

OPTIMIERUNG I

VORLESUNGSSKRIPT, SOMMERSEMESTER 2014

Christian Clason

Stand vom 15. Februar 2016

Fakultät für Mathematik
Universität Duisburg-Essen

INHALTSVERZEICHNIS

I GRUNDLAGEN

- 1 THEORIE DER LINEAREN UNGLEICHUNGEN 5
 - 1.1 Polyeder und ihre Darstellungen 5
 - 1.2 Fourier–Motzkin-Elimination 8
 - 1.3 Das Farkas-Lemma und Trennungssätze 13
- 2 FUNDAMENTALSATZ DER LINEAREN OPTIMIERUNG 16
 - 2.1 Dualität 16
 - 2.2 Komplementarität 20
- 3 GEOMETRIE DER POLYEDER 24

II ANWENDUNGSBEISPIELE

- 4 SPIELTHEORIE 31
- 5 DISKRETE TOMOGRAPHIE 38
- 6 NETZWERKPROBLEME 43

III NUMERISCHE VERFAHREN

- 7 DAS SIMPLEX-VERFAHREN 53
 - 7.1 Herleitung des Verfahrens 54
 - 7.2 Finden eines Startvektors 58
 - 7.3 Vermeidung von Zyklen 61
- 8 DAS DUALE SIMPLEX-VERFAHREN 65

9	PRIMAL-DUALE VERFAHREN	69
9.1	Das primal-duale Simplex-Verfahren	69
9.2	Kombinatorische primal-duale Algorithmen	72
10	GANZZAHLIGE LINEARE OPTIMIERUNG	76
10.1	Schnittebenen-Verfahren	77
10.2	Branch-and-Bound-Verfahren	80

EINLEITUNG

Die mathematische Optimierung beschäftigt sich mit der Aufgabe, Minima bzw. Maxima von Funktionen zu bestimmen. Konkret seien eine Menge X , eine (nicht notwendigerweise echte) Teilmenge $U \subset X$ und eine Funktion $f : X \rightarrow \mathbb{R}$ gegeben. Gesucht ist ein $\bar{x} \in U$ mit

$$f(\bar{x}) \leq f(x) \quad \text{für alle } x \in U,$$

geschrieben

$$f(\bar{x}) = \min_{x \in U} f(x).$$

Man nennt U *zulässige Menge* und einen Punkt $x \in U$ *zulässigen Punkt*; die Forderung $\bar{x} \in U$ wird *Nebenbedingung* genannt. Der optimale Wert $f(\bar{x})$ wird als *Minimum* bezeichnet, \bar{x} selber als *Minimierer*, geschrieben $\bar{x} = \arg \min_{x \in U} f(x)$. Analog spricht man von *Maximum* und *Maximierer*, wenn $f(\bar{x}) \geq f(x)$ für alle $x \in U$ ist. Da gilt

$$\max_{x \in U} f(x) = - \min_{x \in U} -f(x),$$

werden wir in der Regel Minimierer suchen, können aber, wenn es bequemer ist, auch das äquivalente Maximierungsproblem betrachten.

Die Fragen, die wir uns dabei stellen müssen, sind:

1. Hat dieses Problem eine Lösung? Insbesondere: Ist U nichtleer, gibt es also überhaupt zulässige Punkte?
2. Gibt es eine intrinsische Charakterisierung von \bar{x} , d. h. ohne Vergleich mit allen anderen $x \in U$?
3. Wie kann dieses \bar{x} (effizient) berechnet werden?

Aus der Vielzahl der möglichen Beispiele sollen nur kurz folgende erwähnt werden:

1. In *Transport- und Produktionsproblemen* sollen Kosten für Transport minimiert bzw. Gewinn aus Produktion maximiert werden. Dabei beschreibt $x \in \mathbb{R}^n$ die Menge der zu transportierenden bzw. produzierenden verschiedenen Güter und $f(x)$ die dafür nötigen Kosten bzw. aus dem Verkauf erzielten Gewinne. Die Nebenbedingung $x \in U$ beschreibt dabei, dass ein Mindestbedarf gedeckt werden muss bzw. nur endlich viele Rohstoffe zur Produktion zur Verfügung stehen.
2. In *inversen Problemen* sucht man einen Parameter u (zum Beispiel Röntgenabsorption von Gewebe in der Computertomographie), hat aber nur eine (gestörte) Messung y^δ zur Verfügung. Ist ein Modell bekannt, das für gegebenen Parameter u die entsprechende Messung $y = Ku$ liefert, so kann man den unbekannt Parameter rekonstruieren, indem man das Problem

$$\min_{u \in U} \|Ku - y^\delta\|^2 + \alpha \|u\|^2$$

für geeignet gewählte Normen und $\alpha > 0$ löst. Die Menge U kann dabei bekannte Einschränkungen an den Parameter (z. B. Positivität) beschreiben.

3. In der *optimalen Steuerung* ist man zum Beispiel daran interessiert, ein Auto oder eine Raumsonde möglichst effizient von einem Punkt x_0 zu einem anderen Punkt x_1 zu steuern. Beschreibt $x(t) \in \mathbb{R}^3$ die Position zum Zeitpunkt $t \in [0, T]$, so gehorcht sie der Differenzialgleichung

$$(1) \quad \begin{cases} x'(t) = f(t, x(t), u(t)), \\ x(0) = x_0, \end{cases}$$

wobei $u(t)$ die Rolle der Steuerung spielt. Will man dabei den Treibstoffverbrauch (der proportional zu $|u(t)|$ ist) minimieren, führt das auf das Problem

$$\min_{(x,u) \in U} \int_0^T |u(t)| \quad \text{mit} \quad U = \{(x, u) : (1) \text{ ist erfüllt und } x(T) = x_1\}.$$

In dieser Vorlesung behandeln wir *lineare Optimierungsprobleme*, in denen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ linear ist und U ein Polyeder, d. h. durch endlich viele lineare Ungleichungen beschrieben werden kann. Allgemein wird solch ein Problem beschrieben durch einen Vektor $c \in \mathbb{R}^n$, eine Matrix $A \in \mathbb{R}^{m \times n}$ und einen Vektor $b \in \mathbb{R}^m$. Wir suchen dann $x \in \mathbb{R}^n$ als Lösung von

$$(LP) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax \leq b. \end{cases}$$

Solche Probleme tauchen sehr häufig in Wirtschaft und Finanzen auf (siehe Beispiel 1); aber auch viele nichtlineare Probleme können auf lineare Probleme zurückgeführt werden. (Vergleiche die zentrale Rolle von linearen Gleichungssystemem in der numerischen Mathematik.)

Dieses Skriptum basiert vor allem auf den folgenden Werken:

- [1] C. Geiger und C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, Berlin. DOI: [10.1007/978-3-642-56004-0](https://doi.org/10.1007/978-3-642-56004-0)
- [2] P. Gritzmann (2014). *Grundlagen der Mathematischen Optimierung*. Springer, Berlin. DOI: [10.1007/978-3-8348-2011-2](https://doi.org/10.1007/978-3-8348-2011-2)
- [3] M. Grötschel (2010). „Lineare und Ganzzahlige Programmierung (ADM II)“. Vorlesungsskript, Institut für Mathematik, Technische Universität Berlin. URL: <http://www3.math.tu-berlin.de/Vorlesungen/WS09/LinOpt/index.de.html>
- [4] C. Kanzow (2010). „Operations Research“. Vorlesungsskript, Institut für Mathematik, Universität Würzburg. URL: http://www.mathematik.uni-wuerzburg.de/~kanzow/OR_10.html
- [5] S. Ulbrich (2013). „Optimierung I“. Vorlesungsskript, Institut für Mathematik, Technische Universität Darmstadt. URL: <https://www3.mathematik.tu-darmstadt.de/index.php?id=84&evsver=1007&evsid=23&>
- [6] R. Schultz (2013). „Optimierung 1“. Vorlesungsskript, Fakultät für Mathematik, Universität Duisburg-Essen
- [7] A. Schrijver (1986). *Theory of linear and integer programming*. John Wiley & Sons, Ltd., Chichester

Teil I

GRUNDLAGEN

THEORIE DER LINEAREN UNGLEICHUNGEN

Es ist klar, dass die Schwierigkeit an dem linearen Optimierungsproblem (LP) in der Nebenbedingung $Ax \leq b$ liegt. Inhalt dieses Kapitels ist daher, die Lösungsmenge von Systemen linearer Ungleichungen zu charakterisieren; das zentrale Resultat wird ein Verfahren sein, mit dem man entscheiden kann, wann diese Menge leer ist.



1.1 POLYEDER UND IHRE DARSTELLUNGEN

Wir beginnen mit einigen elementaren Notationen und Eigenschaften. Wir bezeichnen üblicherweise Skalare in \mathbb{R} mit griechischen Buchstaben und Vektoren in \mathbb{R}^n mit kleinen Buchstaben, z. B.

$$c = (c_1, \dots, c_n)^T \in \mathbb{R}^n.$$

Wir schreiben $c \geq 0$ genau dann, wenn $c_i \geq 0$ für alle $1 \leq i \leq n$. Das Skalarprodukt von $c, d \in \mathbb{R}^n$ ist

$$c^T d = \sum_{i=1}^n c_i d_i \in \mathbb{R}.$$

Zwei spezielle Vektoren sind der j -te *Einheitsvektor*

$$e_j := (0, \dots, 0, \underset{j-1}{1}, \underset{j}{0}, \dots)^T \in \mathbb{R}^n$$

und der *Eins-Vektor*

$$\mathbb{1} := (1, \dots, 1) \in \mathbb{R}^n.$$

Für eine Matrix $A \in \mathbb{R}^{m \times n}$ schreiben wir A_{ij} für den Eintrag in der i -ten Zeile und j -ten Spalte; die i -te Zeile (aufgefasst als Zeilenvektor) bezeichnen wir mit $A_i \in \mathbb{R}^n$. Damit lässt sich die i -te Zeile des linearen Ungleichungssystems $Ax \leq b$ kurz schreiben als $A_i x \leq b_i$.

Eine Teilmenge $G \subset \mathbb{R}^n$ heisst *Hyperebene*, falls

$$G = \{x \in \mathbb{R}^n : a^T x = \alpha\} \quad \text{für ein } a \in \mathbb{R}^n, \alpha \in \mathbb{R},$$

man bezeichnet a dann als *Normalenvektor* zu G . Eine Teilmenge $H \subset \mathbb{R}^n$ heisst (abgeschlossener) *Halbraum*, falls

$$H = \{x \in \mathbb{R}^n : a^T x \leq \alpha\} \quad \text{für ein } a \in \mathbb{R}^n, \alpha \in \mathbb{R}.$$

Eine Teilmenge $P \subset \mathbb{R}^n$ heisst *Polyeder*, falls

$$P = P(A, b) := \{x \in \mathbb{R}^n : Ax \leq b\} \quad \text{für ein } A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m,$$

und *Polytop*, falls P ein beschränkter Polyeder ist, d. h. $P \subset \{x \in \mathbb{R}^n : \|x\| \leq M\}$ für ein $M > 0$.

Offensichtlich ist jeder Halbraum ein Polyeder, aber auch die leere Menge (denn es gilt $\{x \in \mathbb{R}^n : 0^T x \leq -1\} = \emptyset$). Damit ist jeder Polyeder Durchschnitt endlich vieler Halbräume, denn

$$P(A, b) = \bigcap_{i=1}^m \{x \in \mathbb{R}^n : A_i x \leq b_i\}.$$

Beispiel 1.1. Wir betrachten das System

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

dessen Lösungsmenge die Form $P(A, b)$ hat für

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

Jedes Paar (A, b) definiert einen eindeutigen Polyeder, aber nicht umgekehrt: Da man Ungleichungen mit positiven(!) Skalaren multiplizieren und addieren darf, ist

$$P(A, b) = P(A, b) \cap \{x \in \mathbb{R}^n : (\alpha A_i)x \leq \alpha b_i\} \cap \{x \in \mathbb{R}^n : (A_i + A_j)x \leq (b_i + b_j)\}$$

für beliebige $\alpha > 0$ und $i, j \in \{1, \dots, m\}$. Jeder Polyeder hat also unendlich viele verschiedene Darstellungen.

Oft tauchen in linearen Optimierungsproblemen zusätzlich Gleichheitsbedingungen auf; diese kann man aber einfach als weitere Ungleichungen aufnehmen: Es gilt $a^T x = \beta$ genau dann, wenn $\beta \leq a^T x \leq \beta$ gilt, d. h. wir fügen die Ungleichungen

$$\begin{aligned} a^T x &\leq \beta, \\ -a^T x &\leq -\beta \end{aligned}$$

hinzu. Häufig sind auch Vorzeichenbedingungen an x , die wir ebenfalls als zusätzliche Ungleichungen einfügen können.

Lemma 1.2. Die Lösungsmenge des linearen Systems

$$(1.1) \quad \begin{cases} Ax + By \leq c, \\ Cx + Dy = d, \\ x \geq 0 \end{cases}$$

für dimensionsverträgliche Vektoren c, d, x, y und Matrizen A, B, C, D ist ein Polyeder.

Beweis. Setze

$$A' = \begin{pmatrix} A & B \\ C & D \\ -C & -D \\ -I & 0 \end{pmatrix}, \quad b' = \begin{pmatrix} c \\ d \\ -d \\ 0 \end{pmatrix}, \quad x' = \begin{pmatrix} x \\ y \end{pmatrix},$$

dann ist (1.1) äquivalent mit $A'x' \leq b'$. □

Insbesondere ist

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

ein Polyeder, den man wie in Lemma 1.2 in die Form $P(A', b')$ transformieren kann. Umgekehrt kann man jeden Polyeder $P(A, b)$ in die Form $P^=(A', b')$ bringen, indem man folgende Transformationen verwendet:

(i) Erfüllt $x \in \mathbb{R}^n$ die Ungleichung

$$A_i x \leq b_i,$$

so gilt stets

$$A_i x + y_i = b_i \quad \text{mit} \quad y_i = b_i - A_i x \geq 0.$$

Man bezeichnet $y_i \in \mathbb{R}$ als *Schlupfvariable* (englisch „slack variable“), ebenso den Vektor $y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$.

Erfüllt umgekehrt $x' = (x, y_i)$ die Bedingungen

$$\begin{aligned} A_i x + y_i &= b_i, \\ y_i &\geq 0, \end{aligned}$$

so gilt

$$A_i x \leq b_i.$$

(ii) Jedes $x \in \mathbb{R}^n$ kann man schreiben als

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

wobei

$$x_i^+ := \begin{cases} x_i & x_i \geq 0, \\ 0 & x_i < 0, \end{cases} \quad x_i^- = \begin{cases} 0 & x_i \geq 0, \\ -x_i & x_i < 0. \end{cases}$$

Dann ist $P(A, b)$ äquivalent zu $P(A', b')$ mit

$$A' = (A, -A, I), \quad b' = b$$

in dem Sinne, dass für $x \in P(A, b)$ stets $(x^+, x^-, b - Ax)^T \in P(A', b')$ und für $x' := (u, v, w)^T \in P(A', b')$ stets $x := u - v \in P(A, b)$ gilt. Wir können also je nach Bedarf zwischen beiden Formen wechseln.

1.2 FOURIER-MOTZKIN-ELIMINATION

Wir betrachten nun das Problem zu entscheiden, wann ein System linearer Ungleichungen $Ax \leq b$ eine Lösung hat – man sagt: *konsistent* ist – oder äquivalent, wann $P(A, b)$ nicht leer ist. Für lineare Gleichungssysteme kann man diese Frage mit Hilfe der Gauß-Elimination beantworten: Man bringt $Ax = b$ durch elementare Zeilen-Umformungen auf Stufenform; bleiben danach keine widersprüchlichen Gleichungen vom Typ $0 = 1$ übrig, so hat das System eine Lösung, die man durch Rückwärtssubstitution erhalten kann.

Für Ungleichungen kann man so aber nicht vorgehen, denn Multiplikation mit negativen Skalaren ändert die Richtung einer Ungleichung, weshalb auch die Addition zweier Ungleichungen nicht reversibel ist (d. h. die Lösungsmenge ändert). Addieren wir etwa in $Ax \leq b$ aus Beispiel 1.1 die erste Spalte zu den beiden anderen, so erhält man das System $A'x \leq b'$ mit

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

Dieses neue System hat als Lösung $x = (-1, -1)$, welches aber offensichtlich keine Lösung des ursprünglichen Systems ist; es gilt also $P(A', b') \neq P(A, b)$. Allerdings ist jede Lösung des alten Systems $Ax \leq b$ auch Lösung von $A'x' \leq b'$; wäre daher $P(A', b') = \emptyset$, so müsste auch $P(A, b) = \emptyset$ sein. Wir können also versuchen, durch Addition von (positiven!) Vielfachen von Zeilen von $Ax \leq b$ möglichst viele neue Ungleichungen zu erzeugen, von denen wir einfacher erkennen können, dass sie nicht konsistent sind. Dazu erzeugen wir die neuen Ungleichungen so, dass eine Variable x_j nicht mehr vorkommt. (Dies entspricht der Projektion von $P(A, b)$

Algorithmus 1.1 : Fourier–Motzkin-Elimination

Input : $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $j \in \{1, \dots, n\}$

Output : $D \in \mathbb{R}^{r \times n}$, $f \in \mathbb{R}^r$, $r \in \mathbb{N}$

1 Skalieren jede Zeile von $Ax \leq b$ mit $|A_{ij}|^{-1}$ (falls $\neq 0$) $\rightsquigarrow A', b'$ mit $A'_{ij} \in \{0, 1, -1\}$

2 Sortiere $M = \{1, \dots, m\}$ in

$$Z = \{i \in M : A'_{ij} = 0\}$$

$$P = \{i \in M : A'_{ij} = 1\}$$

$$N = \{i \in M : A'_{ij} = -1\}$$

3 Setze $r = |Z \cup (P \times N)| = |Z| + |P| \cdot |N|$ und wähle Bijektion $\pi : \{1, \dots, r\} \rightarrow Z \cup (P \times N)$

4 **for** $i = 1, \dots, r$ **do**

5 **if** $\pi(i) = s \in Z$ **then**

6 | Setze $D_i = A'_s$ und $f_i = b'_s$

7 **if** $\pi(i) = (s, t) \in P \times N$ **then**

8 | Setze $D_i = A'_s + A'_t$ und $f_i = b'_s + b'_t$

auf den Unterraum $\{x \in \mathbb{R}^n : x_j = 0\}$.) Hat man dadurch ein Paar (oder mehrere Paare) von Ungleichungen der Form

$$\begin{aligned} x_k &\leq \beta, \\ -x_k &\leq \alpha, \end{aligned}$$

erzeugt, so hat das System nur dann eine Lösung, wenn $-\alpha \leq \beta$ ist; jede Wahl $x_k \in [-\alpha, \beta]$ ist dann zulässig, und durch Einsetzen in die ursprünglichen Ungleichungen kann man einen zulässigen Punkt x berechnen. Ansonsten wiederholt man die Prozedur, um weitere Unbekannte zu eliminieren. Dieses Verfahren bezeichnet man als *Fourier–Motzkin-Elimination*.¹

Die Projektion von $P(A, b)$ auf den Unterraum $\{x \in \mathbb{R}^n : x_j = 0\}$ kann dabei mit Algorithmus 1.1 berechnet werden. Für die Bijektion geht man zum Beispiel zuerst Z (in aufsteigender Ordnung) und dann $P \times N$ (in lexikographischer Ordnung) durch. Die erzeugte Matrix D kann dabei leer sein, falls $A_{ij} > 0$ (oder $A_{ij} < 0$) für alle $1 \leq i \leq m$ gilt. In diesem Fall sind keine Ungleichungen zu erfüllen, also ist $P(D, f) = \mathbb{R}^n$. Ansonsten hat D in der j -ten Spalte nur Null-Einträge, und wir können die Prozedur mit $j + 1$ wiederholen, bis wir eine Nullmatrix erhalten.

Bevor wir dies beweisen, soll das Vorgehen anhand des Systems aus Beispiel 1.1 illustriert werden.

¹Nach [Jean Baptiste Joseph Fourier](#) (1768–1830), der es 1827 in einem Aufsatz beschrieben hat (und auch als einer der ersten die Verbindung zwischen linearen Ungleichungen, Polyedern, und Optimierung erkannt hat), und [Theodore Samuel Motzkin](#) (1908–1970), der es 1936 in seiner Doktorarbeit wieder entdeckte.

Beispiel 1.3. 1. Wir beginnen mit $j = 1$ und wenden Algorithmus 1.1 an auf

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

Da alle Einträge von A bereits in $\{0, 1, -1\}$ liegen, können wir Schritt 1 überspringen und setzen $Z = \{3\}$ und $P \times N = \{(1, 2)\}$. Schritt 6 liefert dann

$$D_1 = A_3 = (0, -1), \quad f_1 = b_3 = 0,$$

und Schritt 8 liefert

$$D_2 = A_1 + A_2 = (0, 1), \quad f_2 = b_1 + b_2 = 1.$$

Damit erhalten wir

$$D = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \quad f = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Die erste Spalte von $D \in \mathbb{R}^{2 \times 2}$ ist also Null.

2. Wir wenden nun Algorithmus 1.1 für $j = 2$ auf $A = D$ und $b = f$ an. Wieder überspringen wir Schritt 1 und setzen $Z = \emptyset$ und $P \times N = \{(2, 1)\}$. Schritt 8 ergibt dann

$$D_1 = A_1 + A_2 = (0, 0), \quad f_1 = b_1 + b_2 = 1,$$

und damit

$$D = (0, 0), \quad f = (1).$$

Unser neues D ist also die Null-Matrix, und das System ist reduziert auf die Ungleichung $0 \leq 1$. Das System ist also lösbar.

3. Um eine Lösung zu berechnen, betrachten wir das Ergebnis des vorletzten Schritts, nämlich die Ungleichungen

$$\begin{aligned} -x_2 &\leq 0, \\ x_2 &\leq 1. \end{aligned}$$

Jedes $x_2 \in [0, 1]$ erfüllt das System. Wählen wir zum Beispiel $x_2 = \frac{1}{2}$ und setzen dies in das ursprüngliche System ein, erhalten wir

$$\begin{aligned} x_1 &\leq \frac{1}{2}, \\ -x_1 &\leq 0. \end{aligned}$$

Eine Lösung von $Ax \leq b$ ist also $x = (\frac{1}{2}, \frac{1}{2})$.

Im Allgemeinen wird aber nicht $r < m$ wie in diesem Beispiel sein; tatsächlich r kann sogar viel größer als m werden. Ob dies geschieht, hängt von der Reihenfolge ab, in der man die Unbekannten eliminiert.

Wir beweisen nun, dass das beschriebene Vorgehen tatsächlich korrekt ist. Wir betrachten zuerst die Anwendung von Algorithmus 1.1 für ein festes j .

Satz 1.4. Seien A, B, j und D, f, r wie in Algorithmus 1.1 gegeben. Dann gilt:

(i) Die j -te Spalte von D ist der Null-Vektor.

(ii) Jede Zeile von $Dx \leq f$ ist positive Linearkombination von Zeilen von $Ax \leq b$, d. h. für alle $1 \leq i \leq r$ existiert ein $u_{(i)} \in \mathbb{R}^m$ mit $u_{(i)} \geq 0$ und

$$u_{(i)}^T A = d_i, \quad u_{(i)}^T b = f_i.$$

Es gibt also eine Matrix $U \in \mathbb{R}^{r \times m}$ (deren Zeilen die $u_{(i)}$ sind) mit $U \geq 0$ und

$$UA = D, \quad Ud = f.$$

(iii) Für $x \in \mathbb{R}^n$ mit $x_j = 0$ setze

$$\lambda_i = \frac{1}{A_{ij}}(b_i - A_{i \cdot} x), \quad \text{für alle } i \in P \cup N,$$

$$\alpha = \begin{cases} -\infty & N = \emptyset, \\ \max_{i \in N} \lambda_i & N \neq \emptyset, \end{cases}$$

$$\beta = \begin{cases} \infty & P = \emptyset, \\ \min_{i \in P} \lambda_i & P \neq \emptyset. \end{cases}$$

Dann gilt

a) Ist $x \in P(D, f)$, so ist $x + \lambda e_j = (x_1, \dots, x_{j-1}, \lambda, x_{j+1}, \dots, x_n)^T \in P(A, b)$ für alle $\lambda \in [\alpha, \beta] \neq \emptyset$.

b) Ist $x + \lambda e_j \in P(A, b)$, dann ist $\lambda \in [\alpha, \beta]$ und $x \in P(D, f)$.

