Deaktivierung
2	(5, 0)	22.5	(7, 1)	{3}	(-3, 3)	(0, 0, -4, 0, 0, 0)	Aktivierung
3	(3, 2)	14.5	(5, 3)	{2, 3}	(0, 0)	(0, 2, -5, 0, 0, 0)	Deaktivierung
4	(3, 2)	14.5	(5, 3)	{2}	(-5, 0)	(0, -3, 0, 0, 0, 0)	Aktivierung
5	(0, 2)	4	(2, 3)	{2, 4}	(0, 0)	(0, -5, 0, 2, 0, 0)	Deaktivierung
6	(0, 2)	4	(2, 3)	{4}	$(-\frac{5}{2}, -\frac{5}{2})$	$(0, 0, 0, -\frac{1}{2}, 0, 0)$	Aktivierung
7	(-1, 1)	0	(1, 2)	{1, 4}	(0, 0)	$(\frac{3}{2}, 0, 0, -\frac{1}{2}, 0, 0)$	Deaktivierung
8	(-1, 1)	0	(1, 2)	{1}	$(\frac{1}{2}, -\frac{1}{2})$	$(\frac{5}{2}, 0, 0, 0, 0, 0)$	zulässig
9	$(-\frac{1}{2}, \frac{1}{2})$	-0.25	$(\frac{3}{2}, \frac{3}{2})$	{1}	(0, 0)	$(\frac{3}{2}, 0, 0, 0, 0, 0)$	KKT-Punkt

△

Die beiden folgenden Resultate weisen nach, dass das Verfahren in gewissem Sinne regelmäßig Fortschritt macht.

**Beobachtung 7.23.** Nach jeweils höchstens  $m$  Iterationen von ACTIVESET ist  $\bar{x}$  eine Optimallösung von (QP').

*Beweis.* Gilt  $\bar{x}_\delta = \mathbf{0}$ , dann ist nach Lemma 7.18  $\bar{x}$  KKT-Punkt von (QP'). Ansonsten folgen so lange Aktivierungsschritte, bis  $\bar{x}_\delta = \mathbf{0}$  oder  $\bar{x} + \bar{x}_\delta \in \mathcal{P}(A, \mathbf{b})$ . Im letzteren Fall wird  $\bar{x}$  auf  $\bar{x} + \bar{x}_\delta$  gesetzt, was wiederum nach Lemma 7.18 ein KKT-Punkt von (QP') ist. Es können natürlich nicht mehr als  $m$  Aktivierungsschritte aufeinander folgen.

Nach Satz 7.12 und wegen  $Q$  positiv definit, bzw.  $f(\mathbf{x})$  strikt konvex (Korollar 2.38), ist KKT-Punkt zu sein hinreichend für die Optimalität von  $\bar{x}$  für (QP'). □

Insbesondere werden nur Verbesserungsrichtungen gewählt, also  $\bar{x}_\delta$  mit negativer Richtungsableitung in  $\bar{x}$ .

**Proposition 7.24.** Ist in ACTIVESET  $\bar{x}_\delta \neq \mathbf{0}$ , dann gilt

$$\nabla f(\bar{x})^\top \bar{x}_\delta < 0.$$

*Beweis.* Da  $\begin{pmatrix} \bar{x}_\delta \\ \bar{u}_I \\ \bar{v} \end{pmatrix}$  Lösung des KKT-Systems für (QP') ist, gilt

$$Q\bar{x}_\delta + A_I^\top \bar{u} + C^\top \bar{v} = -\nabla f(\bar{x}), \quad A_I \bar{x}_\delta = \mathbf{0}, \quad C \bar{x}_\delta = \mathbf{0},$$

also

$$-\nabla f(\bar{x})^\top \bar{x}_\delta = \bar{x}_\delta^\top Q \bar{x}_\delta + \bar{x}_\delta^\top A_I^\top \bar{u} + \bar{x}_\delta^\top C^\top \bar{v} = \bar{x}_\delta^\top Q \bar{x}_\delta.$$

Da  $Q$  positiv definit ist, folgt für  $\bar{x}_\delta \neq \mathbf{0}$ , dass

$$\nabla f(\bar{x})^\top \bar{x}_\delta < 0. \quad \square$$

---

Da durchaus  $b_i - A_i \bar{x} = 0$  für  $i \notin I$  gelten kann, kann es im Aktivierungsschritt dazu kommen, dass sich  $\bar{x}$  nicht ändert. Daher kann es passieren, dass ACTIVESET ewig zwischen Aktivierungs- und Deaktivierungsschritten „kreiselt“, ohne dass sich die Lösung verändert. Dies tritt allerdings in der Praxis selten auf. Tritt kein Kreiseln auf, so erreicht der Algorithmus nach Beobachtung 7.23 immer wieder Optimallösungen von (QP') zu verschiedenen aktiven Mengen  $I$ . Der Zielfunktionswert ist dabei nach Proposition 7.24 streng monoton fallend. Da nur endlich viele verschiedene Mengen  $I$  möglich sind, ist in diesem Fall die Terminierung garantiert.

*Bemerkung 7.25.* Man kann zeigen, dass QPs im Allgemeinen NP-schwer zu lösen sind, aber in polynomieller Laufzeit für positiv definites  $Q$  auf Basis der Ellipsoidmethode oder Innerer-Punkte-Verfahren.

*Bemerkung 7.26.* Die Strategie der aktiven Menge überträgt sich auch auf nichtquadratische Zielfunktionen. Dort treten anstelle von (QP') dann nichtlineare Teilprobleme auf (siehe Vorlesung *Nichtlineare Optimierung*).