Beweis. Aussagen (i) und (ii) folgen aus der Konstruktion von D und f ; insbesondere ist für alle $s = \pi(i) \in Z$ der Eintrag $D_{ij} = A_{sj} = 0$, und für $(s, t) = \pi(i) \in P \times N$ ist

$$d_{ij} = a'_{sj} + a'_{tj} = 1 - 1 = 0$$

In beiden Fällen können wir $u_{(i)}$ konkret angeben: Für $\pi(i) = s \in Z$ ist $u_{(i)} = e_s \geq 0$, und für $(s, t) = \pi(i) \in P \times N$ ist $u_{(i)} = |A_{sj}|^{-1} e_s + |A_{tj}|^{-1} e_t \geq 0$.

Bleibt also Aussage (iii). Sei dafür $x \in P(D, f)$ mit $x_j = 0$. Für a) zeigen wir zuerst $\alpha \leq \beta$. Dies gilt trivialerweise, falls N oder P leer sind. Sind N und P dagegen nicht leer, so existiert

ein $(s, t) \in P \times N$ mit $\alpha = \lambda_t$ und $\beta = \lambda_s$ sowie ein $q \in \{1, \dots, r\}$ mit $\pi(q) = (s, t)$. Nach Konstruktion ist dann

$$D_q = A'_s + A'_t = \frac{1}{|A_{sj}|} A_s + \frac{1}{|A_{tj}|} A_t = \frac{1}{A_{sj}} A_s - \frac{1}{A_{tj}} A_t,$$

$$f_q = \frac{1}{A_{sj}} b_s - \frac{1}{A_{tj}} b_t.$$

Für $x \in P(D, f)$ gilt insbesondere $D_q^T x \leq f_q$; daraus folgt nach Einsetzen und Umsortieren die Ungleichung

$$\alpha = \frac{1}{A_{tj}} (b_t - A_t x) \leq \frac{1}{A_{sj}} (b_s - A_s x) = \beta.$$

Sei nun $\lambda \in [\alpha, \beta]$ und $x \in P(D, f)$ mit $x_j = 0$. Wir müssen zeigen, dass $x + \lambda e_j \in P(A, b)$ ist, d. h. dass $A_i(x + \lambda e_j) \leq b_i$ für alle $1 \leq i \leq m$ ist.

1. Fall: $i \in Z$. Dann ist $i = \pi(q)$ mit $D_q = A_i$ und $f_q = b_i$, also gilt wegen $A_{ij} = 0$ die Ungleichung

$$A_i(x + \lambda e_j) = A_i x = D_q x \leq f_q = b_i.$$

2. Fall: $i \in P$. Dann ist $A_{ij} > 0$ und damit gilt nach Definition von β die Ungleichung

$$A_i(x + \lambda e_j) = A_i x + \lambda A_{ij} \leq A_i x + \beta A_{ij} \leq A_i x + \lambda_i A_{ij} = b_i.$$

3. Fall: $i \in N$ geht analog.

Ebenso beweist man die Aussage (iii) b). □

Folgerung 1.5. *Es gilt $P(A, b) = \emptyset$ genau dann, wenn $P(D, f) = \emptyset$ gilt; beziehungsweise $Ax \leq b$ ist konsistent genau dann, wenn $Dx \leq f$ konsistent ist.*

Führt man nun Algorithmus 1.1 sukzessive für $j = 1, \dots, n$ aus, so erhält man eine Folge von Matrizen $D^{(1)}, \dots, D^{(n)}$ und Vektoren $f^{(1)}, \dots, f^{(n)}$. Ist $D^{(j)}$ leer für ein j , so ist $P(D^{(j)}, b^{(j)}) = \mathbb{R}^n$ und damit $Ax \leq b$ konsistent. Ansonsten ist $D^{(n)} = 0$ und daher gilt $Ax \leq b$ genau dann, wenn $0 = D^{(n)}x \leq f^{(n)}$ ist. Nach Satz 1.4 (ii) existiert nun eine Matrix

$$U := U^{(n)} \dots U^{(1)} \in \mathbb{R}^{r^{(n)} \times m}$$

mit

$$U \geq 0, \quad UA = D^{(n)} = 0, \quad f^{(n)} = Ub.$$

Das System $Ax \leq b$ ist also *nicht* konsistent genau dann, wenn $f_i^{(n)} < 0$ für ein $1 \leq i \leq r^{(n)}$ ist. Anders ausgedrückt: Es gibt einen Vektor u (nämlich die entsprechende Zeile u_i von U) mit $u^T A = 0$ und $u^T b < 0$. Wir fassen zusammen:

Folgerung 1.6. Das Ungleichungssystem $Ax \leq b$ für $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$ hat genau dann keine Lösung, wenn ein Vektor $u \in \mathbb{R}^m$ existiert mit

$$u \geq 0, \quad u^T A = 0, \quad u^T b < 0.$$

Mit der Fourier–Motzkin-Elimination haben wir auch ein erstes Verfahren zur Hand, um lineare Optimierungsprobleme der Form (LP) zu lösen. Führen wir als neue Variable $\eta = c^T x$ ein, so kann man Problem (LP) äquivalent schreiben als

$$\begin{cases} \min_{(x, \eta) \in \mathbb{R}^{n+1}} \eta \\ \text{mit } Ax \leq b, \\ c^T x - \eta \leq 0. \end{cases}$$

Wenden wir nun die Fourier–Motzkin-Elimination auf

$$A' = \begin{pmatrix} A & 0 \\ c^T & -1 \end{pmatrix}, \quad b' = \begin{pmatrix} b \\ 0 \end{pmatrix},$$

an um x_1, \dots, x_n zu eliminieren, so erhalten wir ein System von Ungleichungen der Form

$$\begin{aligned} \eta &\leq \beta_i, & i \in P^{(n)}, \\ -\eta &\leq \alpha_i, & i \in N^{(n)}. \end{aligned}$$

Den minimalen Wert erhalten wir dann, indem wir $\eta = \min_{i \in N^{(n)}} (-\alpha_i)$ wählen, und dann durch sukzessives Einsetzen in den Zwischenschritten einen zugehörigen Minimierer $\bar{x} = (x_1, \dots, x_n)^T$ berechnen.

Dieses Verfahren ist allerdings nicht praktikabel; zwar liefert es garantiert nach n Schritten eine Lösung (oder die Information, dass keine existiert), aber die Anzahl der Ungleichungen $r^{(n)}$ kann exponentiell wachsen (etwa wie 2^n), so dass die Schritte nicht mehr in vertretbarer Zeit oder mit vertretbarem Speicheraufwand durchführbar sind.

1.3 DAS FARKAS-LEMMA UND TRENNUNGSSÄTZE

Folgerung 1.6 sieht unscheinbar aus, hat aber weitreichende Auswirkungen. Als Vorgeschmack leiten wir daraus direkt Resultate für Polyeder her, die im allgemeinen Fall (konvexer Mengen) tiefe Werkzeuge der Funktionalanalysis benötigen (nämlich den Satz von Hahn–Banach).

Zuerst formulieren wir die Aussage von Folgerung 1.6 für den „allgemeinen Fall“ aus Lemma 1.2.

Folgerung 1.7 (Farkas-Lemma). Für dimensionsverträgliche Vektoren c, d, x, y, u, v und Matrizen A, B, C, D gilt genau eine der beiden Aussagen

(i) Es existieren x, y mit

$$(1.2) \quad \begin{cases} Ax + By \leq c, \\ Cx + Dy = d, \\ x \geq 0. \end{cases}$$

(ii) Es existieren u, v mit

$$(1.3) \quad \begin{cases} u^T A + v^T C \geq 0, \\ u^T B + v^T D = 0, \\ u \geq 0, \\ u^T c + v^T d < 0. \end{cases}$$

Beweis. Wie im Beweis von Lemma 1.2 schreiben wir (1.2) als $A'x' \leq b'$ mit

$$A' = \begin{pmatrix} A & B \\ C & D \\ -C & -D \\ -I & 0 \end{pmatrix}, \quad b' = \begin{pmatrix} c \\ d \\ -d \\ 0 \end{pmatrix}, \quad x' = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Nach Folgerung 1.6 hat dieses System genau dann keine Lösung (d. h. (1.2) gilt nicht), wenn ein $u' =: (u, v^+, v^-, w) \geq 0$ existiert mit

$$(u')^T A' = 0, \quad (u')^T b' < 0.$$

Ausmultipliziert bedeutet das

$$\begin{cases} u^T A + (v^+)^T C - (v^-)^T C - w^T I = 0, \\ u^T B + (v^+)^T D - (v^-)^T D = 0, \\ u^T c + (v^+)^T d - (v^-)^T d < 0. \end{cases}$$

Setzen wir $v := v^+ - v^-$ und fassen $w \geq 0$ als Schlupfvariable auf, so erhalten wir, dass (1.3) gilt. \square

Durch Spezialisierung erhalten eine Familie von Aussagen.

Folgerung 1.8. Für $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$ gilt jeweils genau eine der Aussagen

- a) Es existiert x mit $Ax \leq b$ oder es existiert $u \geq 0$ mit $u^T A = 0$ und $u^T b < 0$.
- b) Es existiert $x \geq 0$ mit $Ax \leq b$ oder es existiert $u \geq 0$ mit $u^T A \geq 0$ und $u^T b < 0$.
- c) Es existiert $x \geq 0$ mit $Ax = b$ oder es existiert u mit $u^T A \geq 0$ und $u^T b < 0$.
- d) Es existiert x mit $Ax = b$ oder es existiert u mit $u^T A = 0$ und $u^T b < 0$.

Hier sehen wir unser erstes *Dualitätsresultat*: Eine Vorzeichenbedingung für x (bzw. u) taucht genau dann auf, wenn eine Ungleichungsbedingung für u (bzw. x) auftaucht. Fall d) entspricht genau der Rangbedingung aus der Gauß-Elimination: Das auf Zeilen-Stufen-Form transformierte System hat eine Zeile $0 = \beta < 0$.

Daraus erhält man mit vergleichsweise wenig Aufwand Trennsätze für Polyeder.

Satz 1.9 (Strikte Trennung von Polyedern). *Seien $P = P(A, b)$ und $Q = (B, d)$ nichtleere Polyeder in \mathbb{R}^n . Dann sind P und Q strikt getrennt, d. h. es existiert eine Hyperebene $H = \{x \in \mathbb{R}^n : c^T x = \gamma\}$ mit $c \in \mathbb{R}^n \setminus \{0\}$, $P \cup Q \not\subset H$ und*

$$P \subset \{x \in \mathbb{R}^n : c^T x < \gamma\}, \quad Q \subset \{x \in \mathbb{R}^n : c^T x \geq \gamma\},$$

genau dann, wenn $P \cap Q = \emptyset$

Beweis. Sind P und Q strikt getrennt, so ist $P \cap Q = \emptyset$ (sonst gäbe es ein $x \in P \cap Q$ mit $c^T x < \gamma \leq c^T x$). Sei umgekehrt $P \cap Q = \emptyset$. Also existiert kein $x \in \mathbb{R}^n$ mit $Ax \leq b$ und $Bx \leq d$, d. h. das System

$$\begin{pmatrix} A \\ B \end{pmatrix} x \leq \begin{pmatrix} b \\ d \end{pmatrix}$$

hat keine Lösung. Aus Lemma 1.7 folgt daher die Existenz von $(u, v)^T \geq 0$ mit $u^T A + v^T B = 0$ und $u^T b + v^T d < 0$. Setze nun

$$c^T := u^T A = -v^T B \quad \text{und} \quad \gamma := \frac{1}{2}(u^T b - v^T d).$$

Dann gilt wegen $Ax \leq b$ für $x \in P$, dass

$$c^T x = u^T Ax \leq u^T b < u^T b - \frac{1}{2}(u^T b + v^T d) = \gamma$$

und wegen $Bx \leq d$ für $x \in Q$, dass

$$c^T x = -v^T Bx \geq -v^T d > -v^T d + \frac{1}{2}(u^T b + v^T d) = \gamma,$$

d. h. $\{x \in \mathbb{R}^n : c^T x = \gamma\}$ ist eine strikt trennende Hyperebene. □

FUNDAMENTALSATZ DER LINEAREN OPTIMIERUNG

Wir können nun entscheiden, wann die zulässige Menge $P(A, b)$ eines linearen Optimierungsproblems nichtleer ist. Da jeder Polyeder abgeschlossen und das lineare Funktional $c^T x$ stetig ist, kann nur noch schief gehen, dass $c^T x$ auf $P(A, b)$ nach unten unbeschränkt ist, also $\inf_{x \in P(A, b)} c^T x = -\infty$ ist.

2

2.1 DUALITÄT

Wir fragen uns also, ob eine der Ungleichungen aus $Ax \leq b$ erlaubt, eine untere Schranke für $c^T x$ anzugeben. Zunächst gilt für alle $x \in P(A, b)$ und $y \geq 0$, dass

$$-y^T Ax \geq -y^T b$$

ist. Können wir also ein $\bar{y} \geq 0$ finden mit $A^T \bar{y} = -c$, dann ist für alle $x \in P(A, b)$

$$(2.1) \quad c^T x = (-A^T \bar{y})^T x = -\bar{y}^T Ax \geq -\bar{y}^T b = -b^T \bar{y}$$

und wir haben eine untere Schranke gefunden. Damit haben wir auch schon unser erstes Existenzkriterium: Ist $P^*(A^T, -c)$ nichtleer – was wir prinzipiell mit Fourier–Motzkin–Elimination überprüfen können – so hat das Problem

$$(P) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax \leq b. \end{cases}$$

eine Lösung $\bar{x} \in P(A, b)$. Wir können sogar noch mehr sagen: gilt in (2.1) für ein \bar{x} die Gleichheit, so ist \bar{x} Lösung, denn für alle $x \in P(A, b)$ ist dann

$$c^T x \geq -b^T \bar{y} = c^T \bar{x}.$$

Umgekehrt muss dann \bar{y} die beste (d. h. größte) untere Schranke sein, löst also das *duale* Problem

$$(D) \quad \begin{cases} \max_{y \in \mathbb{R}^n} & -b^T y \\ \text{mit} & A^T y = -c, \\ & y \geq 0. \end{cases}$$

Entsprechend wird (P) das *primale* Problem genannt. Diese einfache Beobachtung ist so zentral, dass wir sie festhalten wollen.

Satz 2.1 (Schwache Dualität). *Für alle $x \in P(A, b)$ und $y \in P^=(A^T, -c)$ ist*

$$(2.2) \quad c^T x \geq -b^T y.$$

Gilt für ein $\bar{x} \in P(A, b)$ und ein $\bar{y} \in P^=(A^T, -c)$ Gleichheit, so ist \bar{x} Lösung von (P) und \bar{y} Lösung von (D).

Eine äquivalente Aussage gilt auch für allgemeinere lineare Probleme mit Nebenbedingungen der Form (1.1).

Lemma 2.2. *Für dimensionsverträgliche Vektoren c, d, e, f, x, y und Matrizen A, B, C, D ist das duale Problem zu*

$$(PA) \quad \begin{cases} \min_{x, y} & c^T x + d^T y \\ \text{mit} & Ax + By \geq e, \\ & Cx + Dy = f, \\ & x \geq 0, \end{cases}$$

gegeben durch

$$(DA) \quad \begin{cases} \max_{u, v} & e^T u + f^T v \\ \text{mit} & A^T u + C^T v \leq c, \\ & B^T u + D^T v = d, \\ & u \geq 0. \end{cases}$$

Beweis. Wir schreiben (PA) analog zum Beweis von Lemma 1.2 als $\min_{x' \in P(A', b')} (c')^T x'$ mit

$$x' := \begin{pmatrix} x \\ y \end{pmatrix}, \quad c' := \begin{pmatrix} c \\ d \end{pmatrix}, \quad A' := \begin{pmatrix} -A & -B \\ C & D \\ -C & -D \\ -I & 0 \end{pmatrix}, \quad b' := \begin{pmatrix} -e \\ f \\ -f \\ 0 \end{pmatrix}.$$

Das duale Problem dazu ist dann $\max_{y' \in P^=((A')^T, -c')} -(b')^T y'$ mit $y' = (u, v^+, v^-, w)^T$. Einsetzen von $v := -(v^+ - v^-)$ und Auffassen von w als Schlupfvariable wie im Beweis von Folgerung 1.7 ergibt dann (DA). \square

Damit erkennt man auch, dass das duale Problem zu (DA) wieder (PA) ist. Wie im Farkas-Lemma gehört zu jeder primalen Nebenbedingung eine duale Variable, die nicht-negativ ist genau dann, wenn es sich um eine Ungleichungsnebenbedingung handelt.

Das zentrale Resultat dieses Kapitels – daher auch als *Fundamentalsatz der linearen Optimierung* bezeichnet – ist die Tatsache, dass Gleichheit in (2.1) immer gilt, solange die beiden zulässigen Mengen nichtleer sind.

Satz 2.3 (Starke Dualität). *Beide Probleme (P) und (D) haben eine Lösung \bar{x} bzw. \bar{y} genau dann, wenn die zulässigen Mengen $P(A, b)$ bzw. $P^=(A^T, -c)$ nicht leer sind. In diesem Fall gilt*

$$c^T \bar{x} = -b^T \bar{y}.$$

Beweis. Lösungen müssen natürlich zulässig sein. Seien also $P(A, b)$ und $P^=(A^T, -c)$ nicht leer. Nach Satz 2.1 genügt es zu zeigen, dass zulässige Punkte \bar{x} und \bar{y} mit $c^T \bar{x} \leq -b^T \bar{y}$ existieren, d. h. dass das System

$$(2.3) \quad \begin{cases} \begin{pmatrix} 0 \\ b^T \end{pmatrix} y + \begin{pmatrix} A \\ c^T \end{pmatrix} x \leq \begin{pmatrix} b \\ 0 \end{pmatrix}, \\ A^T y = -c, \\ y \geq 0, \end{cases}$$

eine Lösung hat. Dies ist nach Folgerung 1.7 äquivalent dazu, dass für $u = (w, \gamma) \geq 0$ und v das System

$$(2.4) \quad \begin{cases} \gamma b^T + v^T A^T \geq 0, \\ w^T A + \gamma c^T = 0, \\ w^T b - v^T c < 0, \end{cases}$$

keine Lösung hat. Dies zeigen wir durch Widerspruch. Wir nehmen an, (2.4) hat eine Lösung, und machen eine Fallunterscheidung nach $\gamma \in \mathbb{R}$.

1. Fall: $\gamma = 0$. Einsetzen in (2.4) und Anwenden von Folgerung 1.7 „rückwärts“ auf die ersten beiden (Un-)Gleichungen ergibt dann, dass das System

$$\begin{cases} Ax \leq b, \\ A^T y = -c, \\ y \geq 0. \end{cases}$$

keine Lösung hat. Dafür müsste aber entweder $P(A, b)$ oder $P^=(A^T, -c)$ leer sein, im Widerspruch zur Annahme.

2. Fall: $\gamma > 0$. Dann folgt aus (2.4), dass

$$0 > \gamma(w^T b - v^T c) = (\gamma b^T)w - (\gamma c^T)v \geq (-v^T A^T)w + (w^T A)v = 0$$

und damit ein Widerspruch.

Also kann (2.4) nicht konsistent sein. Damit hat (2.3) eine Lösung, was zu zeigen war. \square

Eine analoge Aussage gilt für (PA) und (DA). Tatsächlich reicht für die starke Dualität aus, dass *eines* der beiden Probleme eine Lösung hat. Manchmal ist es auch einfacher, Lösbarkeit des dualen Problems zu zeigen als Existenz eines primal zulässigen Punkts.

Satz 2.4. Die folgenden Aussagen sind äquivalent:

- (i) (P) hat eine Lösung \bar{x} ;
- (ii) (D) hat eine Lösung \bar{y} ;
- (iii) (P) und (D) haben beide eine Lösung, für die gilt $c^T \bar{x} = -b^T \bar{y}$.

Beweis. Aus (iii) folgt trivialerweise (i) und (ii). Sei also \bar{x} Lösung von (P), d. h. gelte $\bar{x} \in P(A, b)$ und $c^T \bar{x} \leq c^T x$ für alle $x \in P(A, b)$. Es gibt also kein $x \in P(A, b)$ mit $c^T x \leq \gamma$ für beliebiges $\gamma < c^T \bar{x}$, d. h. das System

$$\begin{cases} Ax \leq b, \\ c^T x \leq \gamma \end{cases}$$

ist nicht konsistent. Nach Folgerung 1.8 a) gibt es daher $(u, \beta) \geq 0$ mit

$$\begin{cases} A^T u + \beta c = 0, \\ u^T b + \beta \gamma < 0. \end{cases}$$

Ist $\beta = 0$, so folgt daraus – wieder mit Folgerung 1.8 a) – dass $Ax \leq b$ inkonsistent ist, im Widerspruch zur Annahme $\bar{x} \in P(A, b)$. Also muss $\beta > 0$ gelten, und $y := \beta^{-1} u$ erfüllt

$$\begin{cases} A^T y = -c, \\ y^T b < -\gamma, \\ y \geq 0. \end{cases}$$

Insbesondere ist damit $y \in P^=(A^T, -c)$ und aus Satz 2.3 folgt die Behauptung (iii).

Man argumentiert analog, falls eine Lösung \bar{y} von (D) existiert. \square

Durch Kontraposition erhält man daraus Bedingungen, wann ein Problem *keine* Lösung hat. Das primale Problem (P) heisse *unbeschränkt*, falls

$$\inf_{x \in P(A, b)} c^T x = -\infty,$$

und *nicht zulässig*, falls $P(A, b) = \emptyset$. Analog heisse das duale Problem (D) *unbeschränkt*, falls

$$\sup_{y \in P^=(A^T, -c)} -b^T y = \infty,$$

und *nicht zulässig*, falls $P^=(A^T, -c) = \emptyset$.

Folgerung 2.5. *Es gilt:*

- (i) *Ist (P) unbeschränkt, dann ist (D) nicht zulässig.*
- (ii) *Ist (D) unbeschränkt, dann ist (P) nicht zulässig.*
- (iii) *Ist (P) nicht zulässig, dann ist (D) unbeschränkt oder nicht zulässig.*
- (iv) *Ist (D) nicht zulässig, dann ist (P) unbeschränkt oder nicht zulässig.*

Beweis. Für (i) sei $y \in P^=(A^T, -c)$. Dann folgt aus Satz 2.1

$$\inf_{x \in P(A, b)} c^T x \geq -b^T y > -\infty,$$

im Widerspruch zur Unbeschränktheit. Analog argumentiert man für (ii).

Für (iii) nehmen wir an, dass (P) nicht zulässig und (D) nach oben beschränkt ist, d. h. Maximum wird angenommen (sonst wären wir bereits fertig). Ist nun (D) zulässig, dann hat (P) nach Satz 2.4 eine Lösung, im Widerspruch zur Annahme. Also ist (D) nicht zulässig. Analog argumentiert man für (iv). \square

Dass tatsächlich beide Probleme unzulässig sein können, zeigt das folgende einfache Beispiel.

Beispiel 2.6. Sei

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

Dann ist nach Folgerung 1.8

- $P(A, b) = \emptyset$, da das System $u^T A = 0, u^T b < 0$ die Lösung $\bar{u} = (1, 1)^T \geq 0$ hat;
- $P^=(A^T, -c) = \emptyset$, da das System $u^T A^T \geq 0, u^T (-c) < 0$ die Lösung $\bar{u} = (-1, -1)$ hat.

2.2 KOMPLEMENTARITÄT

Wir beschäftigen uns nun mit der vergleichsfreien Charakterisierung von Lösungen von (P). Dazu verwenden wir, dass eine zulässige Lösung \bar{x} die beste untere Schranke in (2.2) annehmen wird – und diese ist genau durch die Lösung \bar{y} des dualen Problems gegeben.

Satz 2.7 (Schwache Komplementarität). *Es ist \bar{x} Lösung von (P) und \bar{y} Lösung von (D) genau dann, wenn gilt*

$$(2.5) \quad \begin{cases} A\bar{x} \leq b, & \text{(primale Zulässigkeit)} \\ A^T \bar{y} = -c, \quad \bar{y} \geq 0, & \text{(duale Zulässigkeit)} \\ \bar{y}_i (b_i - A_i \bar{x}) = 0 \quad \text{für alle } i = 1, \dots, m. & \text{(Komplementarität)} \end{cases}$$

Man nennt (2.5) auch *Karush–Kuhn–Tucker-Bedingungen* (oder kurz: *KKT-Bedingungen*).

Beweis. Ist \bar{x} Lösung von (P) und \bar{y} Lösung von (D), so sind natürlich insbesondere \bar{x} und \bar{y} zulässig. Nach Satz 2.4 (iii) ist dann wegen der dualen Zulässigkeit

$$0 = c^T \bar{x} - (-b^T \bar{y}) = (-A^T \bar{y})^T \bar{x} + b^T \bar{y} = (b - A\bar{x})^T \bar{y} = \sum_{i=1}^m (b_i - A_i \bar{x}) \bar{y}_i.$$

Aus primaler und dualer Zulässigkeit folgt, dass alle Summanden nicht-negativ sind und daher einzeln verschwinden müssen. Daraus folgt die Komplementarität.

Sind umgekehrt die KKT-Bedingungen erfüllt, dann ist $\bar{x} \in P(A, b)$ und $\bar{y} \in P^*(A^T, -c)$ und aus der Komplementarität folgt wie oben

$$0 = (b - A\bar{x})^T \bar{y} = c^T \bar{x} - (-b^T \bar{y}),$$

und damit nach Satz 2.1 dass \bar{x} Lösung von (P) und \bar{y} Lösung von (D) ist. □

Schwache Komplementarität kann man mit Hilfe von Lemma 2.2 auch für allgemeine lineare Optimierungsprobleme zeigen.

Folgerung 2.8. Seien A, B, C, D und c, d, e, f wie in Lemma 2.2. Dann ist (\bar{x}, \bar{y}) Lösung von (PA) und (\bar{u}, \bar{v}) Lösung von (DA) genau dann, wenn beide zulässig sind und die Komplementaritätsbedingungen

$$\begin{cases} \bar{u}_i (e_i - [A\bar{x} - B\bar{y}]_i) = 0 & \text{für alle } i = 1, \dots, m, \\ \bar{x}_i (c_i - [A^T \bar{u} + C^T \bar{v}]_i) = 0 & \text{für alle } i = 1, \dots, n, \end{cases}$$

erfüllen.

Beweis. Wie im Beweis von Lemma 2.2 transformiert man (PA) auf die Form (P) für geeignetes A', b', c', x' . Dann ist wie im Beweis von Satz 2.7 die Optimalität von (\bar{x}, \bar{y}) und (\bar{u}, \bar{v}) äquivalent mit

$$\begin{aligned} 0 &= c^T \bar{x} + d^T \bar{y} - (e^T \bar{u} + f^T \bar{v}) \\ &= c^T \bar{x} + (B^T \bar{u} + D^T \bar{v})^T \bar{y} - e^T \bar{u} - (C\bar{x} + D\bar{y})^T \bar{v} + (\bar{u}^T A\bar{x} - \bar{u}^T A\bar{x}) \\ &= (c - A^T \bar{u} - C^T \bar{v})^T \bar{x} - (e - A\bar{x} - B\bar{y})^T \bar{u}. \end{aligned}$$

Aus den Ungleichungsbedingungen folgt, dass die erste Klammer (und wegen $\bar{x} \geq 0$ der ganze Term) nicht-negativ und die zweite Klammer (und wegen $\bar{u} \geq 0$ der ganze Term) nicht-positiv ist. Die Differenz kann also nur gleich Null sein, wenn beide Terme separat verschwinden, und daraus folgt wie zuvor die Komplementarität. □

Die Komplementaritätsbedingung sagt, dass $\bar{y}_i = 0$ oder $b_i - A_i \bar{x} = 0$ für jedes $1 \leq i \leq m$ gilt. Dies ist aber kein exklusives Oder – es ist also zugelassen, dass sowohl \bar{y}_i als auch $b_i - A_i \bar{x}$ verschwinden. Man kann jedoch zeigen, dass unter allen Lösungen auch ein Paar existiert, für das immer nur genau eine Gleichheit gilt.

Satz 2.9 (Strikte Komplementarität). *Sind $P(A, b)$ und $P^=(A^T, -c)$ nicht leer, so existiert eine Lösung \bar{x} von (P) und eine Lösung \bar{y} von (D) für die gilt*

$$\bar{y}_i = 0 \quad \text{genau dann, wenn} \quad A_i \bar{x} < b_i \quad \text{für alle } i = 1, \dots, m.$$

Beweis. Wir zerlegen die Menge der Ungleichungen wie folgt. Definiere für $A \in \mathbb{R}^{m \times n}$ die Zeilenmenge $M := \{1, \dots, m\}$ und

$$\begin{aligned} N &:= \{i \in M : A_i \bar{x} < b_i \text{ für eine Lösung } \bar{x} \text{ von (P)}\}, \\ B &:= \{i \in M : \bar{y}_i > 0 \text{ für eine Lösung } \bar{y} \text{ von (D)}\}. \end{aligned}$$

Die zu beweisende Aussage ist dann äquivalent mit der Behauptung, dass $N \cap B = \emptyset$ und $M = N \cup B$ ist.

Ersteres folgt direkt aus der schwachen Komplementarität: Wäre $i \in N \cap B$, so würde gelten $A_i \bar{x} < b_i$ für eine Lösung \bar{x} des primalen Problems und $\bar{y}_i > 0$ für eine Lösung \bar{y} des dualen Problems, im Widerspruch zu Satz 2.7.

Für die zweite Aussage zeigen wir zuerst, dass es Lösungen gibt, die alle strikten Ungleichungen gleichzeitig erfüllen. Wir wählen zu jedem $i \in N$ eine der Lösungen $\bar{x}^{(i)}$, die nach Definition die entsprechende strikte Ungleichung erfüllt, und bilden die Konvexkombination

$$\bar{x} := \frac{1}{|N|} \sum_{i \in N} \bar{x}^{(i)}.$$

Da die Lösungsmenge eines linearen Optimierungsproblems konvex ist, ist \bar{x} wieder Lösung von (P). Weiterhin gilt für alle $j \in N$

$$A_j \bar{x} = \sum_{i \in N} \frac{1}{|N|} A_j \bar{x}^{(i)} < \frac{1}{|N|} b_j + \frac{|N| - 1}{|N|} b_j = b_j,$$

da alle $\bar{x}^{(i)}$ zulässig sind und für $j = i$ die Ungleichung nach Konstruktion sogar strikt ist. Genauso definieren wir \bar{y} mit $\bar{y}_i > 0$ für alle $i \in B$.

Wir schreiben nun A_N für die Matrix, die nur die Zeilen A_i mit $i \in N$ enthält, und b_N für den Vektor mit Einträgen b_i mit $i \in N$; wir definieren analog A_B und b_B . Mit $N \cap B = \emptyset$ erhalten wir dann

$$\begin{aligned} A_N \bar{x} &< b_N, & \bar{y}_B &> 0, \\ A_B \bar{x} &= b_B, & \bar{y}_N &= 0. \end{aligned}$$

Angenommen, $J := M \setminus (B \cup N)$ wäre nicht leer, d. h. es existiert ein $j \in J$. Nach Definition gilt dann $\tilde{y}_j = 0$ und $(b - A\tilde{x})_j = 0$. Wir zeigen nun, dass das System

$$(2.6) \quad \begin{cases} A_{J \setminus \{j\}} x \leq 0, \\ A_B x = 0, \\ A_j x < 0, \end{cases}$$

keine Lösung haben kann. Wäre nämlich \tilde{x} eine Lösung, so gäbe es wegen $A_i \tilde{x} < b_i$ für alle $i \in N$ ein $\varepsilon > 0$ klein genug, so dass $A_i(\tilde{x} + \varepsilon \tilde{x}) \leq b_i$ für alle $i \in M$ gilt. Da $A_i \tilde{x} \leq 0$ für $i \notin N$ ist, wäre also auch $\tilde{x} + \varepsilon \tilde{x}$ zulässig für (P). Wegen $\tilde{y}_j = 0$ erfüllt $\tilde{x} + \varepsilon \tilde{x}$ auch die schwachen Komplementaritätsbedingungen, ist also nach Satz 2.7 sogar eine Lösung. Da aus (2.6) sogar

$$A_j(\tilde{x} + \varepsilon \tilde{x}) < A_j \tilde{x} \leq b_j$$

und damit die strikte Komplementarität in j folgt, erhalten wir $j \in N$ und damit ein Widerspruch.

Also hat (2.6) keine Lösung, und aus dem Farkas-Lemma (Folgerung 1.7 mit $C^T = -A_{J \setminus \{j\}}$, $D^T = -A_B$, $b^T = A_j$) erhalten wir die Existenz von $v \geq 0$ und w mit

$$(2.7) \quad -(A_{J \setminus \{j\}})^T v - (A_B)^T w = A_j.$$

Definiere nun $\tilde{y} \in \mathbb{R}^m$ mit

$$\tilde{y}_N = 0, \quad \tilde{y}_B = w, \quad \tilde{y}_{J \setminus \{j\}} = v, \quad \tilde{y}_j = 1,$$

dann gilt wegen (2.7), dass $A^T \tilde{y} = 0$ ist. Ausserdem ist $\tilde{y}_i \geq 0$ für alle $i \in J \cup N$ sowie $\tilde{y}_j > 0$. Da $\tilde{y}_B > 0$ ist, erfüllt für $\varepsilon > 0$ klein genug also $\tilde{y} + \varepsilon \tilde{y}$ das System

$$\begin{aligned} A^T(\tilde{y} + \varepsilon \tilde{y}) &= -c, \\ (\tilde{y} + \varepsilon \tilde{y}) &\geq 0, \end{aligned}$$

ist damit zulässig für (D). Wegen $(b - A\tilde{x})_j = 0$ erfüllt $\tilde{y} + \varepsilon \tilde{y}$ auch die schwachen Komplementaritätsbedingungen, ist also nach Satz 2.7 sogar eine Lösung. Wieder folgt aus $\tilde{y}_j > 0$ die strikte Komplementarität in j und damit $j \in B$, im Widerspruch zur Annahme. Also ist $J = \emptyset$ und damit $N \cup B = M$, was zu zeigen war. \square

Auch dieses Resultat kann wie in Folgerung 2.8 auf allgemeine lineare Optimierungsprobleme erweitern, was wir hier aber nicht ausführen.

GEOMETRIE DER POLYEDER

3

Aus Satz 2.9 folgt, dass wenn ein lineares Optimierungsproblem eine Lösung hat, eine Lösung existiert für die in einigen Ungleichungen (nämlich denen, die einer strikt positiven dualen Lösung entsprechen) sogar Gleichheit gilt. Anschaulich heisst das, dass diese Lösung auf einer Seite des zulässigen Polyeders liegt. Wir wollen nun zu zeigen, dass unter bestimmten Bedingungen die Lösung sogar in einer Ecke liegt. Dafür müssen wir insbesondere den Begriff von Seite und Ecke mathematisch präzisieren.

Sei $P \subset \mathbb{R}^n$ ein Polyeder. Eine Ungleichung $a^T x \leq \beta$ mit $a \in \mathbb{R}^n$ und $\beta \in \mathbb{R}$ heisst *gültig* für P , falls

$$P \subset \{x \in \mathbb{R}^n : a^T x \leq \beta\}$$

gilt. Wir nennen $F \subset P$ *Seite* von P , falls es eine gültige Ungleichung $a^T x \leq \beta$ gibt mit

$$F = \{x \in P : a^T x = \beta\}.$$

Eine Seite F heisst *nichttrivial*, falls F weder leer noch gleich P ist.

Mit dieser Definition erhalten wir sofort das Gewünschte.

Lemma 3.1. *Die Lösungsmenge von $\min_{x \in P} c^T x$ ist eine Seite von P .*

Beweis. Ist die Lösungsmenge leer, so gilt die Aussage trivialerweise. Ist die Menge nichtleer, so setzen wir $\eta := \min_{x \in P} c^T x$. Dann ist $-c^T x \leq -\eta$ gültig für P , und die Lösungsmenge ist darstellbar als $\{x \in P : c^T x = \eta\}$. \square

Dieses Resultat ist noch nicht besonders überzeugend, da nicht klar ist, ob dann auch Gleichheit in einer Ungleichung aus $Ax \leq b$ gelten muss. Dies zeigt aber der nächste Satz.

Satz 3.2. *Sei $P := P(A, b)$ ein Polyeder und $F \neq \emptyset$. Dann ist F Seite von P genau dann, wenn es ein Teilsystem $A'x \leq b'$ von $Ax \leq b$ gibt mit $F = \{x \in P : A'x = b'\}$.*

Beweis. Sei F Seite von P , d. h. es existieren $a \in \mathbb{R}^n$, $\beta \in \mathbb{R}$ mit

$$F = \{x \in P : a^T x = \beta\} \quad \text{und} \quad \beta = \max_{x \in P} a^T x.$$

Also ist $a^T x \leq \beta$ eine gültige Ungleichung für P . Die Seite F ist daher Lösungsmenge von

$$\max_{x \in P(A, b)} a^T x = - \min_{x \in P(A, b)} (-a^T x).$$

Nach Satz 2.4 hat das duale Problem $\min_{y \in P=(A^T, a)} -b^T y$ eine Lösung \bar{y} . Satz 2.7 ergibt dann, dass $A_i \bar{x} = b_i$ für alle i mit $\bar{y}_i > 0$ gilt. Dies ergibt das gewünschte Teilsystem.

Sei umgekehrt $F = \{x \in P : A'x = b'\}$. Setze $a := (A')^T \mathbb{1}$ (d. h. a ist die Summe der Zeilen von A'), dann für alle $x \in P$

$$a^T x = ((A')^T \mathbb{1})^T x = \mathbb{1}^T (A'x) \leq \mathbb{1}^T b =: \beta,$$

also ist $a^T x \leq \beta$ gültig für P . Weiter gilt für alle $x \in F$

$$a^T x = \mathbb{1}^T A'x = \mathbb{1}^T b' = \beta,$$

und damit ist F Seite von P . □

Folgerung 3.3. Seien P ein Polyeder und $F \subset P$ eine Seite. Dann gilt:

- (i) P hat endlich viele Seiten.
- (ii) F ist wieder ein Polyeder.
- (iii) $F' \subset F$ ist Seite von P genau dann, wenn F' Seite von F ist.

Dies ist noch nicht völlig befriedigend, da eine Seite durch mehrere Ungleichungen dargestellt werden kann, ohne dass klar ist ob ihre Anzahl etwas über die Dimension der Seite aussagt. Wir suchen daher eine sparsamere Darstellung. Eine nichttriviale Seite, die nicht Teilmenge einer anderen nichttrivialen Seite ist, nennen wir *Facette*. Diese charakterisieren wir nun. Dafür nennen wir eine Ungleichung $A_i x \leq b_i$ von $Ax \leq b$

- *implizite Gleichung* (in $Ax \leq b$), falls $A_i x = b_i$ für alle $x \in P(A, b)$ gilt;
- *redundant* (in $Ax \leq b$), falls sie durch positive Linearkombination anderer Ungleichungen in $Ax \leq b$ dargestellt werden kann.

Wir bezeichnen das System der impliziten Gleichungen mit $A^=x \leq b^=$ und das der verbleibenden nicht-redundanten Ungleichungen mit $A^+x \leq b^+$. Wie im Beweis von Satz 2.9 zeigt man nun, dass alle diese Ungleichungen gleichzeitig strikt sein können.

Lemma 3.4. *Ist $P(A, b)$ nichtleer, so existiert ein $x \in P(A, b)$ mit $A^-x = b^-$ und $A^+x < b^+$.*

Beweis. Ist A^+ die leere Matrix, so ist die Behauptung trivial. Sei also A^+ nicht leer. Für jede Zeile $A_i x \leq b_i$ aus $A^+x \leq b^+$ muss dann ein $x^{(i)} \in P(A, b)$ existieren mit $A_i x^{(i)} < b_i$ und $A^-x^{(i)} = b^-$. Bezeichne wieder N die Menge der strikten Ungleichungen und definiere $x := \frac{1}{|N|} \sum_{i \in N} x^{(i)} \in P(A, b)$. Dann gilt $A^+x < b^+$ und $A^-x = b^-$. \square

Jeder (nichttriviale) Polyeder besitzt also ein nichtleeres relatives Inneres.

Satz 3.5. *Eine Seite F ist Facette von $P(A, b)$ genau dann, wenn*

$$F = \{x \in P(A, b) : A_i x = b_i\}$$

ist für eine Ungleichung aus $A^+x \leq b^+$.

Beweis. Sei $F = \{x \in P : A'x = b'\}$ eine Facette von $P := P(A, b)$, wobei $A'x \leq b'$ ein Teilsystem von $A^+x \leq b^+$ ist (ein solches existiert, da $F \neq P$ nach Annahme), und sei $A_i x \leq b_i$ eine Ungleichung aus $A'x \leq b'$. Dann ist $F' = \{x \in P : A_i x = b_i\}$ nach Satz 3.2 eine Seite von P mit $F \subset F' \subset P$. Weil $A_i x \leq b_i$ keine implizite Gleichung von P ist, gilt $F' \neq P$, und aus der Maximalität von F folgt $F = F'$.

Sei umgekehrt $A_i x \leq b_i$ eine Ungleichung aus $A^+x \leq b^+$, und habe F die angegebene Darstellung. Nach Satz 3.2 ist F eine Seite von P ; wir haben also zu zeigen, dass F maximal ist. Seien dafür $A'x \leq b'$ die verbleibenden Ungleichungen. Nach Lemma 3.4 existiert dann ein \bar{x} mit $A^-\bar{x} = b^-$ und $A^+\bar{x} < b^+$. Da $A_i x \leq b_i$ nach Voraussetzung nichtredundant ist, gibt es – wieder mit Lemma 3.4 – ein \hat{x} mit $A^-\hat{x} = b^-$, $A'\hat{x} < b'$ und $A_i \hat{x} > b_i$. Aus der Stetigkeit von $x \mapsto A_i x$ und der Konvexität von Polyedern folgt dann die Existenz eines $t \in (0, 1)$ so dass für $x(t) := t\hat{x} + (1-t)\bar{x} \in P$ gilt

$$A^-x(t) = b^-, \quad A_i x(t) = b_i, \quad A'x(t) < b',$$

d. h. $x(t) \in F$. Sei nun $F' = \{x \in P : A''x = b''\}$ eine beliebige Seite von P mit $F \subset F'$, wobei $A''x \leq b''$ ein Teilsystem von $A^+x \leq b^+$ ist. Dann ist $x(t) \in F \subset F'$. Wegen $A'x(t) < b'$ ist aber $A_i x \leq b_i$ die einzige Ungleichung, für die Gleichheit in $x(t)$ gilt. Damit muss aber $A'' = A_i$ und $b'' = b_i$ sein, woraus $F' = F$ und damit Maximalität von F folgt. Also ist F Facette von P . \square

Jede Seite von P ausser P selbst ist also Schnitt von Facetten, den maximalen Seiten. Umgekehrt wollen wir Ecken charakterisieren als minimale Seiten. Dabei ist nicht klar, ob P überhaupt Ecken hat (Halbräume sind ja auch Polyeder). Wir erinnern: Mengen von der Form $\{x \in \mathbb{R}^n : Ax = b\}$ heissen *affiner Unterraum* von \mathbb{R}^n .

Folgerung 3.6. Ein Polyeder P hat genau dann keine nichttrivialen Seiten, wenn P ein affiner Unterraum ist.

Beweis. Ein Polyeder $P = P(A, b)$ ist genau dann ein affiner Unterraum, wenn $A^+x \leq b^+$ leer ist. Dies ist nach Satz 3.5 dann und nur dann der Fall, wenn es keine Facetten gibt. Da jede Seite Schnitt von Facetten ist, ist dies aber äquivalent dazu, dass es keine nichttrivialen Seiten gibt. \square

Daraus folgt sofort, dass eine Seite (die ja selber ein Polyeder ist) genau dann keine weiteren nichtleeren Seiten enthält – wir sagen: eine *minimale Seite* ist – wenn sie ein affiner Unterraum ist.

Satz 3.7. Eine nichtleere Menge F ist genau dann minimale Seite von $P(A, b)$, wenn $F \subset P(A, b)$ ist und die Form $F = \{x \in \mathbb{R}^n : A'x = b'\}$ hat für ein Teilsystem $A'x \leq b'$ von $Ax \leq b$.

Beweis. Hat F die angegebene Darstellung, so ist sie ein affiner Teilraum und damit minimale Seite. Sei umgekehrt $F \neq \emptyset$ eine minimale Seite. Dann ist

$$F = \{x \in \mathbb{R}^n : A''x \leq b'', A'x = b'\}$$

für zwei Teilsysteme von $A^+x \leq b^+$, wobei $A''x \leq b''$ minimal sein und insbesondere keine redundanten Ungleichungen in $A''x \leq b''$, $A'x = b'$, enthalten soll. Da F minimal ist, enthält es keine nichttrivialen Seiten und ist daher nach Folgerung 3.6 ein affiner Unterraum. Also muss $A''x \leq b''$ leer sein, und F daher die gewünschte Darstellung haben. \square

Das nächste Lemma charakterisiert die minimalen Seiten. Wir erinnern: der Rang einer Matrix $\text{rang } A$ ist die Anzahl der linear unabhängigen Zeilen.

Lemma 3.8. Alle nichttrivialen minimalen Seiten von $P(A, b)$ haben die Dimension $n - \text{rang } A$.

Beweis. Sei wie in Satz 3.7 eine Seite $F = \{x \in \mathbb{R}^n : A'x = b'\}$ für ein Teilsystem $A'x \leq b'$ gegeben. Aus der linearen Algebra wissen wir, dass die Lösungsmenge eines linearen Gleichungssystems $A'x = b'$ die Dimension $n - \text{rang } A'$ hat. Wir zeigen nun durch Widerspruch, dass $\text{rang } A' = \text{rang } A$ gilt. Angenommen, es wäre $\text{rang } A > \text{rang } A'$. (Da es sich um ein Teilsystem handelt, kann der Rang ja nicht kleiner sein.) Dann existiert eine Zeile A_i von A , die linear unabhängig zu allen Zeilen von A' ist. Damit gilt aber wegen $F \subset P$, dass

$$F \subset \{x \in \mathbb{R}^n : A'x = b' \text{ und } A_ix \leq b_i\} \subsetneq \{x \in \mathbb{R}^n : A'x = b'\} = F,$$

ein Widerspruch. \square

Damit kommen wir endlich zu der gesuchten Charakterisierung von Ecken: Eine Menge $F \subset P$ heisst *Ecke* von P , wenn F eine (minimale) Seite der Dimension Null ist. Ein Polyeder, der Ecken hat, heisst *spitz*.

Satz 3.9. *Folgende Aussagen über ein Polyeder $P \subset \mathbb{R}^n$ sind äquivalent:*

- (i) P ist spitz;
- (ii) $\text{rang } A = n$;
- (iii) jede nichtleere Seite von P ist spitz.

Beweis. Die Äquivalenz von (i) und (ii) folgt aus der Definition und Lemma 3.8. Hat eine nichtleere Seite F von P eine Ecke, so ist diese (als minimale Seite) nach Folgerung 3.3 (iii) auch Ecke von P , also ist F spitz. Sei umgekehrt P spitz und $F \subset P$ eine nichtleere Seite. Angenommen, eine minimale Seite F_0 von F ist keine Ecke, d. h. F_0 ist ein affiner Teilraum der Dimension größer Null. Da F_0 nach Folgerung 3.3 (iii) auch Seite von P ist, hat auch P eine minimale Seite der Dimension größer Null. Nach Lemma 3.8 haben aber alle minimalen Seiten von P den selben Rang, und damit kann P keine Ecken haben. Damit ist (i) äquivalent zu (iii). \square

Nach all der Vorarbeit erhalten wir nun das zentrale Resultat dieses Kapitels.

Satz 3.10. *Ist P ein spitzer Polyeder und ist $\min_{x \in P} c^T x$ lösbar, so existiert eine Lösung, die Ecke von P ist.*

Beweis. Nach Lemma 3.1 ist die Lösungsmenge des Optimierungsproblems eine (nichtleere) Seite von P , die nach Satz 3.9 spitz ist, also eine Ecke enthält. Diese ist die gesuchte Lösung. \square

Dies ist auch der Grundgedanke der Simplex-Methode, die wir in einem folgenden Kapitel betrachten werden: Um eine Lösung zu finden, muss man nur an den Ecken von P suchen – die dafür natürlich existieren müssen. Für Polyeder der Form $P^=(A, b) = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ ist das stets der Fall.

Satz 3.11. *Ein nichtleerer Polyeder $P^=(A, b)$ ist spitz.*

Beweis. Wir können $P^=(A, b)$ darstellen als $P(D, f)$ mit

$$D = \begin{pmatrix} A \\ -A \\ -I \end{pmatrix}, \quad f = \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}.$$

Auf Grund des Identitätsblocks ist $\text{rang } D = n$, und aus Satz 3.9 folgt die Behauptung. \square

Schliesslich charakterisieren wir die Ecken von $P^=(A, b)$. Bezeichne dafür für $x \in \mathbb{R}^n$

$$\text{supp}(x) := \{j \in \{1, \dots, n\} : x_j > 0\}$$

den Träger von x .

Satz 3.12. *Die folgenden Aussagen sind äquivalent:*

- (i) $\{x\} \subset \mathbb{R}^n$ ist Ecke von $P^=(A, b)$;
- (ii) Die Spaltenvektoren $A^j, j \in \text{supp}(x)$, von A sind linear unabhängig.

Beweis. Wir schreiben $P^=(A, b)$ wie im Beweis von Satz 3.11 als $P(D, f)$. Dann ist $\{x\}$ nach Satz 3.7 und Satz 3.9 Ecke von $P^=(A, b)$ genau dann, wenn ein Teilsystem $D'x \leq f'$ existiert mit

$$\{x \in \mathbb{R}^n : D'x = f'\} = \{x\}$$

und $\text{rang } D' = n$. Betrachten wir nun die Struktur von D' genauer, so bedeutet das, dass $Ax = b$ diejenigen Komponenten von x , die im Träger von x liegen, eindeutig festlegt. (Alle anderen sind ja nach Definition Null, und $-Ax = -b$ ist offensichtlich linear abhängig von $Ax = b$.) Dies ist aber genau dann der Fall, wenn die Spalten von A , die zu den Komponenten x_j im Träger $\text{supp}(x)$ gehören, linear unabhängig sind. (Für die anderen ist $x_j = 0$, so dass die entsprechenden Spalten keine Rolle spielen; es reicht also, dass die reduzierte Matrix vollen Rang hat.) □

Teil II

ANWENDUNGSBEISPIELE

SPIELTHEORIE

4

Wir betrachten in diesem und den nächsten Kapiteln einige Anwendungen der linearen Optimierung, insbesondere der Dualitätstheorie. Das Ziel ist dabei nicht unbedingt, dass man die betrachteten Probleme mit den im letzten Teil behandelten Algorithmen lösen kann – oft existieren andere, an das Problem speziell angepasste Verfahren die deutlich effizienter sind – sondern deutlich machen, wie hilfreich der Blickwinkel der Dualität sein kann. Wir beginnen mit dem Nobelpreis-ergiebigsten Gebiet der Mathematik: der Spieltheorie.

Inhalt der Spieltheorie ist die mathematische Modellierung von rationalem, interaktivem Verhalten in Entscheidungssituationen. Ein typisches Problem ist die Frage nach optimalen Strategien in Gegenwart eines Antagonisten, der auf die eigenen Strategien reagieren kann. Das einfachste – und unverfänglichste – Beispiel für eine solche Situation stellen übliche Gesellschaftsspiele dar (daher der Name), aber das Interesse an der Spieltheorie begründet sich natürlich an den vielzähligen (komplexeren) Anwendungen in Wirtschaft und Politik. Hier betrachten wir die einfachste Variante: Zweipersonenspiele, bei denen jeder *Spieler* (die wir hier I und J nennen) gleichzeitig und unabhängig eine von mit $1, \dots, m$ bzw. $1, \dots, n$ indizierten vorgegebenen Aktionen wählen kann. Je nach Kombination (i, j) muss einer der Spieler einen bestimmten Betrag an den anderen Spieler zahlen. Damit lassen sich schon viele Situationen modellieren; wir betrachten wieder nur die einfachsten.

Beispiel 4.1 (Parität). Das erste Beispiel ist eine Zwei-Spieler-Variante von „Kopf oder Zahl“. Jeder Spieler nennt unabhängig eine natürliche Zahl; die genannten Zahlen werden dann summiert. Ist die Summe gerade, so muss I an J einen Einsatz zahlen; ist die Summe ungerade, zahlt J an I den selben Betrag.

Offenbar kommt es für das Ergebnis nur darauf an, ob ein Spieler eine gerade oder eine ungerade Zahl nennt. Schreibt man die möglichen Aktionen von I als Zeilen und die von J als Spalten und trägt für alle möglichen Kombinationen den Betrag, den I an J zahlen muss, ein, kann man das Spiel wie folgt charakterisieren:

I spielt \ J spielt	gerade	ungerade
gerade	1	-1
ungerade	-1	1

Beispiel 4.2 (Schere, Stein, Papier). Dieses universell bekannte Spiel ist wohl um die letzte Jahrhundertwende über Japan nach Europa und von dort weiter in die USA verbreitet worden und stellt das Musterbeispiel für ein Spiel mit zyklisch dominierenden Aktionen dar: „Schere schlägt Stein schlägt Papier schlägt Schere“. Die Auszahlungstabelle ist wie folgt (Unentschieden bei gleicher Wahl bedeutet keine Auszahlung):

I spielt \ J spielt	Stein	Schere	Papier
Stein	0	-1	1
Schere	1	0	-1
Papier	-1	1	0

Solche Spiele kann man also vollständig durch Matrizen beschreiben. Ein *Matrixspiel* besteht also aus einem Tripel (m, n, A) mit $m, n \in \mathbb{N}$ und $A \in \mathbb{R}^{m \times n}$. Jeder Zeilen- (bzw. Spalten-)Index heisst *Zeilen-* (bzw. *Spalten-*)*Aktion*. Die Einträge A_{ij} heissen *Auszahlungen* (von I and J), die Matrix A *Auszahlungsmatrix*. Einem Paar (i, j) entspricht also ein Gewinn A_{ij} von J und ein gleicher Verlust von I. (Negative Einträge entsprechen natürlich einem Verlust von J bzw. einem Gewinn von I.) Die Frage ist nun, wie ein Spieler seine Aktionen wählen kann, um seine Auszahlung auch beim bestmöglichen Gegenspiel zu nutzen, insbesondere wenn mehrere (unabhängige) Spiele in Reihe gespielt werden. Da es dabei wegen der Annahme der Unabhängigkeit nicht auf die Reihenfolge der Spiele ankommt, reduziert sich die Frage darauf, wie verhältnismässig oft die verschiedenen Aktionen gewählt werden sollen – unter der Annahme, dass der Gegenspieler diese Information besitzt! Wir illustrieren das Prinzip am einfachsten Beispiel.

Beispiel 4.3 (Parität). Angenommen, der Spieler I wählt in jedem Spiel mit Wahrscheinlichkeit $p \in [0, 1]$ „gerade“ und mit Wahrscheinlichkeit $1 - p$ „ungerade“. Abhängig von seiner Wahl bekommt J dann in Erwartung ausgezahlt:

- J spielt „gerade“: $1 \cdot p + (-1) \cdot (1 - p) = 2p - 1$;
- J spielt „ungerade“: $(-1) \cdot p + 1 \cdot (1 - p) = 1 - 2p$.

Ist also $p \leq 1/2$, so wird J „ungerade“ spielen und damit einen Gewinn von $1 - 2p \geq 0$ erzielen; ist $p > 1/2$ wählt J „gerade“ und gewinnt $2p - 1 > 0$. Die einzige Möglichkeit, wie I einen Verlust verhindern kann, ist also $p = 1/2$ zu wählen. Da die Auszahlungsmatrix symmetrisch ist, wird J für sich genauso wählen.

Auch dies formalisieren wir. Sei (m, n, A) ein Matrixspiel und definiere für $k \in \mathbb{N}$ den *k-Simplex*

$$S^k := \{x \in \mathbb{R}^k : x \geq 0 \text{ und } \mathbf{1}^T x = 1\}.$$

Jeder Vektor $x \in X := S^n$ bzw. $y \in Y := S^m$ heisst *Strategie* für J bzw. I. Ist (z. B.) $x = e_j$ für ein $1 \leq j \leq n$, so spricht man von einer *reinen Strategie*, ansonsten von einer *gemischten Strategie*. (Reine Strategien entsprechen genau den Ecken von S^n .) Für zwei Strategien x und

y ist die *mittlere Auszahlung* (d. h. der Erwartungswert der Auszahlung als Zufallsvariable) definiert als

$$y^T Ax = \sum_{i=1}^m \sum_{j=1}^n A_{ij} x_j y_i.$$

Wir versuchen zuerst, in einem einfachen Beispiel eine optimale Strategie herzuleiten.

Beispiel 4.4 (abstraktes Matrixspiel). Wir betrachten das Matrixspiel $(2, 3, A)$ für

$$A = \begin{pmatrix} 2 & -2 & 0 \\ -3 & 4 & -2 \end{pmatrix}.$$

Auf den ersten Blick scheint das Spiel I zu bevorzugen, denn der Betrag der Summe der negativen Auszahlungen (die ja seinem Gewinn entsprechen) ist kleiner als die Summe der positiven Auszahlungen (die dem Gewinn von J entsprechen). Wie sich zeigen wird, trügt das aber!

Bei diesem Spiel ist die mittlere Auszahlung für ein gegebenes Paar (x, y) an Strategien

$$y^T Ax = 2x_1 y_1 - 2x_2 y_1 - 3x_1 y_2 + 4x_2 y_2 - 2x_3 y_2.$$

Stellen wir uns zuerst auf den Standpunkt von I, der diese Auszahlung (an J) bei bekannter Strategie x von J minimieren möchte. Dies führt auf Lösung des Problems

$$\left\{ \begin{array}{l} \min_y (2x_1 - 2x_2)y_1 + (-3x_1 + 4x_2 - 2x_3)y_2 \\ \text{mit } y_1 + y_2 = 1, \\ y_1 \geq 0, \\ y_2 \geq 0. \end{array} \right.$$

Da S^2 nichtleer und beschränkt ist, hat dieses Problem eine Lösung \bar{y} , die nach Satz 3.10 in einer Ecke liegt. Wir müssen daher nur die beiden Strategien $y = (1, 0)$ und $y = (0, 1)$ vergleichen, d. h. der minimale Wert ist

$$\min\{2x_1 - 2x_2, -3x_1 + 4x_2 - 2x_3\}.$$

Der Spieler J, der diese Rechnung verfolgt hat, möchte dieses Minimum – welcher Term auch immer es sein wird – so groß wie möglich halten. Da er von Aktion 3 wegen $x_3 \geq 0$ ja nie profitieren kann, wird er $x_3 = 0$ wählen. Verwendet man noch $x_2 = 1 - x_1$, so reduziert sich das Problem darauf, das Maximum über alle x_1 von

$$\min\{-2 + 4x_1, 4 - 7x_1\}$$

zu bilden. Das Minimum ist aber maximal, wenn beide Terme zusammenfallen (sonst könnte man einen der Terme noch vergrößern). Auflösen nach x_1 und Einsetzen liefert die Strategie $\bar{x} = \frac{1}{11}(6, 5, 0)^T$ und den zugehörigen Wert $2/11$ – und damit einen mittleren Gewinn für J! Analoges Vorgehen mit vertauschten Rollen liefert die Strategie $\bar{y} = \frac{1}{11}(7, 4)^T$. Natürlich war die Herleitung heuristisch, und die Optimalität dieser Strategie muss noch bewiesen werden.

Wir wenden nun den Ansatz aus Beispiel 4.4 auf den allgemeinen Fall an. Spielt J die Strategie x , so hat I das lineare Optimierungsproblem

$$\min_{y \in Y} y^T A x$$

zu lösen. Der Spieler J dagegen ist bestrebt, diese Auszahlung an ihn durch Wahl von x zu maximieren, d. h. betrachtet

$$(M_J) \quad \max_{x \in X} \min_{y \in Y} y^T A x.$$

Vertauschen wir die Rollen von I und J, so erhalten wir analog die Aufgabe

$$(M_I) \quad \min_{y \in Y} \max_{x \in X} y^T A x.$$

(Solche Probleme bezeichnet man als *Sattelpunktprobleme*.) Die Frage ist nun: Macht dies für die optimalen Strategien einen Unterschied? Kann dies sogar zu unterschiedlichen Auszahlungen führen? Der folgende Satz¹ gibt die erhoffte negative Antwort.

Satz 4.5 (Minimax). *Sei (m, n, A) ein Matrixspiel. Dann gilt*

$$\max_{x \in X} \min_{y \in Y} y^T A x = \min_{y \in Y} \max_{x \in X} y^T A x,$$

und die Gleichheit wird in einem Paar (\bar{x}, \bar{y}) (genannt Sattelpunkt) genau dann angenommen mit $\bar{y}^T A \bar{x} = \bar{v}$, wenn (\bar{x}, \bar{v}) Lösung ist von

$$(P_J) \quad \begin{cases} \max_{x, v} v \\ \text{mit } A x - v \mathbf{1} \geq 0, \\ \mathbf{1}^T x = 1, \\ x \geq 0, \end{cases}$$

und (\bar{y}, \bar{v}) Lösung ist von

$$(P_I) \quad \begin{cases} \min_{y, \mu} \mu \\ \text{mit } A^T y - \mu \mathbf{1} \leq 0, \\ \mathbf{1}^T y = 1, \\ y \geq 0. \end{cases}$$

Beweis. Wir nehmen zuerst den Standpunkt von J ein und betrachten für eine gegebene Strategie $x \in X$ das Problem $\min_{y \in Y} y^T A x$, das I zu lösen hat. Da das Optimum – wenn es existiert – nach Satz 3.10 in einer Ecke angenommen wird, muss gelten

$$\min_{y \in Y} y^T A x = \min_{1 \leq i \leq m} e_i^T A x = \min_{1 \leq i \leq m} A_i x.$$

¹Dieser ist laut [John von Neumann](#), der ihn bewies, der Beginn der mathematischen Spieltheorie.

Dieses Minimum lässt sich also charakterisieren als das maximale $v \in \mathbb{R}$ mit $A_i x \geq v$ für alle $1 \leq i \leq n$. Das Sattelpunktproblem (M_J) – in dem zusätzlich über alle $x \in X$ maximiert wird – ist daher äquivalent zum linearen Optimierungsproblem (P_J) . Analog hat aus Sicht von I der Spieler J das Problem $\max_{1 \leq j \leq m} (A^T)_j y$ (Skalarprodukt von y und der j -ten Spalte von A) zu lösen, dessen Maximum charakterisiert wird als minimales μ mit $(A^T)_j y \leq \mu$ für alle $1 \leq j \leq m$, woraus die Äquivalenz von (M_I) und (P_I) folgt. Nun sind (P_J) und (P_I) nach Lemma 2.2 dual zueinander (nach Übergang von y zu $-y$). Da die zulässigen Mengen nichtleer sind (wähle z. B. $x = e_i, y = e_j, v = \mu = A_{ij}$), folgt aus Satz 2.3, dass beide Probleme eine Lösung (\bar{x}, \bar{v}) bzw. $(\bar{y}, \bar{\mu})$ besitzen, und dass gilt $\bar{\mu} = \bar{v}$. Daraus folgt die behauptete Gleichheit.

Bleibt zu zeigen, dass die Lösungen des primalen bzw. dualen Optimierungsproblems genau die Sattelpunkte sind. Sei zuerst (\bar{x}, \bar{v}) Lösung von (P_J) und $(\bar{y}, \bar{\mu})$ Lösung von (P_I) . Dann folgt aus der schwachen Dualität $\bar{v} = \bar{\mu}$, und daher gilt wegen $\mathbb{1}^T \bar{x} = 1 = \mathbb{1}^T \bar{y}$

$$\bar{y}^T A \bar{x} \geq \bar{y}^T (\bar{v} \mathbb{1}) = \bar{v} = \bar{\mu} = \bar{x}^T (\bar{\mu} \mathbb{1}) \geq \bar{x}^T A^T \bar{y} = \bar{y}^T A \bar{x},$$

d. h. der optimale Wert des Sattelpunktproblems wird in (\bar{x}, \bar{y}) angenommen. Sei umgekehrt $(\bar{x}, \bar{y}) \in X \times Y$ ein Sattelpunkt mit Wert $\bar{y}^T A \bar{x} = \bar{v}$. Dann sind insbesondere (\bar{x}, \bar{v}) zulässig für (P_J) und (\bar{y}, \bar{v}) zulässig für (P_I) . Weiterhin gilt

$$(4.1) \quad \begin{cases} \bar{y}^T (A \bar{x} - v \mathbb{1}) = \bar{y}^T A \bar{x} - \bar{v} = 0, \\ \bar{x}^T (v \mathbb{1} - A^T \bar{y}) = \bar{v} - \bar{x}^T A^T \bar{y} = 0. \end{cases}$$

Wegen der Zulässigkeit von (\bar{x}, \bar{v}) und (\bar{y}, \bar{v}) sind deshalb die Komplementaritätsbedingungen aus Folgerung 2.8 erfüllt, und damit (\bar{x}, \bar{v}) Lösung von (P_J) und (\bar{y}, \bar{v}) Lösung von (P_I) . \square

Daraus folgt sofort der *Hauptsatz der Spieltheorie*.

Folgerung 4.6. Für jedes Matrixspiel (m, n, A) existiert ein Paar optimaler Strategien (\bar{x}, \bar{y}) , die gleichzeitig für (M_J) und für (M_I) optimal sind.

Damit können wir den Wert eines Matrixspiels (m, n, A) eindeutig definieren als

$$\bar{v} = \min_{y \in Y} \max_{x \in X} y^T A x = \max_{x \in X} \min_{y \in Y} y^T A x.$$

Ist $\bar{v} = 0$, so heisst das Spiel *fair* oder *Nullsummenspiel*. Der Wert eines Spiels gibt also die erwartete Auszahlung bei Anwendung optimaler Strategien an; er lässt sich auch auffassen als Gebühr, die J an I zahlen muss, damit dieser überhaupt mitspielt. Die Komplementaritätsbedingungen (4.1) lassen sich so interpretieren, dass I eine Aktion i sicher nicht spielen wird (d. h. $\bar{y}_i = 0$), wenn dies bei optimaler Strategie von J zu einer suboptimalen Auszahlung (d. h. $A_i \bar{x} > \bar{v}$) führen würde.

Eine andere Sichtweise für die Aussage von Satz 4.5 ist die folgende: Ein Sattelpunkt zeichnet sich dadurch aus, dass es sich für J nicht lohnt, bei fester Wahl von \bar{y} von der Strategie \bar{x}

abzuweichen (da sich sonst die Auszahlung an ihn verringern würde). Umgekehrt lohnt es sich für I nicht, bei fester Wahl von \bar{x} die Strategie \bar{y} zu verlassen (sonst würde sich sein Verlust, d. h. Auszahlung an J vergrößern). Mathematisch formuliert gilt also

$$\bar{y}^T A x \leq \bar{y}^T A \bar{x} \leq y^T A \bar{x} \quad \text{für alle } x \in X, y \in Y.$$

Ein Paar (\bar{x}, \bar{y}) , dass diese Bedingung erfüllt, nennt man *Nash-Gleichgewicht*.² Satz 4.5 sagt also aus, dass jedes Matrixspiel ein Nash-Gleichgewicht besitzt.

Folgerung 4.7. Sei (m, n, A) ein Matrixspiel. Ein Paar (\bar{x}, \bar{y}) ist genau dann ein Nash-Gleichgewicht mit Wert \bar{v} , wenn (\bar{x}, \bar{v}) Lösung von (P_I) und (\bar{y}, \bar{v}) Lösung von (P_{II}) ist.

Beweis. Sind (\bar{x}, \bar{v}) und (\bar{y}, \bar{v}) Lösungen, so gilt für $x \in X$ und $y \in Y$, dass

$$y^T A \bar{x} \geq \bar{v}(y^T \mathbf{1}) = \bar{v} = \bar{v}(x^T \mathbf{1}) \geq x^T A^T \bar{y} = \bar{y}^T A x.$$

Nach Satz 4.5 ist $\bar{v} = \bar{y}^T A \bar{x}$ und daher (\bar{x}, \bar{y}) ein Nash-Gleichgewicht. Umgekehrt gilt für alle $x \in X$ und $y \in Y$, dass

$$\max_{x \in X} y^T A x \geq y^T A x \geq \min_{y \in Y} y^T A x.$$

Nimmt man das Minimum über alle y in der linken Ungleichung und das Maximum über alle x in der rechten, so erhält man

$$\min_{y \in Y} \max_{x \in X} y^T A x \geq y^T A x \geq \max_{x \in X} \min_{y \in Y} y^T A x.$$

Geht man analog in der Definition eines Nash-Gleichgewichts (\bar{x}, \bar{y}) vor, erhält man

$$\min_{y \in Y} \max_{x \in X} y^T A x \leq \bar{y}^T A \bar{x} \leq \max_{x \in X} \min_{y \in Y} y^T A x.$$

Also gilt Gleichheit, und aus Satz 4.5 folgt die Charakterisierung als Lösung der Optimierungsprobleme. \square

Zurück zu Beispiel 4.4: Hier verifiziert man leicht, dass $(\bar{x}, \bar{v}) = \frac{1}{11}(6, 5, 0, 2)^T$ und $(\bar{y}, \bar{v}) = \frac{1}{11}(7, 4, 2)^T$ die Komplementaritätsbedingungen (4.1) erfüllen und damit ein Nash-Gleichgewicht für dieses Matrixspiel bilden. Mit etwas mehr Aufwand zeigt man, dass dies tatsächlich die einzigen zulässigen Lösungen der Komplementaritätsbedingungen sind. Es ist also in der Tat optimal für J, die Aktion 3 nie zu wählen.

Zum Abschluss analysieren wir noch „Schere, Stein, Papier“. Dafür ist das folgende Resultat nützlich.

²Nach John Forbes Nash, der Satz 4.7 – für Spiele mit mehr als zwei Spielern und ohne Nullsummenbedingung – im Rahmen seiner Doktorarbeit 1949 bewies und dafür 1994 den Nobelpreis erhielt. Tatsächlich liegt der Wert des Begriff des Nash-Gleichgewichts liegt darin, dass er – anders als der des Paares optimaler Strategien – sinnvoll auf mehr als zwei Spieler verallgemeinert werden kann. Für den Beweis der Existenz von Gleichgewichten benötigt man jedoch andere Techniken; Nash selbst verwendete ein Fixpunktargument, nachzulesen in [Nash 1950] (auf nicht mal einer Seite).

Lemma 4.8. Sei (n, n, A) ein Matrixspiel mit Wert \bar{v} . Ist A schiefssymmetrisch (d. h. $A^T = -A$), dann gilt:

(i) $x^T Ax = 0$ für alle $x \in X$;

(ii) $\bar{v} = 0$;

(iii) für $\bar{x} \in X$ mit $A\bar{x} \geq 0$ ist (\bar{x}, \bar{x}) ein Nash-Gleichgewicht.

Beweis. Zu (i): Ist A schiefssymmetrisch, so ist $A_{ij} = -A_{ji}$ und insbesondere $A_{ii} = 0$ für alle $1 \leq i, j \leq n$. Also ist

$$x^T Ax = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j = \sum_{i=1}^n \sum_{j=i+1}^n (A_{ij} + A_{ji}) x_i x_j = 0.$$

Zu (ii): Sei (\bar{x}, \bar{y}) ein Paar optimaler Strategien, d. h. $\bar{y}^T A\bar{x} = \bar{v}$. Dann ist wegen (i)

$$0 = \bar{x}^T A\bar{x} \geq \bar{v} \geq \bar{y}^T A\bar{y} = 0.$$

Zu (iii): Sei $\bar{x} \in X$ mit $A\bar{x} \geq 0 = \bar{v}$. Dann ist (\bar{x}, \bar{v}) zulässig sowohl für (P_I) als auch für (P_{II}) und wegen Übereinstimmung des Werts sogar optimal. Also ist (\bar{x}, \bar{x}) nach Folgerung 4.7 ein Nash-Gleichgewicht. \square

Spiele mit schiefssymmetrischer Matrix sind also fair, und beide Spieler haben die gleiche optimale Strategie. Solche Spiele werden daher (kontraintuitiv) *symmetrisch* genannt. Da „Schere, Stein, Papier“ symmetrisch ist, ist der Wert somit Null. Die Bedingung $A\bar{x} \geq 0$ liefert die drei Ungleichungen

$$\bar{x}_3 \geq \bar{x}_2 \geq \bar{x}_1 \geq \bar{x}_3,$$

was zusammen mit $\bar{x}_1 + \bar{x}_2 + \bar{x}_3 = 1$ als Nash-Gleichgewicht für beide Spieler die Strategie $\bar{y} = \bar{x} = (1/3, 1/3, 1/3)^T$ liefert. Es ist also für keinen Spieler vorteilhaft, eine der Aktionen zu bevorzugen, da das sofort vom Gegner ausgenutzt werden könnte.³

³All das gilt natürlich nur unter der Voraussetzung, dass die konkrete Wahl der Aktion in einem Spiel wirklich zufällig und insbesondere unabhängig von dem Ergebnis des vorherigen Spiels ist – eine Annahme, die in der Praxis selten erfüllt ist, siehe [Wang u. a. 2014].

DISKRETE TOMOGRAPHIE

5

In der Computertomographie durchleuchtet man einen Patienten mit Röntgenstrahlen und misst die Strahlungsintensität nach dem Durchgang, um ein Absorptionsprofil in dieser Richtung zu bestimmen. (Dies entspricht der Projektion des räumlich variablen Absorptionskoeffizienten im Gewebe auf eine Ebene hinter dem Körper.) Wiederholt man dies für eine genügend große Zahl verschiedener Richtungen, kann man dadurch die räumliche Verteilung des Absorptionskoeffizienten (der sich mit verschiedenen Gewebetypen ändert) im Inneren des Körpers rekonstruieren. Hier betrachten wir ein diskretes Problem, in dem statt eines variablen Koeffizienten diskrete Punkte auf einem Gitter sitzen, die entlang eines „Strahles“ abgezählt werden. Solche Situationen treten zum Beispiel bei der Bestimmung von atomaren Kristallgittern durch Transmissionselektronenmikroskopie auf. Die Frage ist dann, unter welchen Bedingungen an die Strahlen dadurch eindeutig bestimmt werden kann, auf welchen Gitterpunkten Atome sitzen. In dieser diskreten Situation kann dieses Problem als lineares Optimierungsproblem formuliert und die Frage mit Hilfe der Dualität beantwortet werden.

Wir formulieren das Problem zuerst allgemein. Sei $d \in \mathbb{N}$, $d > 1$ (die Raumdimension) und F^* eine *endliche* Teilmenge von \mathbb{R}^d (die Positionen der Atome in einer gegebenen Probe). Sei $m \in \mathbb{N}$, $m > 1$ (die Anzahl der Strahlen, die wir durch die Probe schicken) und für $i \in \{1, \dots, m\}$ sei $s_i \in \mathbb{R}^d \setminus \{0\}$ (die Richtung des Strahls i), $S_i := \{\lambda s_i : \lambda \in \mathbb{R}\}$ (die entsprechende Gerade), und

$$\mathcal{T}_i = \{x + S_i : x \in \mathbb{R}^d \text{ mit } (x + S_i) \cap F^* \neq \emptyset\}$$

(die Menge aller Strahlen in Richtung s_i , auf denen zumindest ein Punkt von F^* liegt). Wir setzen voraus, dass die s_i paarweise verschieden sind. Setze nun

$$\mathcal{T} := \bigcup_{i=1}^m \mathcal{T}_i$$

(die Menge aller Strahlen, auf denen ein Punkt von F^* liegt) und

$$G := \bigcap_{i=1}^m \bigcup_{T \in \mathcal{T}_i} T$$

(die Menge aller Schnittpunkte der Strahlen). Wir nennen $G = G(F^*)$ das durch F^* bestimmte *tomographische Raster*. Nach Konstruktion ist dann $F^* \subset G$ (da durch jeden Punkt f in F^* ja ein $T_i \in \mathcal{T}_i$ für alle $i \in \{1, \dots, m\}$ geht und daher f Schnittpunkt dieser T_i sein muss). Schliesslich sei

$$\beta_{F^*} : \mathcal{T} \rightarrow \mathbb{N}, \quad \beta_{F^*}(T) = |F^* \cap T|$$

(die Anzahl der Punkte in F^* , die auf dem Strahl T liegen) unsere *Messung*. Die Frage ist nun, ob es eine weitere, von F^* verschiedene, endliche Menge $F \subset G$ geben kann, für die $\beta_{F^*} = \beta_F$ ist. Solche Mengen nennen wir *tomographisch äquivalent*.

Durch Beschränkung auf eine endliche Menge können wir diese Frage als lineares Optimierungsproblem formulieren. Dazu identifizieren wir eine endliche Teilmenge $F \subset G$ mit ihrem *Inzidenzvektor* $x \in \{0, 1\}^{|G|}$ gemäß

$$x_g = \begin{cases} 1 & g \in F, \\ 0 & g \notin F. \end{cases}$$

Insbesondere sei x^* der Inzidenzvektor zu F^* . Ist F nun tomographisch äquivalent zu F^* (stimmen also die Messungen überein), so muss

$$(5.1) \quad \begin{cases} \sum_{g \in G \cap T} x_g = \beta_{F^*}(T) & \text{für alle } T \in \mathcal{T}, \\ x_g \in \{0, 1\} & \text{für alle } g \in G \end{cases}$$

gelten. Um dies in Matrixform zu bringen, nummerieren wir die Strahlen von 1 bis $p := |\mathcal{T}|$ und die Gitterpunkte von 1 bis $q := |G|$ durch, und definieren $A \in \mathbb{R}^{p \times q}$ und $b \in \mathbb{R}^p$ durch

$$A_{ij} = \begin{cases} 1 & g_j \in T_i, \\ 0 & g_j \notin T_i, \end{cases} \quad b_i = \beta_{F^*}(T_i).$$

Dann ist (5.1) äquivalent zu

$$Ax = b, \quad 0 \leq x \leq \mathbb{1}, \quad x \in \mathbb{N}_0^q.$$

Es gibt also genau dann eine zu F^* tomographisch äquivalente Menge, wenn das Polytop

$$P := \{x \in \mathbb{R}^q : Ax = b, 0 \leq x \leq \mathbb{1}\}$$

einen von x^* verschiedenen Punkt mit ganzzahligen Koeffizienten enthält.

Das folgende Kriterium stellt eine Verbindung zur kontinuierlichen Tomographie dar: Eine endliche Menge F^* heisst *additiv*, wenn es ein $\eta : \mathcal{T} \rightarrow \mathbb{R}$ gibt, so dass für alle $g \in G$ gilt

$$\sum_{\substack{T \in \mathcal{T} \\ \text{mit } g \in T}} \eta(T) \begin{cases} \geq 1 & g \in F^*, \\ \leq -1 & g \notin F^*. \end{cases}$$

In diesem Fall gilt nämlich für $\mu(T) := \eta(T)/\beta_{F^*}(T)$, dass $g \in F^*$ ist genau dann, wenn

$$\sum_{\substack{T \in \mathcal{T} \\ \text{mit } g \in T}} \mu(T) \beta_{F^*}(T) \geq 1.$$

Dies kann man so interpretieren: Geeignet gewichtet, kann man die Messdaten für jeden Strahl T auf die auf ihm liegenden Gitterpunkte g verteilen, so dass in jedem Punkt g die Summe aller beitragenden Messdaten größer als der Schwellwert 1 ist dann und nur dann, wenn $g \in F^*$. Ein ähnliches Konzept der (gewichteten) *gefilterten Rückprojektion* ist auch die Grundlage der Bildgewinnung in der klassischen Computertomographie.

Man sollte nun annehmen, dass es von der spezifischen Wahl der Gewichte abhängt, ob man die korrekte Lösung bekommt. Auch die Einschränkung auf ganzzahlige Lösungen scheint problematisch. Tatsächlich ist dies (im Diskreten) nicht der Fall: Die Existenz *eines* solchen Gewichts ist bereits ausreichend zu garantieren, dass P keine von x^* verschiedenen Punkte enthält.

Satz 5.1. *Es gilt $P = \{x^*\}$ genau dann, wenn F^* additiv ist.*

Beweis. Um den Zusammenhang mit der Additivität zu sehen, transformieren wir zuerst das Problem von $\{0, 1\}^q$ auf $\{-1, 1\}^q$. Definiere dafür $c \in \{-1, 1\}^q$ durch

$$c_j = 2x_j^* - 1, \quad 1 \leq j \leq q.$$

Sei nun $x \in P \setminus \{x^*\}$. Dann existieren $j \in \{1, \dots, q\}$ mit $x_j \neq x_j^*$, für die gilt

$$c_j(x_j^* - x_j) = (2x_j^* - 1)(x_j^* - x_j) = \begin{cases} (-1)(-1) & x_j^* = 0, \\ (1)(1) & x_j^* = 1. \end{cases}$$

Also ist $c_j(x_j^* - x_j) > 0$ und damit

$$c^T x < c^T x^*.$$

Da Gleichheit nur für $x = x^*$ gilt, ist $P = \{x^*\}$ genau dann, wenn

$$(5.2) \quad \min_x c^T x = c^T x^*.$$

Ebenso können wir die Additivität mit Hilfe von A ausdrücken: Es gibt ein $y^* \in \mathbb{R}^p$ (dessen Einträge genau die Gewichte $\eta(T_i)$ sind), so dass für alle $j \in \{1, \dots, q\}$ gilt

$$(5.3) \quad (A^T)_j y^* \begin{cases} \geq 1 & x_j^* = 1, \\ \leq -1 & x_j^* = 0. \end{cases}$$

Wir zeigen nun mit Hilfe von Dualität, dass x^* die Bedingung (5.2) erfüllt genau dann, wenn ein y^* mit (5.3) existiert. Betrachte dafür das lineare Optimierungsproblem

$$(5.4) \quad \begin{cases} \min_x c^T x \\ \text{mit } Ax = b, \\ x \leq 1, \\ x \geq 0, \end{cases}$$

sowie das dazu duale Problem

$$(5.5) \quad \begin{cases} \max_{y,z} b^T y - \mathbb{1}^T z, \\ \text{mit } A^T y - z \leq c, \\ z \geq 0. \end{cases}$$

Da (5.4) eine Lösung besitzt (der Inzidenzvektor x^* ist zulässig und die Nebenbedingungen bilden ein Polytop, also ist $c^T x$ beschränkt), hat nach Satz 2.4 auch (5.5) eine Lösung, und nach Satz 2.7 müssen alle Paare (x, y) von Lösungen die Komplementaritätsbedingungen

$$(5.6) \quad \begin{cases} x_j(c_j - (A^T)_j y + z_j) = 0 & \text{für alle } 1 \leq j \leq q, \\ z_j(1 - x_j) = 0 & \text{für alle } 1 \leq j \leq q, \end{cases}$$

erfüllen.

Sei zuerst F^* additiv, d. h. es existiert ein y^* mit (5.3). Wir zeigen, dass der Inzidenzvektor x^* zu F^* eine Lösung von (5.4) ist – und daher (5.2) gilt – indem wir nachweisen, dass (x^*, y^*) die Komplementaritätsbedingungen erfüllen. Dafür konstruieren wir $z^* \in \mathbb{R}^q$ durch

$$z_j^* = \begin{cases} (A^T)_i y^* - c_j & x_j^* = 1, \\ 0 & x_j^* = 0. \end{cases}$$

Nach Konstruktion sind damit die Komplementaritätsbedingungen (5.6) erfüllt; es bleibt zu zeigen, dass (y^*, z^*) zulässig ist für (5.5) ($x^* \in \{0, 1\}^q$ ist zulässig, da b genau die Messdaten zu x^* sind): Für alle j mit $x_j^* = 1$ haben wir $(A^T)_i y^* - z_j^* = c_j$ und

$$z_j^* = (A^T)_i y^* - c_j = (A^T)_i y^* - (2x_j^* - 1) \geq 1 + (1 - 2x_j^*) = 0.$$

Für alle j mit $x_j^* = 0$ gilt wegen der Additivität

$$(A^T)_i y^* - z_j = (A^T)_i y^* \leq -1 = (2x_j^* - 1) = c_j$$

sowie nach Definition $z_j^* = 0$. Nach Satz 2.7 ist also x^* Lösung von (5.4).

Sei umgekehrt x^* Lösung von (5.4) und (y^*, z^*) Lösung von (5.5). Dann sind beide zulässig und erfüllen die Komplementaritätsbedingungen (5.6). Betrachte zuerst j mit $x_j^* = 0$ beliebig. Dann muss $z_j^* = 0$ sein, und aus der dualen Zulässigkeit folgt

$$(A^T)_i y^* - z_j^* = (A^T)_i y^* \leq c_j = 2x_j^* - 1 = -1.$$

Ist dagegen j mit $x_j^* = 1$, so folgt analog aus Komplementarität und Zulässigkeit

$$0 = c_j - (A^T)_i y^* + z_j = 2x_j^* - 1 - (A^T)_i y^* + z_j \geq 1 - (A^T)_i y^*$$

und damit $(A^T)_i y^* \geq 1$. Also gilt (5.3) und damit ist F^* additiv. \square

Damit folgt sofort das gewünschte Eindeutigkeitsresultat.

Folgerung 5.2. *Ist F^* additiv und $\beta_F(T) = \beta_{F^*}(T)$ für alle $T \in \mathcal{T}$, so ist $F = F^*$.*

NETZWERKPROBLEME

6

Wir betrachten nun ein weiteres diskretes Optimierungsproblem: Bestimmen des maximalen Flusses durch ein Netzwerk. Die ursprüngliche Motivation war die folgende:

Betrachte ein Schienennetz, welches zwei Städte über eine gewisse Anzahl dazwischen liegender Städte verbindet, wobei jeder Verbindung eine Zahl zugeordnet ist, die ihrer Kapazität entspricht. Finde, unter der Annahme eines stationären Zustands, den maximalen Fluss von einer der gegebenen Städte zur jeweils anderen. (übers. aus [Ford und Fulkerson 1956])

Eine modernere Anwendung ist das Routing in Computer-Netzwerken.

Wir formulieren das Problem nun in graphentheoretischer Form. Ein *gerichteter Graph* ist ein Paar $G = (V, K)$, wobei V eine endliche Menge (der *Knoten*) und K eine Teilmenge der Menge aller geordneten Paare (u, v) mit $u, v \in V$ (der *Kanten*, wobei (u, v) der Kante entspricht, die u mit v verbindet). Beachte, dass $(u, v) \neq (v, u)$ und insbesondere für $(u, v) \in K$ nicht unbedingt auch $(v, u) \in K$ ist. Jeder Kante $k_j = (u, v) \in K$ sei eine *Kapazität* $c_j \geq 0$ zugeordnet (die maximale Transportmenge von u nach v). Weiter seien zwei Knoten $s \in V$ (die *Quelle*) und $t \in V$ (die *Senke*) ausgezeichnet. Wir suchen nun den maximalen Fluss von s nach t entlang der Kanten von G , wobei die Kapazität jeder Kante nicht überschritten werden darf und auf jedem dabei durchquerten Knoten v die *Flussbedingung* gilt: Die Summe der Ein-Flüsse (über Kanten (p, q) mit $q = v$) muss gleich der der Aus-Flüsse (über Kanten (p, q) mit $p = v$) sein. Aus Buchhaltungs-Gründen führen wir noch eine zusätzliche Kante (t, s) ein. Fordert man nun, dass auch in s und t die Erhaltungsbedingung gilt, so entspricht der Fluss entlang (t, s) genau dem Gesamtfluss von s nach t durch das Netzwerk. Dieser Kante ordnen wir eine Kapazität zu, die groß genug ist, den Gesamtfluss nicht (künstlich) zu begrenzen (etwa größer als die Summe aller anderen Kapazitäten).

Um dies als lineares Optimierungsproblem zu formulieren, definieren wir schliesslich für $G = (V, K)$, $V = \{v_1, \dots, v_m\}$, $K = \{k_1, \dots, k_n\}$, noch die *Inzidenzmatrix* $A \in \mathbb{R}^{m \times n}$ mit

$$A_{ij} = \begin{cases} -1 & k_j = (v_i, q) \text{ für ein } q \in V, \\ 1 & k_j = (p, v_i) \text{ für ein } p \in V, \\ 0 & \text{sonst.} \end{cases}$$

Jedem Knoten von A entspricht also eine Zeile, in der die auslaufenden Kanten mit -1 und die einlaufenden Kanten mit 1 markiert sind. Wir vereinbaren, dass die erste Zeile dabei dem Knoten $s =: v_1$ und die zweite dem Knoten $t =: v_2$ zugeordnet sind. Umgekehrt entspricht jede Spalte von A einer Kante $k_j = (p, q)$, auf der p mit 1 und q mit -1 markiert sind (also nur zwei Einträge von 0 verschieden sind). Wir vereinbaren, dass die erste Spalte der „Buchhaltungskante“ $(t, s) =: k_1$ zugeordnet ist. Führt man für jede Kante k_j die auf ihr transportierte Menge $x_j \geq 0$ ein, so kann man die Flussbedingung algebraisch als $Ax = 0$ formulieren. Ein Fluss $x \geq 0$ mit $Ax = 0$ heisst (s, t) -Fluss. Der Gesamtfluss ist als gleich dem Fluss entlang $k_1 = (t, s)$, d. h. gleich $x_1 = e_1^T x$. Führen wir den Kapazitätsvektor $c \in \mathbb{R}^n$ mit $c \geq 0$ und $c_1 > \sum_{j=2}^n c_j$ ein, so erhalten wir das lineare Optimierungsproblem

$$(6.1) \quad \begin{cases} \max_{x \in \mathbb{R}^n} e_1^T x, \\ \text{mit } Ax = 0, \\ x \leq c, \\ x \geq 0. \end{cases}$$

Zu jedem linearen Problem gehört ein duales Problem; in diesem Fall ist dies

$$(6.2) \quad \begin{cases} \min_{\substack{z \in \mathbb{R}^m \\ u, w \in \mathbb{R}^m}} c^T u, \\ \text{mit } A^T z + u - w = e_1, \\ u, w \geq 0. \end{cases}$$

Es soll also eine geeignet gewichtete Summe der Kapazitäten minimiert werden. Ein erstaunliches Resultat ist nun, dass auch dieses Problem eine interessante graphentheoretische Interpretation hat: als *minimalen Schnitt* des Netzwerkes. Dabei ist ein *Schnitt* eines Graphen $G = (V, K)$ bezüglich einer Menge $W \subset V$ definiert als

$$\delta(W) := \{(p, q) \in K : p \in W, q \notin W\}.$$

Gilt $s \in W$ und $t \notin W$, so heisst $\delta(W)$ ein (s, t) -Schnitt. Anschaulich zerlegt ein (s, t) -Schnitt einen Graphen durch „Zerschneiden“ aller Kanten in D , so dass kein Fluss von s nach t mehr möglich ist. Der minimale Schnitt erreicht mit möglichst geringem Aufwand, d. h. er minimiert über alle W die *Kapazität*

$$\kappa(W) := \sum_{\substack{k_j = (p, q) \in K \\ p \in W, q \notin W}} c_j$$

des zugehörigen Schnittes. (Wir werden später sehen, warum es reicht, hier nur die aus W herausführenden Kanten zu betrachten.) Vergleicht man diese Definition mit (6.2), so stellt man fest, dass hier nur über alle $u \in \{0, 1\}^n$ minimiert wird.

Solche Fragestellungen treten häufig in der Bildverarbeitung auf, etwa bei der *Segmentierung* von Bildern. Dabei soll ein Bild zerlegt werden in zwei Regionen (z. B. Vordergrund und

Hintergrund in Stereobildern, oder Tumor und Organ in der Computertomographie). Dafür modelliert man ein (Digital-)Bild als Pixelmatrix, d. h. man ordnet jedem Pixel (i, j) , $1 \leq i \leq N$, $1 \leq j \leq M$, seinen Grauwert u_{ij} zu. Als Graphen nimmt man dann für V die Menge aller Pixel und für K die Menge

$$\{(p, q) : p = (i, j), q = (i', j') \text{ mit } i' - i = 1 \text{ oder } j' - j = 1\}$$

(d. h. jedes Pixel bekommt eine gerichtete Kante zu seinem rechten oder unteren Nachbarn). Zusätzlich führt man die Knoten s und t ein, sowie Kanten von s zu jedem $v \in V$ und von jedem $v \in V$ zu t . Ein (s, t) -Schnitt zerlegt die Menge der Pixel dann in eine Menge W , die durch unzer�rennte Kanten mit s verbunden sind – diese interpretiert man z. B. als Vordergrund – und eine Menge $V \setminus W$, die durch unzer�rennte Kanten mit t verbunden sind (die dann als Hintergrund interpretiert werden). Ein minimaler Schnitt soll dabei Pixel mit ähnlichem Grauwert zur gleichen Menge zuordnen; je ähnlicher der Grauwert benachbarter Pixel, desto höher also die Kapazität der verbindenden Kante; man wählt zum Beispiel $c_{(p,q)} = \exp(-(u_p - u_q)^2)$. Vorgaben, welche Pixel Vorder- bzw. Hintergrund sein sollten, berücksichtigt man durch besonders hohe Kapazitäten der entsprechenden Kanten von s bzw. von t .

Wir zeigen nun (analog zur schwachen Dualität), dass die Kapazität eines Schnitts nie kleiner als der Gesamtfluss durch das Netzwerk sein kann.

Lemma 6.1. Sei $G = (V, K)$ ein gerichteter Graph mit Kapazität $c \geq 0$, $x \in \mathbb{R}^n$ ein (s, t) -Fluss und $\delta(W)$ ein (s, t) -Schnitt. Dann gilt:

$$(i) \quad \sum_{\substack{k_j=(p,q) \in K \\ p \in W, q \notin W}} x_j - \sum_{\substack{k_j=(p,q) \in K \setminus \{k_1\} \\ p \notin W, q \in W}} x_j = e_1^T x,$$

$$(ii) \quad e_1^T x \leq \kappa(W).$$

Beweis. Sei x ein (s, t) -Fluss, d. h. es gilt $Ax = 0$. Wir summieren nun alle Zeilen von $Ax = 0$, die einem Knoten in $V \setminus W$ entsprechen, und sortieren die Beiträge nach Kanten: Kanten, die zwei Knoten in $V \setminus W$ verbinden, kommen zweimal vor (einmal für ihren Aus-Fluss- und einmal für ihren Ein-Fluss-Knoten), und zwar mit positivem bzw. negativem Vorzeichen. Kanten in $\delta(W)$ kommen dagegen nur einmal vor, wobei das Vorzeichen davon abhängt, ob der Aus- oder der Ein-Fluss-Knoten in $V \setminus W$ liegt. Wir erhalten also

$$0 = \sum_{\substack{k_i=(p,q) \notin K \\ p \in W, q \notin W}} (x_i - x_i) + \sum_{\substack{k_i=(p,q) \in K \\ p \in W, q \notin W}} x_i + \sum_{\substack{k_i=(p,q) \in K \\ p \notin W, q \in W}} (-x_i).$$

Ziehen wir aus der letzten Summe den Term $x_1 = e_1^T x$ für die Kante, die aus $V \setminus W$ nach s führt, heraus, so folgt (i). Schätzen wir x_j in der ersten Summe nach oben durch c_j und in der letzten Summe nach unten durch 0 ab, so folgt (ii). \square

Der minimale Schnitt ist dann dadurch charakterisiert, dass in (ii) Gleichheit angenommen wird.

Satz 6.2 (Max-Flow–Min-Cut¹). *Der Gesamtfluss eines maximalen (s, t) -Flusses ist gleich der Kapazität eines minimalen (s, t) -Schnittes.*

Beweis. Das Maximal-Fluss-Problem (6.1) ist zulässig (z. B. ist $x = 0$ zulässiger Punkt) und beschränkt (da $0 \leq x \leq c$), hat also eine Lösung \bar{x} . Nach Satz 2.4 hat also auch das duale Problem (6.2) eine Lösung $(\bar{u}, \bar{w}, \bar{z})$. Wir konstruieren nun aus der dualen Lösung einen Schnitt $\delta(W)$ mit $e_1^T \bar{x} = c^T \bar{u} = \kappa(W)$, indem wir für W die Menge aller Knoten v_i nehmen, für die $\bar{z}_i > \tau$ für einen geeigneten Schwellwert $\tau > 0$ gilt. Der Schnitt besteht dann aus den Kanten k_j mit $\bar{u}_j > 0$ (falls k_j Kante *aus* W) oder $\bar{w}_j > 0$ (falls k_j Kante *nach* W). Dafür verwenden wir, dass diese Lösungen nach Folgerung 2.8 die Komplementaritätsbedingungen

$$(6.3) \quad \begin{cases} \bar{u}_j(c_j - \bar{x}_j) = 0 & 1 \leq j \leq n, \\ \bar{x}_j \bar{w}_j = 0 & 1 \leq j \leq n, \end{cases}$$

erfüllen. (Für $c_j \neq 0$ folgt daraus $\bar{u}_j \bar{w}_j = 0$; die obige Wahl der Kanten für den Schnitt ist also eindeutig.)

Wir bestimmen zuerst den Schwellwert τ , indem wir die Kante (t, s) betrachten. Nach Wahl von c_1 können wir $x_1 < c_1$ annehmen, weshalb $u_1 = 0$ sein muss. Aus der ersten Zeile von $A^T \bar{z} + \bar{u} - \bar{w} = e_1$ und $\bar{w} \geq 0$ folgt damit

$$1 = [A^T \bar{z}]_1 - \bar{v}_1 = \bar{z}_1 - \bar{z}_2 - \bar{w}_1 \leq \bar{z}_1 - \bar{z}_2,$$

und deshalb $\bar{z}_1 \geq 1 + \bar{z}_2$. Wir setzen also $\tau := 1 + \bar{z}_2$, d. h.

$$W := \{v_i \in V : \bar{z}_i \geq 1 + \bar{z}_2\}.$$

Dann ist nach Konstruktion $s = v_1 \in W$ und $t = v_2 \notin W$.

Sei nun $k_j = (v_i, v_r) \in \delta(W)$. Wir unterscheiden, ob die Kante aus W heraus- oder in W hineinläuft:

1. Fall: $v_i \in W, v_r \notin W$. Dann ist nach Konstruktion

$$\bar{z}_i \geq 1 + \bar{z}_2 > \bar{z}_r.$$

Aus der j -ten Zeile von $A^T \bar{z} + \bar{u} - \bar{w} = e_1$ und $\bar{w} \geq 0$ folgt damit

$$0 = -\bar{z}_i + \bar{z}_r + \bar{u}_j - \bar{w}_j < \bar{u}_j - \bar{w}_j \leq \bar{u}_r.$$

Also ist $\bar{u}_j > 0$ und aus der ersten Komplementaritätsbedingung in (6.3) folgt $\bar{x}_j = c_j$.

¹Dieser Zusammenhang zwischen maximalen Flüssen und minimalen Schnitten wurde in [Ford und Fulkerson 1956] gezeigt. Erst viel später – nachdem der Bericht, auf den für die eingangs zitierte Motivation verwiesen wurde, nicht mehr der Geheimhaltung unterlag – stellte sich heraus, dass es dabei um das *sowjetische* Schienennetz ging. Da dies zur Zeit des Kalten Krieges war, ging es – anders als der Titel „Maximal flow through a network“ suggeriert – also nicht primär um maximale Flüsse. Und der Begriff des minimalen Schnitts gewinnt eine deutlich konkretere und unfreundlichere Interpretation...

2. Fall: $v_i \notin W, v_r \in W$. Dann ist nach Konstruktion

$$\bar{z}_r \geq 1 + \bar{z}_2 > \bar{z}_i.$$

Dann ist wie oben wegen $\bar{u} \geq 0$

$$0 = -\bar{z}_i + \bar{z}_r + \bar{u}_j - \bar{w}_j > \bar{u}_j - \bar{w}_j \geq -\bar{w}_j.$$

Also ist $\bar{w}_j > 0$ und aus der zweiten Komplementaritätsbedingung in (6.3) folgt $\bar{x}_j = 0$.

Damit erhalten wir nach Lemma 6.1

$$e_1^T \bar{x} = \sum_{\substack{k_j=(p,q) \in K \\ p \in W, q \notin W}} \bar{x}_j - \sum_{\substack{k_j=(p,q) \in K \setminus \{k_1\} \\ p \notin W, q \in W}} \bar{x}_j = \sum_{\substack{k_j=(p,q) \in K \\ p \in W, q \notin W}} c_j = \kappa(W). \quad \square$$

Man kann dieses Resultat auch so interpretieren, dass das duale Problem (6.2) stets eine *binäre* Lösung $\bar{u} \in \{0, 1\}^n$ besitzt. (In der Tat kann man über den im Beweis konstruierten minimalen Schnitt eine komplett binäre Lösung $(\bar{u}, \bar{w}, \bar{z})$ finden.)

Aus dem Beweis folgt insbesondere, dass der minimale Schnitt über Kanten k_j verläuft, in denen der Fluss in Richtung t maximal (d. h. $\bar{x}_j = c_j$) und in Richtung s minimal (d. h. $\bar{x}_j = 0$) ist. Dies ist die Basis für ein Verfahren zur gleichzeitigen Konstruktion maximaler Flüsse und minimaler Schnitte: den *Ford–Fulkerson-Algorithmus*. Die Grundidee dabei ist, einen gegebenen zulässigen (s, t) -Fluss (z. B. $x = 0$) schrittweise so zu modifizieren, dass $x_j = c_j$ für möglichst viele Kanten k_j in Richtung t und $x_j = 0$ für möglichst viele Kanten in Richtung s ist. Dabei müssen natürlich die Flussbedingungen eingehalten werden, so dass man dafür nicht willkürlich Kanten herausgreifen kann. Man geht daher wie folgt vor:

1. Wähle einen beliebigen (ungerichteten) Weg von s nach t entlang von Kanten aus K .
2. Überprüfe, ob der Fluss auf *diesen* Kanten bereits maximal ist (d. h. irgendwo durch die Kapazität beschränkt ist).
3. Ist dies nicht der Fall, so erhöhe den Fluss auf diesen Kanten, indem er um den selben Betrag auf allen Kanten in Richtung t erhöht und auf allen Kanten in Richtung s reduziert wird; hier wählt man natürlich den maximalen Betrag, der von allen Schranken zugelassen wird.

Dies wiederholt man so lange, bis keine Wege mehr gefunden werden können, auf denen eine solche Erhöhung möglich ist. Für Netzwerke mit ganzzahligen Kapazitäten liefert dieses Verfahren nach endlich vielen Schritten einen maximalen Fluss (und damit auch einen maximalen Schnitt).

Um dies zu zeigen, müssen wir die verwendeten Begriffe mathematisch präzisieren. Ein *ungerichteter Weg* in einem Graphen (V, K) ist eine endliche Folge $w = \{v_0, k_1, v_1, k_2, \dots, v_r\}$ von alternierenden Knoten und Kanten mit den beiden Eigenschaften

- (i) Die Knoten sind paarweise verschieden: $v_i \neq v_j$ für $i \neq j$;
- (ii) Jede Kante beginnt und endet mit den benachbarten Knoten: $k_i = (v_{i-1}, v_i)$ (Vorwärtskante) oder $k_i = (v_i, v_{i-1})$ (Rückwärtskante).

Ist $v_0 = s$ und $v_r = t$, so nennt man w einen (s, t) -Weg; dabei schliessen wir aus, dass die „Buchhaltungskante“ (t, s) Teil von w ist. Ein (s, t) -Weg w heisst *erhöhend* für einen Fluss x und einen Kapazitätsvektor c , wenn für alle $k_j \in w$ gilt:

- (i) $x_j < c_j$ falls k_j Vorwärtskante;
- (ii) $x_j > 0$ falls k_j Rückwärtskante.

Damit können wir nun die Korrektheit des oben geschilderten Verfahrens beweisen.

Satz 6.3. *Ein (s, t) -Fluss x ist genau dann maximal, wenn für ihn keine erhöhenden (s, t) -Wege existieren.*

Beweis. Wir zeigen zuerst, dass für einen maximalen Fluss keine erhöhenden Wege existieren können. Angenommen, x ist ein (s, t) -Fluss und w ist ein erhöhender Weg dazu. Dann existieren für alle $k_j \in w$ positive Zahlen

$$\varepsilon_j := \begin{cases} c_j - x_j & k_j \text{ Vorwärtskante,} \\ x_j & k_j \text{ Rückwärtskante.} \end{cases}$$

Wir setzen nun $\varepsilon := \min_j \varepsilon_j > 0$ und für alle $k_j \in w$

$$x'_j = \begin{cases} x_j + \varepsilon & k_j \in w, k_j \text{ Vorwärtskante,} \\ x_j - \varepsilon & k_j \in w, k_j \text{ Rückwärtskante,} \\ x_j & k_j \notin w, j \neq 1, \\ x_j + \varepsilon & j = 1. \end{cases}$$

(Letzteres aus Buchhaltungsgründen, damit der Gesamtfluss erhalten bleibt.) Dann ist nach Konstruktion $0 \leq x'_j \leq c_j$ für alle $1 \leq j \leq n$. Ausserdem sind die Flussbedingungen erfüllt: Für jeden Knoten $v_j \in V$ gilt

1. $v_j \notin w \cup \{s, t\}$: Dann liegt v_j nicht auf dem erhöhenden Weg, es gilt also $x'_i = x_i$ für alle angrenzenden Kanten k_i .
2. $v_j \in w$ mit $k_j = (v_{j-1}, v_j)$ und $k_{j+1} = (v_j, v_{j+1})$: Dann sind k_j und k_{j+1} Vorwärtskanten, d. h.

$$x'_{j+1} - x'_j = (x_{j+1} + \varepsilon) - (x_j + \varepsilon) = x_{j+1} - x_j.$$

Für die restlichen angrenzenden Kanten gilt $x'_i = x_i$.

3. $v_j \in w$ mit $k_j = (v_{j-1}, v_j)$ und $k_{j+1} = (v_{j+1}, v_j)$: Dann ist k_j Vorwärtskante und k_{j+1} Rückwärtskante, d. h.

$$-x'_{j+1} - x'_j = -(x_{j+1} - \varepsilon) - (x_j + \varepsilon) = -x_{j+1} - x_j.$$

Für die restlichen angrenzenden Kanten gilt $x'_i = x_i$.

4. Analog argumentiert man für die restlichen beiden Fälle für $v_j \in w$.
5. $v_j \in \{s, t\}$: Da w ein (s, t) -Weg ist, müssen k_1 und k_r Vorwärtskanten sein. Die Buchhaltungskante (t, s) ist dagegen Rückwärtskante, so dass auch in s und t die Flussdifferenz von x' mit der von x übereinstimmt.

Da x nach Voraussetzung die Flussbedingungen erfüllt, muss das also auch für x' gelten. Damit haben wir einen zulässigen Fluss x' mit Gesamtfluss

$$e_1^T x' = x'_1 = x_1 + \varepsilon > x_1$$

konstruiert. Der Fluss x ist daher nicht maximal.

Angenommen, \bar{x} ist ein (s, t) -Fluss, für den kein erhöhender (s, t) -Weg existiert. Wir setzen

$$W := \{s\} \cup \{v \in V : \text{Es existiert erhöhender } (s, v)\text{-Weg}\}.$$

Dann ist $s \in W$ und $t \notin W$. Weiter gilt für jede Kante $k_j = (p, q)$ mit $p \in W$ und $q \notin W$, dass $\bar{x}_j = c_j$ gilt (da k_j Vorwärtskante ist, könnte man sonst den erhöhenden (s, p) -Weg nach q verlängern); analog ist $\bar{x}_j = 0$ für jede Rückwärtskante $k_j = (p, q)$ mit $p \notin W$ und $q \in W$. Wie im Beweis von Satz 6.2 folgt nun

$$e_1^T \bar{x} = \sum_{\substack{k_j=(p,q) \in K \\ p \in W, q \notin W}} c_j = \kappa(W),$$

woraus nach demselben Satz folgt, dass \bar{x} maximaler Fluss und $\delta(W)$ minimaler Schnitt ist. □

Sind sämtliche Kapazitäten (und der Ausgangsfluss $x = 0$) ganzzahlig, so sind auch alle ε_i stets ganzzahlig. Also führt jeder gefundene erhöhende Weg zu einer echten Erhöhung des Gesamtflusses um $\varepsilon \geq 1$. Da der maximale Gesamtfluss endlich ist, erreichen wir diesen also in endlich vielen Schritten.

Folgerung 6.4. Für $c \in \mathbb{N}^n$ liefert das Ford–Fulkerson-Verfahren nach endlich vielen Schritten einen maximalen Fluss $\bar{x} \in \mathbb{N}^n$.

Das Finden erhöhender Wege ist mit Algorithmen der Graphentheorie effizient durchführbar. In Algorithmus 6.1 ist das komplette Ford–Fulkerson-Verfahren unter Verwendung einer Breitensuche angegeben. Dabei ist U als *Warteschlange* zu verstehen, d. h. wird v_j in Schritt 9 oder Schritt 14 hinzugefügt, so kommt er an das Ende der Schlange, während er in Schritt 4 vom Anfang der Schlange genommen wird.

Algorithmus 6.1 : Ford–Fulkerson-Verfahren**Input** : $(V, K), c \in \mathbb{N}^n$ **Output** : $x \in \mathbb{N}^n, W$

```

1 Setze  $x = 0$  // Initialisierung mit zulässigem Fluss
2 Setze  $W := \{v_1\}, U := \{v_1\}, \varepsilon_1 = \infty, \sigma = 0$  // Suche erhöhenden Weg von  $s = v_1$  . . .
3 repeat
4   for  $v_j \in U$  do // Prüfe Knoten, ob Teil erhöhenden Weges
5     for  $k_i = (v_j, v_r) \in K$  mit  $v_r \notin W$  do // Prüfe alle Vorwärtskanten
6       if  $x_i < c_i$  then // Nicht gesättigt
7         Setze  $\varepsilon_r = \min\{\varepsilon_j, c_i - x_i\}$  // Bislang maximale Erhöhung
8         Setze  $\sigma_r = j$  // Vorwärtskante
9         Setze  $W := W \cup \{v_r\}, U := U \cup \{v_r\}$  //  $v_j$  liegt auf Weg
10      for  $k_i = (q, v_j) \in K$  mit  $q \notin W$  do // Prüfe alle Rückwärtskanten
11        if  $x_i > 0$  then // Nicht gesättigt
12          Setze  $\varepsilon_r = \min\{\varepsilon_j, x_i\}$  // Bislang maximale Erhöhung
13          Setze  $\sigma_r = -j$  // Rückwärtskante
14          Setze  $W := W \cup \{v_r\}, U := U \cup \{v_r\}$  //  $v_j$  liegt auf Weg
15      Setze  $U := U \setminus \{v_j\}$  //  $v_j$  geprüft
16      if  $v_2 \in W$  then // . . . nach  $t = v_2$ 
17        Setze  $r := 2$  // Starten in  $t = v_2$ 
18        repeat // Gehe Weg zurück
19           $j := |\sigma_r|$  // Voriger Knoten auf Weg
20          Finde  $i$  mit  $k_i = (v_j, v_r)$  // Zugehörige Kante
21          if  $j > 0$  then // Vorwärtskante
22            Setze  $x_i := x_i + \varepsilon_2$  // Fluss erhöhen
23          else // Rückwärtskante
24            Setze  $x_i := x_i - \varepsilon_2$  // Fluss verringern
25          Setze  $r := j$  // Weiter zu nächstem Knoten
26        until  $r = 1$  // Ende in  $s = v_1$ 
27        Setze  $U := \{v_1\}, \varepsilon_1 = \infty, \sigma = 0$  // Suche neuen erhöhenden Weg
28 until  $U = \emptyset$  // Keine erhöhenden Wege mehr  $\rightarrow$  Fluss maximal

```

Teil III

NUMERISCHE VERFAHREN

DAS SIMPLEX-VERFAHREN

7

Wir betrachten nun das erste¹ – und immer noch berühmteste² – Verfahren zur Lösung linearer Optimierungsprobleme der Form $\min_{x \in P} c^T x$. Die Grundidee des Simplex-Verfahrens ist es, ausgehend von einer gegebenen Ecke des zulässigen Polyeders P (die immer existiert, wenn $P = P^=(A, b)$ ist) eine neue Ecke zu finden, für die der Wert der Zielfunktion $c^T x$ strikt kleiner ist. Dabei vermeiden wir die historisch übliche Tableau-Form und leiten das (revidierte) Simplex-Verfahren gleich in Matrix-Schreibweise her.

Sei daher ein lineares Optimierungsproblem in *Normalform*

$$(P) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \end{cases}$$

für $A \in \mathbb{R}^{m \times n}$ gegeben. Wir nehmen an, dass $n > m$ ist (sonst hat $Ax = b$ keine oder eine eindeutig bestimmte Lösung, und es gibt nichts zu optimieren), und dass $\text{rang } A = m$ ist (sonst entfernen wir linear abhängige Zeilen). Das zu (P) duale Problem lautet nach Lemma 2.2

$$(D) \quad \begin{cases} \max_{y \in \mathbb{R}^m} b^T y \\ \text{mit } A^T y \leq c. \end{cases}$$

Nach Folgerung 2.8 ist \bar{x} Lösung von (P) und \bar{y} Lösung von (D) – man nennt dann (\bar{x}, \bar{y}) ein *primal-duales Paar* genau dann, wenn \bar{x} und \bar{y} jeweils zulässig sind und die Komplementaritätsbedingungen

$$(7.1) \quad x_j (c_j - [A^T y]_j) = 0 \quad \text{für alle } j = 1, \dots, n$$

erfüllen.

¹Eigentlich das zweite, denn das Simplex-Verfahren wurde 1947 von [George Dantzig](#) während seiner Arbeit für die US Air Force entwickelt; die [Fourier–Motzkin–Elimination](#) ist daher (mindestens) elf Jahre älter. Die Grundidee geht jedoch ebenfalls auf Fourier zurück.

²siehe <https://www.siam.org/pdf/news/637.pdf>

7.1 HERLEITUNG DES VERFAHRENS

Aus Satz 3.10 wissen wir, dass falls (P) lösbar ist, eine Lösung \bar{x} existiert, die Ecke von $P^=(A, b)$ ist. Nach Satz 3.12 sind in diesem Fall die Spalten von A , die zu strikt positiven Komponenten von \bar{x} gehören, linear unabhängig. Dies motiviert die folgende Definition: Ein Vektor $x \in P^=(A, b)$ heisst für $P^=(A, b)$ zulässiger *Basisvektor*, falls $x \in P^=(A, b)$ liegt und eine Indexmenge $B \subset \{1, \dots, n\}$ mit $|B| = m$ (die *Basis*) existiert, so dass gilt:

- (i) $x_j = 0$ für alle $j \notin B$, und
- (ii) die Spalten $A^j, j \in B$, sind linear unabhängig.

Wir schreiben x_B für den Vektor mit Einträgen $x_j, j \in B$ (den *Basisvariablen*) sowie $A_B \in \mathbb{R}^{m \times m}$ für die *Basismatrix* mit den Spalten A^j für $j \in B$, die nach Konstruktion invertierbar ist. Analog definieren wir $N := \{1, \dots, n\} \setminus B$ sowie die *Nichtbasisvariablen* x_N und die *Nichtbasismatrix* A_N . Es gilt dann für alle $z \in \mathbb{R}^n$

$$Az = A_B z_B + A_N z_N,$$

und insbesondere für den zugehörigen Basisvektor x

$$A_B x_B = b, \quad x_N = 0.$$

Wir zeigen nun, dass es sich bei den zulässigen Basisvektoren genau um die Ecken von $P^=(A, b)$ handelt.

Lemma 7.1. *Sei $A \in \mathbb{R}^{m \times n}$ mit $\text{rang}(A) = m < n$. Dann ist $\{x\}$ genau dann Ecke von $P^=(A, b)$, wenn x ein für $P^=(A, b)$ zulässiger Basisvektor ist.*

Beweis. Sei $\{x\} \subset P^=(A, b)$ eine Ecke und $I = \text{supp}(x)$. Dann sind nach Satz 3.12 die Spalten $A^j, j \in I$, linear unabhängig. Wegen $\text{rang } A = m$ gilt $|I| \leq m$. Es ist daher entweder $|I| = m$ (dann sind wir fertig), oder es existieren $m - |I|$ weitere (zueinander und zu A_I) linear unabhängige Spalten A^j . Ergänzen wir I mit diesen j zu B , so ist $A_B \in \mathbb{R}^{m \times m}$ regulär und $x_j = 0$ für $j \notin B$ (da $\text{supp}(x) = I \subset B$). Also ist x ein Basisvektor.

Sei umgekehrt $x \in P^=(A, b)$ ein Basisvektor. Dann existiert eine Basis B mit invertierbarer Basismatrix $A_B \in \mathbb{R}^{m \times m}$ und $x_j = 0$ für $j \notin B$. Also ist $\text{supp}(x) \subset B$, und die Spalten $A^j, j \in \text{supp}(x)$, sind linear unabhängig. Nach Satz 3.12 ist daher $\{x\}$ eine Ecke von $P^=(A, b)$. \square

Sei nun ein ein Basisvektor x mit Basis B gegeben. Unser Ziel ist nun, einen neuen Vektor z zu konstruieren, so dass gilt: (i) $c^T z < c^T x$; (ii) z ist zulässig; (iii) z ist Basisvektor. Dabei nutzen wir aus, dass die Basisvariablen von z durch die Gleichungsnebenbedingungen eindeutig

festgelegt werden: Damit z zulässig ist, muss insbesondere $b = Az = A_B z_B + A_N z_N$ gelten; wegen der Invertierbarkeit von A_B und mit $A_B x_B = b$ bedeutet dies³

$$(7.2) \quad z_B = A_B^{-1} b - A_B^{-1} A_N z_N = x_B - A_B^{-1} A_N z_N.$$

Wir überlegen nun, wie wir durch Wahl von z_N die Bedingung (i) erreichen können. Aus (7.2) und $x_N = 0$ folgt.

$$\begin{aligned} c^T z &= c_B^T z_B + c_N^T z_N \\ &= c_B^T (x_B - A_B^{-1} A_N z_N) + c_N^T z_N \\ &= c_B^T x_B + (c_N^T - c_B^T A_B^{-1} A_N) z_N \\ &= c^T x + (c_N - A_N^T (A_B^T)^{-1} c_B)^T z_N. \end{aligned}$$

Setzen wir nun $y := (A_B^T)^{-1} c_B \in \mathbb{R}^m$, d. h. $y \in \mathbb{R}^m$ löst das lineare Gleichungssystem

$$(7.3) \quad A_B^T y = c_B,$$

und

$$u_N := c_N - A_N^T y$$

(der Übersichtlichkeit halber schleppen wir $u_B = 0$ mit, d. h. $u \in \mathbb{R}^n$), so erhalten wir

$$(7.4) \quad c^T z = c^T x + u_N^T z_N = c^T x + \sum_{j \in N} u_j z_j.$$

Um den Zielfunktionswert zu verkleinern, muss also die Summe auf der rechten Seite negativ sein. Ist nun $u_j \geq 0$ für alle $j \in N$, so müsste dafür mindestens ein z_j negativ sein – aber damit wäre z nicht zulässig. In diesem Falle ist x also bereits der beste zulässige Vektor.

Lemma 7.2. *Gilt $u_N \geq 0$, so ist der zugehörige Basisvektor $x \in \mathbb{R}^n$ eine Lösung von (P) und $y := (A_B^T)^{-1} c_B \in \mathbb{R}^m$ eine Lösung von (D).*

Beweis. Nach Voraussetzung ist x zulässig für (P). Aus $c_N - A_N^T y = u_N \geq 0$ folgt mit (7.3)

$$[A^T y]_j = (A^j)^T y \begin{cases} \leq c_j & j \in N, \\ = c_j & j \in B, \end{cases}$$

und damit Zulässigkeit von y für das duale Problem (D). Insbesondere gilt $c_j - [A^T y]_j = 0$ für alle $j \in B$, und damit sind wegen $x_j = 0$ die Komplementaritätsbedingungen (7.1) erfüllt. Also ist (x, y) nach Satz 2.7 ein primal-duales Paar. \square

³In Folge ist stets $A_B^{-1} = (A_B)^{-1}$ bzw. $A_B^T = (A_B)^T$ zu lesen.

Wir nehmen nun an, es existiert ein $r \in N$ mit $u_r < 0$, und machen für z_N den Ansatz

$$z_j = \begin{cases} t & j = r \\ 0, & j \in N \setminus \{r\}. \end{cases}$$

Für alle $t > 0$ gilt dann

$$c^T z = c^T x + t u_r < c^T x,$$

womit Bedingung (i) erfüllt wäre.

Weiterhin gilt

$$A_N z_N = \sum_{j \in N} z_j A^j = t A^r.$$

Damit z zulässig ist, muss insbesondere $Az = b$ gelten; nach (7.2) bedeutet das, dass wir

$$z_B = x_B - t A_B^{-1} A^r$$

wählen müssen. Führen wir w_B ein als Lösung von

$$(7.5) \quad A_B w_B = A^r$$

(wieder mit $w_N = 0$, d. h. $w \in \mathbb{R}^n$), so lässt sich das schreiben als

$$z_B = x_B - t w_B.$$

Es bleibt noch, $z \geq 0$ zu erreichen. Offensichtlich ist $z_N = t e_r \geq 0$ für alle $t \geq 0$. Ist $w_B \leq 0$, so gilt wegen $x_B \geq 0$ auch $z_B \geq 0$ für alle $t \geq 0$; damit ist $z = z(t)$ für alle $t > 0$ zulässig, aber

$$c^T z(t) = c^T x + t u_r \rightarrow -\infty \quad \text{für } t \rightarrow \infty,$$

das Problem (P) ist also unbeschränkt. Wir halten dies als Lemma fest.

Lemma 7.3. *Gilt $w_i \leq 0$ für alle $i \in B$, so hat das Problem (P) keine Lösung.*

Wir nehmen nun an, es existiert (mindestens) ein $i \in B$ mit $w_i > 0$. Um $z_B = x_B - t w_B \geq 0$ zu gewährleisten, muss also gelten

$$(7.6) \quad t \leq \frac{x_i}{w_i} \quad \text{für alle } i \in B \text{ mit } w_i > 0.$$

Jede Wahl von $t \geq 0$, die (7.6) erfüllt, ergibt also einen zulässigen Vektor z .

Allerdings ist z im Allgemeinen kein Basisvektor, denn für $r \in N$ ist ja $z_r = t > 0$ und $z_B > 0$ möglich; und damit haben wir mehr als die $n - m$ erlaubten Indizes im Träger von z . Wir müssen also t so wählen, dass $z_s = x_s - t w_s = 0$ ist für ein $s \in B$. Wir setzen daher

$$(7.7) \quad t = \min_{i \in B, w_i > 0} \frac{x_i}{w_i} = \frac{x_s}{w_s}.$$

Wir haben damit einen neuen Basisvektor, für den der Index s die Basis verlassen und dafür der neue Index r hereingekommen ist.

Lemma 7.4. *Der Vektor z ist ein Basisvektor von $P^=(A, b)$ mit Basis $B' := (B \cup \{r\}) \setminus \{s\}$.*

Beweis. Nach Konstruktion ist $|B'| = |B| = m$ und $z_j = 0$ für $j \notin B'$. Es bleibt zu zeigen, dass die entsprechenden Spalten linear unabhängig sind. Seien $\alpha_i, i \in B'$, gegeben mit $\sum_{i \in B'} \alpha_i A^i = 0$. Wir zeigen nun $\alpha_i = 0$: Wegen (7.5) ist

$$\begin{aligned} 0 &= \sum_{i \in B'} \alpha_i A^i = \sum_{i \in B, i \neq s} \alpha_i A^i + \alpha_r A^r \\ &= \sum_{i \in B, i \neq s} \alpha_i A^i + \alpha_r A_B w_B \\ &= \sum_{i \in B, i \neq s} \alpha_i A^i + \alpha_r \sum_{i \in B} w_i A^i \\ &= \sum_{i \in B, i \neq s} (\alpha_i + \alpha_r w_i) A^i + \alpha_r w_s A^s. \end{aligned}$$

Da B eine Basis ist, sind nach Voraussetzung die Spalten $A^i, i \in B$, linear unabhängig. Also gilt $\alpha_i + \alpha_r w_i = 0$ für alle $i \in B \setminus \{s\}$ und $\alpha_r w_s = 0$. Wegen $w_s > 0$ muss aber $\alpha_r = 0$ und damit auch $\alpha_i = 0$ für alle $i \in B \setminus \{s\}$ gelten. \square

Wir fassen zusammen: Wählen wir r mit $u_r < 0$ und $t > 0$ beliebig, so ist Bedingung (i) erfüllt; wählen wir s und t nach (7.7), so sind Bedingungen (ii) und (iii) erfüllt. Damit alle drei Bedingungen gleichzeitig gelten, muss daher $x_s/w_s > 0$ sein. Zwar ist stets $x_s \geq 0$ und $w_s > 0$, aber auch für $s \in B$ kann $x_s = 0$ sein (vergleiche den Beweis von Lemma 7.1 im Fall $|\text{supp}(x)| < m$). In diesem Fall nennen wir x einen *entarteten Basisvektor*. Ist x aber nicht entartet (d. h. $x_B > 0$), so ist das Minimum in (7.7) strikt positiv, und wir erhalten eine echte Reduktion im Zielfunktionswert. Setzen wir nun $x = z$ und wiederholen die Prozedur, so ergibt dies das Simplex-Verfahren; die notwendigen Schritte sind in Algorithmus 7.1 zusammengefasst.

Algorithmus 7.1 : Simplex-Verfahren

Input : $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, Basisvektor $x \in P^=(A, b)$ mit Basis B

Output : Primal-duales Paar (\bar{x}, \bar{y}) oder Information „primales Problem unbeschränkt“

```

1 while nicht fertig do
2   Bestimme Nichtbasis  $N$ , Matrizen  $A_B, A_N$ 
3   Löse  $A_B^T y = c_B$ 
4   Setze  $u_N = c_N - A_N^T y$ 
5   if  $u_N \geq 0$  then return  $(x, y)$  //  $x, y$  optimal
6   Wähle  $r \in N$  mit  $u_r < 0$ 
7   Löse  $A_B w_B = A^r$ 
8   if  $w_B \leq 0$  then return nicht lösbar // Problem unbeschränkt
9   Wähle  $t = \min_{i \in B, w_i > 0} \frac{x_i}{w_i} = \frac{x_s}{w_s}$ 
10  Setze  $x_B = x_B - t w_B$ 
       $x_r = t$ 
       $B = (B \cup \{r\}) \setminus \{s\}$ 

```

Satz 7.5. *Sind alle in Algorithmus 7.1 auftretenden Basisvektoren nicht entartet, so bricht das Verfahren nach endlich vielen Schritten ab. Bei Abbruch in Zeile 5 ist der aktuelle Vektor x Lösung von (P) und der in Schritt 3 berechnete Vektor y Lösung von (D); bei Abbruch in Zeile 8 hat (P) keine Lösung.*

Beweis. Bezeichne $\{x^k\}_{k \in \mathbb{N}}$ die Folge der Iterierten im Simplex-Verfahren. Sind alle x^k nicht entartet, so gilt nach Herleitung stets $c^T x^{k+1} < c^T x^k$. Insbesondere ist damit $x^{k+1} \neq x^j$ für alle $j \leq k$; ein Basisvektor wird also höchstens einmal vorkommen. Da es nach Lemma 7.1 nur endlich viele Basisvektoren gibt, muss das Verfahren abbrechen. Die weiteren Aussagen folgen aus Lemma 7.2 und Lemma 7.3. \square

Aus praktischer Hinsicht bleiben aber noch Fragen offen:

1. Wie berechnet man den ursprünglichen Basisvektor, von dem das Verfahren startet?
2. Wie wählt man die Indizes r, s in Schritt 6 und Schritt 9, falls mehrere zur Auswahl stehen?
3. Wie löst man die Gleichungssysteme in Schritt 3 und Schritt 7 effizient?

Hier konzentrieren wir uns auf die ersten beiden Fragen; insbesondere wird die Antwort auf die zweite Frage erlauben, auch bei degenerierten Basisvektoren nach endlich vielen Schritten zum Optimum zu kommen. Die Antwort auf die letzte Frage, die auf die Aktualisierung einer LU- oder QR-Zerlegung einer Matrix bei Modifikation einer Spalte hinausläuft, überlassen wir der numerischen Linearen Algebra; siehe etwa [Golub und Van Loan 2013, Chapter 6.5] und [Geiger und Kanzow 2002, Kapitel 3.4.3].

7.2 FINDEN EINES STARTVEKTORS

Ein zulässiger Basisvektor für Polyeder der Form $P = (A, b)$ ist in der Regel nicht einfach zu erraten; eine Ausnahme sind Polyeder, die aus den Bedingungen $Ax \leq b, x \geq 0$ durch Einführung einer Schlupfvariablen z erzeugt wurden. Ist nämlich $b \geq 0$, so sind diese Bedingungen äquivalent dazu, dass für $A' = (A, I)$ und $x' := (x, z)^T$ gilt

$$A'x' = b, \quad x' \geq 0,$$

und $(x, z) = (0, b)$ ist ein zulässiger Basisvektor mit Basis $B = \{n + 1, \dots, n + m\}$. Für den allgemeinen Fall kann man nun einen Basisvektor berechnen, indem man das Simplex-Verfahren auf ein Hilfsproblem mit diesem Basisvektor anwendet. (Dies wird oft als „Phase I“ des Simplex-Verfahrens bezeichnet; der eigentliche Algorithmus 7.1 ist dann „Phase II“.) Tatsächlich taugt schon das obige Beispiel als Hilfsproblem. Sei wieder angenommen, dass $b \geq 0$ ist (da wir in (P) reine Gleichungsnebenbedingungen betrachten, ist die Annahme

$b \geq 0$ keine Einschränkung, denn wir können Zeilen A_i mit $b_i < 0$ mit -1 multiplizieren, ohne das Problem zu verändern), und betrachte das Hilfsproblem

$$(H) \quad \begin{cases} \min_{\substack{x \in \mathbb{R}^n \\ z \in \mathbb{R}^m}} \mathbb{1}^T z \\ \text{mit } Ax + z = b, \\ x, z \geq 0 \end{cases}$$

Hier ist z keine Schlupfvariable; man spricht von einer *künstlichen Variablen*.

Satz 7.6. Für das Problem (H) gelten folgende Aussagen:

- (i) Der Vektor $(x, z) = (0, b)$ ist zulässiger Basisvektor mit Basis $B = \{n+1, \dots, n+m\}$.
- (ii) Das Problem hat eine Lösung (\bar{x}, \bar{z}) .
- (iii) Ist (\bar{x}, \bar{z}) ein Basisvektor, so ist im Fall $\bar{z} \neq 0$ das ursprüngliche Problem (H) nicht zulässig; andernfalls ist \bar{x} ein für $P=(A, b)$ zulässiger Basisvektor.

Beweis. Aussage (i) folgt wie oben aus $A'_B = I \in \mathbb{R}^{m \times m}$. Ausserdem ist 0 zulässig für das zu (H) duale Problem, und damit folgt Aussage (ii) aus Satz 2.3.

Für Aussage (iii) sei die Lösung (\bar{x}, \bar{z}) ein Basisvektor. Sei $\bar{z} \neq 0$ und \bar{x} ein für (P) zulässiger Vektor. Dann ist aber auch $(x, z) = (\bar{x}, 0)$ zulässig für (H) mit Funktionswert $\mathbb{1}^T z = 0 < \mathbb{1}^T \bar{z}$, im Widerspruch zur Optimalität von (\bar{x}, \bar{z}) . Sei nun $\bar{z} = 0$. Dann ist $A\bar{x} = b$, und die zu positiven Komponenten von \bar{x} gehörenden Spalten sind linear unabhängig. Sind dies m Stück, so sind wir fertig; ansonsten können wir diese wegen $\text{rang } A = m$ zu m linear unabhängigen Spalten ergänzen und erhalten eine Basis für (P). \square

Zu beachten ist, dass für $\bar{z} = 0$ zwar \bar{x} ein Basisvektor für (P) ist, aber die zu (\bar{x}, \bar{z}) gehörige Basis B nicht unbedingt eine Basis für \bar{x} ist (nämlich falls in B Indizes $i > n$ enthalten sind; dies entspricht einer degenerierten Basislösung). Vor Start der zweiten Phase muss also eventuell die Basis neu bestimmt werden. Dies wird in der im Folgenden beschriebenen Methode vermieden; hier wird das ursprüngliche Problem so modifiziert, dass der Start-Basisvektor für (H) auch für dieses Problem ein Basisvektor ist, die Lösung aber (unter bestimmten Voraussetzungen) mit der von (P) übereinstimmt. Wir betrachten für ein gegebenes $\gamma \geq 0$ das Problem

$$(P_\gamma) \quad \begin{cases} \min_{\substack{x \in \mathbb{R}^n \\ z \in \mathbb{R}^m}} c^T x + \gamma \mathbb{1}^T z \\ \text{mit } Ax + z = b, \\ x, z \geq 0 \end{cases}$$

Da die zulässige Menge die gleiche wie für (H) ist, ist natürlich $(0, b)$ wieder ein Basisvektor mit Basis $\{n+1, \dots, n+m\}$. Wir können also das Simplex-Verfahren für $A' = (A, I)$ und $x' = (x, z)^T$ mit diesem Vektor starten, und erhalten (falls das Verfahren abbricht) eine

Lösung (\bar{x}, \bar{z}) . Die überraschende Aussage ist nun, dass für γ gross genug \bar{x} bereits eine Lösung von (P) ist.

Lemma 7.7. *Ist (\bar{x}, \bar{z}) Lösung von (P_γ) mit $\bar{z} = 0$, so ist \bar{x} Lösung von (P).*

Beweis. Ist (\bar{x}, \bar{z}) Lösung mit $\bar{z} = 0$, so ist der Vektor zulässig und erfüllt daher $A\bar{x} = b$. Damit ist \bar{x} auch zulässig für (P). Umgekehrt ist für jeden zulässigen Vektor $x \in P^=(A, b)$ der Vektor $(x, 0)$ zulässig für (P_γ) . Aus der Optimalität von (\bar{x}, \bar{z}) für (P_γ) folgt dann

$$c^T \bar{x} = c^T \bar{x} + \gamma \mathbf{1}^T \bar{z} \leq c^T x + \gamma \mathbf{1}^T 0 = c^T x,$$

d. h. \bar{x} minimiert $c^T x$ über $P^=(A, b)$. □

Satz 7.8. *Es existiert ein $\bar{\gamma} \geq 0$, so dass für alle $\gamma > \bar{\gamma}$ das Problem (P_γ) lösbar ist, und für jede Lösung (\bar{x}, \bar{z}) gilt $\bar{z} = 0$.*

Beweis. Wir betrachten das zu (P_γ) duale Problem (vergleiche (D))

$$\left\{ \begin{array}{l} \max_{y \in \mathbb{R}^m} b^T y \\ \text{mit } \begin{pmatrix} A^T \\ I \end{pmatrix} y \leq \begin{pmatrix} c \\ \gamma \mathbf{1} \end{pmatrix}, \end{array} \right.$$

beziehungsweise

$$(D_\gamma) \quad \left\{ \begin{array}{l} \max_{y \in \mathbb{R}^m} b^T y \\ \text{mit } A^T y \leq c \\ y_i \leq \gamma, \quad 1 \leq i \leq m. \end{array} \right.$$

Nun ist (P) lösbar, und deshalb existiert nach Satz 2.4 eine Lösung \bar{y} von (D); diese erfüllt insbesondere $A^T \bar{y} \leq c$. Setzen wir $\bar{\gamma} := \max_i \bar{y}_i$, so ist für alle $\gamma > \bar{\gamma}$ diese Lösung \bar{y} auch zulässig für (D_γ) . Da der zulässige Bereich für (D_γ) kleiner ist als der für (D), kann der optimale Zielfunktionswert $b^T y$ sicher nicht größer sein als $b^T \bar{y}$, d. h. \bar{y} ist auch Lösung von (D_γ) . Wieder nach Satz 2.4 hat deshalb (P_γ) eine Lösung (\bar{x}, \bar{z}) , und diese erfüllt die Komplementaritätsbedingungen

$$\begin{aligned} \bar{x}_j (c_j - [A^T \bar{y}]_j) &= 0, & 1 \leq j \leq n, \\ \bar{z}_i (\gamma - \bar{y}_i) &= 0, & 1 \leq i \leq m. \end{aligned}$$

Da γ so gewählt war, dass $\bar{y}_i \leq \bar{\gamma} < \gamma$ für alle $1 \leq i \leq m$ gilt, folgt aus der zweiten Bedingung $\bar{z} = 0$, was zu zeigen war. □

In der Praxis startet man mit einer hinreichend grossen Zahl γ und erhöht diese z. B. durch Multiplikation mit einem festen Faktor, falls Algorithmus 7.1 mit einer Lösung (\bar{x}, \bar{z}) mit $\bar{z} \neq 0$ (oder wegen Unzulässigkeit) abbricht. (Da γ in den Nebenbedingungen nicht vorkommt, bleibt (\bar{x}, \bar{z}) ein zulässiger Basisvektor, mit dem man den nächsten Durchlauf starten kann.)

7.3 VERMEIDUNG VON ZYKLEN

Wir untersuchen nun die Wahl von r und s in Schritt 6 und Schritt 9. Für beide Schritte ist eine Vielzahl von Strategien vorgeschlagen worden. Beginnen wir mit der Wahl des Index $r \in N$, der neu in die Basis aufgenommen wird. Wir erinnern, dass wir für u_N ein r suchen mit $u_r < 0$, da dann im nächsten Schritt gilt $c^T z = c^T x + u_N^T z_N < c^T x$. Mögliche Strategien sind die

- (a) *Erster-Index-Regel*: wähle r als den ersten Index $j \in N$, für den $u_j < 0$ ist;
- (b) *Kleinster-Index-Regel*: wähle r als den *kleinsten* Index $j \in N$, für den $u_j < 0$ ist (da die Einträge in N nicht sortiert sein müssen, kann diese Regel zu einem anderen Ergebnis als (a) führen);
- (c) *Größter-Abstieg-Regel*: wähle r so, dass $u_r = \min\{u_i : i \in N, u_i < 0\}$ ist;
- (d) *Steilster-Abstieg-Regel*: wähle r so, dass für das entsprechende w aus Schritt 7 der Wert $tu_r/\|w\|$ minimal wird. (Anstatt dem größten Abstieg suchen wir den größten Abstieg pro Änderung von x .)

Diese Regeln sind nach wachsendem Aufwand sortiert; insbesondere kann man für (a) und (b) die Komponenten $u_j = c_j - A_j y$ nacheinander berechnen und abbrechen, sobald zum ersten Mal $u_j < 0$ ist (wobei für (b) die Menge N zuerst sortiert werden muss). Bei (c) muss dagegen stets der gesamte Vektor u_N berechnet werden, und (d) verlangt sogar jedesmal die Lösung eines Gleichungssystems. Dafür kann man erwarten, dass die aufwändigeren Regeln (die ja zu größerem Abstieg pro Schritt führen sollen) weniger Iterationen benötigen. Ob unter dem Strich weniger komplizierte Iterationen schneller sind als mehr einfache, hängt von der Problemgröße ab. (Erfahrungen favorisieren die einfachen Regeln für kleine bis mittlere Probleme (etwa 10^2 Variablen) und die komplizierteren Regeln für grössere Probleme).

Für die Wahl des Index s , der für r die Basis verlässt, existieren weniger Regeln. Zur Erinnerung: Hier suchen wir ein t so, dass $z_s = x_s - tw_s = 0$ ist. Verbreitet sind die

- (i) *Kleinster-Index-Regel*: wähle s als den *kleinsten* Index j , für den $\frac{x_j}{w_j} = \min_{i \in B, w_i > 0} \frac{x_i}{w_i}$ ist;
- (ii) *Lexikographische Regel*: wähle s als den Index $i \in B$ mit $w_i > 0$, für den $\frac{(A_B^{-1}A)_i}{w_i}$ lexikographisch minimal ist (ein Vektor x ist lexikographisch kleiner als y , falls $x \neq y$ und $x_i < y_i$ für den kleinsten Index i , für den $x_i \neq y_i$; z. B. ist $(1, 2)$ kleiner als $(2, 1)$).

Der Wert dieser Regeln ist, dass sie sicher stellen können, dass das Simplex-Verfahren auch bei Auftauchen von degenerierten Basisvektoren terminiert.

Dazu betrachten wir zuerst genauer, was einen degenerierten Basisvektor auszeichnet. Jeder Vektor $x \in \mathbb{R}^n$ wird durch n linear unabhängige Gleichungen eindeutig bestimmt; dies gilt insbesondere für jeden Basisvektor. Für einen nichtdegenerierten Basisvektor kann man diese Gleichungen eindeutig zerlegen in m Gleichungen $A_B x_B = b$ und $n - m$ Gleichungen

$x_N = 0$. (Geometrisch bedeutet dies, dass die Ecke $\{x\}$ Schnitt von genau n Facetten ist, von denen $n - m$ *Basisflächen* des positiven Orthanten $\{x \geq 0\}$ sind.) Ein degenerierter Basisvektor x erfüllt immer noch $A_B x_B = b$ und $x_N = 0$, jedoch ist zusätzlich $x_i = 0$ für (mindestens) ein $i \in B$. Damit erfüllt x die $n + 1$ Gleichungen $A_B x_B = b$, $x_B = 0$ und $x_i = 0$; x ist also überbestimmt. (Geometrisch: Die Ecke $\{x\}$ ist Schnitt von mehr als n Facetten.) Jede Wahl von n dieser Gleichungen ist also eine unterschiedliche Beschreibung desselben Basisvektors (wobei man den Fall, dass $x_i = 0$ als Gleichung in $A_B x_B = b$ auftaucht, vom Fall $x_i = 0$ wegen Gleichheit in $x \geq 0$ unterscheiden muss).

Was bedeutet das für das Simplex-Verfahren? In Schritt 7 bestimmen wir das *minimale* t , so dass eine der neuen Basisvariable $z_i = 0$ mit $i \in B$ erfüllt (nur so können wir garantieren, dass nicht mehrere Indizes die Basis verlassen). Da aber bereits eine alte Basisvariable $x_i = 0$ mit $i \in B$ erfüllt, passiert gar nichts – der Basisvektor ändert sich also nicht. Allerdings hat sich die Basis geändert (denn s wird aus der alten Basis gewählt, die r noch nicht enthält, und damit ist garantiert $r \neq s$). Insbesondere hat der störende Index i in jedem Fall die Basis verlassen. Die Karten sind also neu gemischt, und es ist möglich, dass in der nächsten Iteration die degenierte Ecke verlassen wird (wenn in der neuen Basis kein Index i mit $x_i = 0$ mehr enthalten ist). Allerdings ist es genauso möglich, dass der gerade entfernte Index als r wieder hinzugenommen wird – und das Spiel geht von vorne los. Da das Simplex-Verfahren deterministisch ist, wird es sich bei gleicher Ausgangssituation wieder gleich verhalten, und der Algorithmus gerät in einen *Zyklus*, in dem sich für den gleichen Basisvektor zwei oder mehr unterschiedliche Basen abwechseln. Wir müssen also verhindern, dass sich die Wahl von (r, s) wiederholen kann.

Eine Möglichkeit ist, r und s jeweils nach der Kleinster-Index-Regel auszuwählen; dies wird als *Regel von Bland* bezeichnet. Für diese Regel kann man beweisen, dass das Simplex-Verfahren in jedem Fall terminiert.

Satz 7.9. *Unter Berücksichtigung der Regel von Bland terminiert Algorithmus 7.1 nach endlich vielen Schritten.*

Beweis. Wir nehmen an, dass in Algorithmus 7.1 ein Zyklus auftritt; es gibt also eine Folge von Iterierten $x^k = x^{k+1} = \dots = x^{k+p}$ mit zugehörigen Basen B^k, \dots, B^{k+p} für die gilt $B^{k+p} = B^k$ und $B^{k+l} \neq B^k$ für $1 \leq l < p$. Sei r der *größte* Index, der während dieses Zyklus in die Basis aufgenommen wird. Um unnötige Notation zu vermeiden, bezeichnen wir die in diesem Schritt auftauchenden Größen einfach mit x, B, y, u, w , etc. Da r nach der Kleinste-Index-Regel gewählt wurde, gilt dann $u_r < 0$ und $u_j \geq 0$ für alle $j \in B$ mit $j < r$. Da es sich um einen Zyklus handelt, muss r die Basis irgendwann wieder zugunsten eines neuen Index verlassen. Die Größen in diesem Schritt bezeichnen wir mit B', y', u', w' , etc. Insbesondere ist $s' = r$ und

$$t' = \min_{i \in B', w'_i > 0} \frac{x_i}{w'_i} = \frac{x_{s'}}{w'_{s'}} = 0$$

(denn sonst wäre $t' > 0$ und wir würden den Zyklus verlassen); es gilt also $x_{s'} = 0$. Da auch s' nach der Kleinste-Index-Regel gewählt wurde, muss $w'_i \leq 0$ oder $x_i > 0$ für alle $i \in B'$ mit $i < s'$ gelten. Weiterhin gilt deshalb für

$$t'' := \begin{cases} 1 & \text{es gibt kein } i < s' \text{ mit } w'_i > 0, \\ \min_{i \in B', w'_i > 0, i < s'} \frac{x_i}{w'_i} & \text{sonst,} \end{cases}$$

dass $t'' > t' = 0$ (sonst wäre t' nicht der kleinste Index, für den das Minimum angenommen wird). Wir definieren nun $z \in \mathbb{R}^n$ durch

$$z_i := \begin{cases} x_i - t''w'_i & i \in B', \\ t'' & i = r', \\ 0 & \text{sonst.} \end{cases}$$

Dann ist $z_{r'} > 0$ und $z_i \geq 0$ für alle $i \in B'$ mit $i < s'$, da entweder $w'_i \leq 0$ oder t'' nach Konstruktion klein genug ist. Weiter gilt

$$\begin{aligned} Az &= \sum_{i \in B'} (x_i - t''w'_i)A^i + t''A^{r'} \\ &= \sum_{i \in B'} x_i A^i - t'' \left(\sum_{i \in B'} w'_i A^i - A^{r'} \right) \\ &= A_{B'} x_{B'} - t''(A_{B'} w'_{B'} - A^{r'}) \\ &= b, \end{aligned}$$

da B' Basis zu x ist und die Klammer nach Konstruktion von $w'_{B'}$ in Schritt 7 gleich Null ist.

Analog zur Herleitung von (7.4) (wofür die Bedingung $z \geq 0$ nicht verwendet wurde) erhalten wir daraus sowohl für das Paar (x, B) als auch für (x, B')

$$c^T z - c^T x = u_N^T z_N = (u'_{N'})^T z_{N'}.$$

Wir untersuchen nun die einzelnen Summanden in $u_N^T z_N = \sum_{j \in N} u_j z_j$ auf ihr Vorzeichen und machen dabei eine Fallunterscheidung nach $j \in N$.

1. Fall: $j = r$: Dann ist $u_r < 0$ nach Wahl von r und wegen $r = s' \in B'$ und $w'_{s'} > 0$ sowie $x_{s'} = 0$ ist

$$z_r = x_r - t''w'_r = 0 - t''w'_{s'} < 0,$$

und damit $u_r z_r > 0$.

2. Fall: $j < r$: Da r nach der Kleinste-Index-Regel gewählt wurde, ist $u_j \geq 0$. Ausserdem gilt nach Definition $z_j \geq 0$ (trivialerweise für $j \in N'$ und nach Bemerkung oben sonst) und damit $u_j z_j \geq 0$.

3. Fall: $j > r$: Da r nach Annahme der größte Index ist, der während des Zyklus aus der Nichtbasis in die Basis genommen wurde, gilt $j \in N'$. Insbesondere gilt $r' \leq r < j$ und damit $z_j = 0$.

Zusammen erhalten wir $u_N^\top z_N > 0$.

Für die zweite Summe erhalten wir aber sofort aus der Definition von z , dass

$$(u_{N'}')^\top z_{N'} = \sum_{j \in N'} u_j' z_j = u_{r'}' t'' < 0$$

gilt, da $u_{r'}' < 0$ nach Wahl von r' und $t'' > 0$ nach Konstruktion gilt. Damit haben wir einen Widerspruch, und es kann daher kein Zyklus auftreten. \square

Wählt man s nach der lexikographischen Regel, können unabhängig von der Wahlregel für r ebenfalls keine Zyklen auftreten. Diese Regel entspricht der Störung der rechten Seite b durch ein geeignetes $\eta \in \mathbb{R}^m$, $0 < \eta \ll 1$, so dass der gestörte Polyeder $P = (A, b + \eta)$ keine degenerierten Ecken enthält; siehe [Gritzmann 2014, Kapitel 5.4].⁴

⁴In der Praxis treten Zyklen selten auf; man kann das Rechnen mit endlicher Genauigkeit als solch eine kleine Störung auffassen – da man Zahlen nur bis auf Maschinengenauigkeit vergleichen kann, wird in der Regel nie ein x_j exakt gleich Null sein. Trotzdem muss man in Programmen diese Möglichkeit natürlich berücksichtigen.

DAS DUALE SIMPLEX-VERFAHREN

8

Das Simplex-Verfahren liefert gleichzeitig eine primale Lösung \bar{x} und eine duale Lösung \bar{y} , indem so lange eine Folge zulässiger Vektoren x und dazugehöriger komplementärer Vektoren $y := (A_B^T)^{-1} c_B$ erzeugt wird, bis y auch dual zulässig (d. h. $u_N := c_N - A_N^T y \geq 0$) ist. Im *dualen* Simplex-Verfahren geht man umgekehrt vor: man hält y dual zulässig und komplementär zu x , und versucht Zulässigkeit von x zu erreichen. Eine Möglichkeit dafür ist die Anwendung von Algorithmus 7.1 auf das duale Problem (D), das wir dafür in Normalform transformieren müssen. Durch Einführen der Schlupfvariablen $z \geq 0$ erhalten wir

$$(8.1) \quad \begin{cases} \max_{y \in \mathbb{R}^m, z \in \mathbb{R}^n} & b^T y \\ \text{mit } & A^T y + z = c, \\ & z \geq 0. \end{cases}$$

Wir nehmen wieder an, dass A vollen Zeilenrang hat, d. h. $\text{rang}(A) = \text{rang}(A^T) = m \leq n$. Um das Simplex-Verfahren auf (8.1) anzuwenden, führen wir $D = (A^T, I) \in \mathbb{R}^{n \times (m+n)}$ ein. Ein Basisvektor (y, z) hat als Basis also eine Teilmenge $H \subset \{1, \dots, m+n\}$ mit $|H| = n$; die entsprechende Nichtbasis bezeichnen wir mit M . Da die Variablen $y \in \mathbb{R}^m$ keine Vorzeichenbedingungen erfüllen müssen, können sie nur durch Gleichheitsnebenbedingungen festgelegt werden. Existiert also ein optimaler Basisvektor (\bar{y}, \bar{z}) , so muss in der dazugehörigen Basis \bar{H} die Menge $\{1, \dots, m\}$ enthalten sei; ansonsten wäre zumindest ein y_i frei wählbar und damit $b^T y$ unbeschränkt. Wir schränken daher unsere Suche nach einer Lösung auf die entsprechenden Ecken ein. Es kommen also noch $n - m \geq 0$ Variablen aus $z \in \mathbb{R}^n$ für die Basis in Frage. Wir bezeichnen die entsprechenden Indizes aus $\{1, \dots, n\}$ mit N und setzen $B := \{1, \dots, n\} \setminus N$, und erhalten damit die Basismatrix $D_H = (A^T, I_N)$ und die Nichtbasismatrix $D_M = I_B$. Die Nichtbasisvariablen sind also genau die z_B . (Der Grund für diese vom primalen Simplex-Verfahren abweichende Notation wird später klar.)

Ein Basisvektor (y, z) mit zugehöriger Basis H ist genau dann zulässig, wenn $A^T y + z = c$ und $z \geq 0$ gilt. Die Gleichung können wir nun zerlegen in Zeilen $i \in B$ und Zeilen $i \in N$:

$$A_B^T y + z_B = c_B, \quad A_N^T y + z_N = c_N.$$

Für die Nichtbasisvariablen gilt nun nach Definition $z_B = 0$. Weiter gilt, dass die Spalten von D_H linear unabhängig sind; damit sind es insbesondere die Spalten von A^T und auch von A_B^T . Also sind die Zeilen von $A_B \in \mathbb{R}^{m \times m}$ linear unabhängig und damit A_B regulär. Der Vektor y ist daher eindeutig als Lösung von $A_B^T y = c_B$ bestimmt. Einsetzen in die zweite Gleichung liefert dann $z_N = c_N - A_N^T y$. Der Basisvektor (y, z) ist also genau dann zulässig für (8.1), wenn $z_N \geq 0$ gilt.

Haben wir einen zulässigen Basisvektor (y, z) gegeben, können wir Algorithmus 7.1 auf Problem (8.1) anwenden. Dabei wollen wir wegen $n \gg m$ vermeiden, mit der künstlich erweiterten Matrix D und der zugehörigen Basis H zu operieren. Wir betrachten also der Reihe nach die Schritte in Algorithmus 7.1:

SCHRITT 3: Löse $D_H^T x = (b, 0)^T$ für $x \in \mathbb{R}^n$. Nach Einsetzen von $D_H = (A^T, I_N)$ ist dies äquivalent zu $Ax = b$ und $x_N = 0$, womit sich die erste Gleichung vereinfacht zu $A_B x_B = b$.

SCHRITT 4: Setze $u_M = (b, 0)_M^T - D_M^T x$. Da y stets zu den Basisvariablen gehört, ist $(b, 0)_M = 0_B = 0$, und damit muss nur $u_B = -I_B x = -x_B$ betrachtet werden.

SCHRITT 5: Gilt $u_B \leq 0$ (im dualen Problem wird ja maximiert), so ist $x_B \geq 0$ und mit der Wahl $x_N = 0$ daher x zulässig. Wegen $A_B^T y = c_B$ sind auch die Komplementaritätsbedingungen erfüllt, und damit ist x Lösung von (P) und y Lösung von (D).

SCHRITT 6: Andernfalls wähle ein $r \in B$ mit $x_B < 0$. Der Index r tritt also in die duale Basis N ein (und verlässt damit die primale Basis B). Beachte, dass der entsprechende Index in M dann $m + r$ ist.

SCHRITT 7: Löse $D_H w_H = D^{m+r}$. Mit $w = (p, q)^T$ ist $w_H = (p, q_N)$. Weiter gilt $D_H = (A^T, I_N)$, und wegen $r \notin N$ kann man die Gleichung zerlegen in

$$\begin{pmatrix} A_B^T & 0 \\ A_N^T & I \end{pmatrix} \begin{pmatrix} p \\ q_N \end{pmatrix} = \begin{pmatrix} D_B^{m+r} \\ D_N^{m+r} \end{pmatrix} \cdot \begin{pmatrix} (e_r)_B \\ 0 \end{pmatrix}$$

mit $e_r \in \mathbb{R}^m$. Die erste Gleichung ergibt $A_B^T p = (e_r)_B$, und Einsetzen in die zweite $q_N = -A_N^T p$.

SCHRITT 8: Überprüfe, ob $w_H \leq 0$. Da wir stets alle Variablen in y in der Basis halten wollen, kommt für den Basistausch nur ein Index aus N in Frage. Wir müssen also lediglich $q_N \leq 0$ überprüfen. Ist dies der Fall, ist das duale Problem unbeschränkt (und damit das primale Problem nach Folgerung 2.5 (ii) nicht zulässig).

SCHRITT 9: Wähle $t = \min_{i \in H, w_i > 0} \frac{(y, z)_i}{w_i}$. Wegen der Beschränkung auf die Variable z reduziert sich die Wahl auf

$$t = \min_{i \in N, q_i > 0} \frac{z_i}{q_i} = \frac{z_s}{q_s}.$$

Der Index s verlässt also die duale Basis N und tritt damit in die primale Basis B ein; der entsprechende Index in H ist $m + s$.

SCHRITT 10: Setze $(y, z)_H = (y, z)_H - tw_H$, $(y, z)_{m+r} = t$, und $H = (H \cup \{m+r\}) \setminus \{m+s\}$.
 Nach Festlegung von H bedeutet dies

$$y = y - tp, \quad z_N = z_N - tq_N, \quad z_r = t, \quad N = (N \cup \{r\}) \setminus \{s\}.$$

Damit erhalten wir das duale Simplex-Verfahren in Algorithmus 8.1. Verwendet man zur Wahl von r und s die Regel von Bland, folgt aus Satz 7.5 und den obigen Überlegungen zu den Abbruchbedingungen die Korrektheit des Verfahrens.

Algorithmus 8.1 : Duales Simplex-Verfahren

Input : $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$,

Basisvektor $x \in \mathbb{R}^n$ (nicht unbedingt zulässig) mit Basis B, Nichtbasis N,
 dualen Basisvektor $(y, z) \in \mathbb{R}^m \times \mathbb{R}^n$ mit $A_B^T y = c_B$ und $z_N := c_N - A_N^T y \geq 0$,
 $z_B = 0$

Output : Primal-duales Paar (\bar{x}, \bar{y}) oder Information „primales Problem unzulässig“

```

1 while nicht fertig do
2   Bestimme Nichtbasis B, Matrizen  $A_B, A_N$ 
3   Löse  $A_B x_B = b$ 
4   if  $x_B \geq 0$  then return  $(x, y)$  // x, y optimal
5   Wähle  $r \in B$  mit  $x_r < 0$ 
6   Löse  $A_B^T p = (e_r)_B$  und setze  $q = -A_N^T p$ 
7   if  $q_N \leq 0$  then return nicht lösbar // Problem unzulässig
8   Wähle  $t = \min_{i \in N, q_i > 0} \frac{z_i}{q_i} = \frac{z_s}{q_s}$ 
9   Setze  $z_N = z_N - tq_N$ 
        $z_r = t$ 
        $y = y - tp$ 
        $N = (N \cup \{r\}) \setminus \{s\}$ 
    
```

Der Vorteil des dualen Simplex-Verfahren gegenüber der primalen Variante ist, dass es einfacher sein kann, einen dual zulässigen Basisvektor zu finden als einen primal zulässigen. Dies ist insbesondere dann der Fall, wenn zu einem bereits gelösten Problem neue Restriktionen hinzugefügt werden: Angenommen, wir haben für das Problem (P) mit Hilfe des (primalen oder dualen) Simplex-Verfahrens einen optimalen Basisvektor $\bar{x} \in \mathbb{R}^n$ mit Basis B und zugehöriger dualer Lösung $\bar{y} \in \mathbb{R}^m$ gefunden, und werden mit dem modifizierten Problem

$$(8.2) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \\ a^T x \leq \beta, \end{cases}$$

für $a \in \mathbb{R}^n$ und $\beta \in \mathbb{R}$ konfrontiert. Einführen einer Schlupfvariablen $\xi \in \mathbb{R}$ transformiert dieses Problem in die Normalform

$$(8.3) \quad \begin{cases} \min_{x' \in \mathbb{R}^{n+1}} (c')^T x' \\ \text{mit } A' x' = b', \\ x' \geq 0 \end{cases}$$

mit

$$x' = \begin{pmatrix} x \\ \xi \end{pmatrix}, \quad c' = \begin{pmatrix} c \\ 0 \end{pmatrix}, \quad A' = \begin{pmatrix} A & 0 \\ a^T & 1 \end{pmatrix}, \quad b' = \begin{pmatrix} b \\ \beta \end{pmatrix}.$$

Dann ist

$$x' := \begin{pmatrix} \bar{x} \\ \beta - a^T \bar{x} \end{pmatrix}$$

ein (möglicherweise degenerierter) Basisvektor zu (8.3) mit Indexmenge $B' = B \cup \{n+1\}$, denn für A_B invertierbar ist auch

$$A'_{B'} = \begin{pmatrix} A_B & 0 \\ a_B^T & 1 \end{pmatrix}$$

invertierbar. Ein zulässiger dualer Basisvektor ist dann

$$y' := \begin{pmatrix} \bar{y} \\ 0 \end{pmatrix}, \quad z' := \begin{pmatrix} \bar{z} \\ 0 \end{pmatrix}$$

mit $\bar{z}_N = c_N - A_N^T \bar{y}$ und $\bar{z}_B = 0$, denn wegen $(n+1) \in B'$ und der Zulässigkeit von (\bar{y}, \bar{z}) gilt $z'_{N'} = \bar{z}_N \geq 0$, $z'_{B'} = (\bar{z}_B, 0)^T = 0$ sowie

$$(A'_{B'})^T y' = \begin{pmatrix} A_B^T & a_B \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{y} \\ 0 \end{pmatrix} = \begin{pmatrix} A_B^T \bar{y} \\ 0 \end{pmatrix} = \begin{pmatrix} c_B \\ 0 \end{pmatrix} = c'_{B'}.$$

Ist nun $a^T \bar{x} \leq \beta$, so ist \bar{x} zulässig für (8.2) und damit optimal (denn durch die zusätzliche Restriktion kann der minimale Funktionswert nur grösser werden). Ansonsten ist x' ein zulässiger und optimaler Basisvektor für (8.3), und man stellt mit einem oder mehreren Schritten des dualen Simplexverfahrens, ausgehend von N' und (y', z') , die primale Zulässigkeit und damit die Optimalität von x' her.

Ähnlich geht man bei Änderungen der rechten Seite vor: Angenommen, \bar{x} ist ein optimaler Basisvektor für (P) mit Basis B , und in (P) wird b durch ein b' ersetzt. Gilt $x'_B := A_B^{-1} b' \geq 0$, so ist x' mit $x'_N = 0$ zulässig. Da b nicht in die Charakterisierung der dualen Zulässigkeit eingeht und sich die Basis nicht geändert hat, bleibt \bar{y} zulässig und komplementär zu x' ; damit ist x' eine Lösung des modifizierten Problems. Ansonsten hat man einen zulässigen dualen Basisvektor \bar{y} und kann mit dem dualen Simplex-Verfahren die primale Zulässigkeit von x' erreichen.

PRIMAL-DUALE VERFAHREN

Die Grundidee des (primalen oder dualen) Simplex-Verfahrens ist, solange von Ecke zu Ecke des (primalen oder dualen) Zulässigkeitsbereiches zu springen, bis die Komplementaritätsbedingungen erfüllt sind und daher ein primal-duales Paar optimaler Lösungen gefunden ist. Nun kann man für jede Index-Wahl-Regel ein lineares Optimierungsproblem konstruieren – z. B. den sogenannten „Klee–Minty-Würfel“ – für den das entsprechende Simplex-Verfahren jede Ecke des zulässigen Polyeders absuchen muss, bevor das Optimum erreicht wird. Primal-duale Verfahren suchen dagegen für einen gegebenen (üblicherweise dual) zulässigen Vektor (der kein Basisvektor sein muss!) direkt einen zulässigen komplementären Vektor, der dann natürlich mit dem gegebenen Vektor ein primal-duales Lösungspaar ergibt. Existiert kein solcher Vektor, so wird der ursprüngliche Vektor geeignet modifiziert und das Verfahren wiederholt.

9

9.1 DAS PRIMAL-DUALE SIMPLEX-VERFAHREN

Wir betrachten wieder das primale Problem in Normalform

$$(9.1) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b \\ x \geq 0 \end{cases}$$

mit zugehörigem dualen Problem

$$(9.2) \quad \begin{cases} \max_{y \in \mathbb{R}^n} b^T y \\ \text{mit } A^T y \leq c \end{cases}$$

und Komplementaritätsbedingungen

$$(9.3) \quad x_j (c_j - [A^T y]_j) = 0 \quad \text{für alle } j = 1, \dots, n$$

Wir nehmen wieder an, dass $A \in \mathbb{R}^{m \times n}$ vollen Rang hat und – ohne Einschränkung – dass $b \geq 0$ gilt.

Sei nun $y \in \mathbb{R}^m$ ein dual zulässiger Vektor und bezeichne mit

$$J := \{j \in \{1, \dots, n\} : [A^T y]_j = c_j\}$$

die in y *aktive Menge* der Ungleichungen. (Da J keine Basis sein muss, ist weder $|J| = m$ noch Regularität von A_J gefordert!) Existiert nun ein $x \in P^=(A, b)$ mit $x_j = 0$ für alle $j \notin J$, so sind die Komplementaritätsbedingungen (9.3) erfüllt und damit x Lösung von (9.1) und y Lösung von (9.2). Da die aktive Menge bereits feststeht, kann man $x_j = 0$ für $j \notin J$ natürlich festsetzen; wir müssen also lediglich $x_J \in \mathbb{R}^{|J|}$ finden mit

$$A_J x_J = b, \quad x_J \geq 0.$$

Wir gehen dafür analog zur Phase I im Simplex-Verfahren vor, indem wir das *reduzierte primale Problem*

$$(9.4) \quad \left\{ \begin{array}{l} \min_{x_J, z} \mathbb{1}^T z \\ \text{mit } A_J x_J + z = b \\ \quad \quad \quad x_J \geq 0 \\ \quad \quad \quad z \geq 0 \end{array} \right.$$

betrachten. Dieses Problem ist zulässig, da wegen $b \geq 0$ zum Beispiel $(x_J, z) = (0, b)$ ein zulässiger Vektor ist. Die Zielfunktion ist wegen $z \geq 0$ ausserdem nach unten durch 0 beschränkt, so dass (9.4) eine Lösung (\bar{x}_J, \bar{z}) mit Optimalwert $\bar{\mu} := \mathbb{1}^T \bar{z} = |\bar{z}|_1$ hat. Ist $\bar{\mu} = 0$, so gilt daher $\bar{z} = 0$ und damit $A_J \bar{x}_J = b$, d. h. $\bar{x}_J \geq 0$ ist unser gesuchter Vektor und (\bar{x}, y) mit $\bar{x}_j = 0$ für $j \notin J$ ist ein primal-duales Lösungspaar.

Ist dagegen $\bar{\mu} > 0$, so existiert kein x_J mit den gesuchten Eigenschaften, und y kann nicht zu einem primal-dualen Paar ergänzt werden. Dies ist auch nur für Lösungen von (9.2) möglich. Wir suchen daher einen neuen zulässigen Vektor y' mit $b^T y' > b^T y$, wofür wir wie im Simplexverfahren den Ansatz $y' = y + tu$ mit $t > 0$ und $u \in \mathbb{R}^m$ geeignet machen. Da aber y' nun kein Basisvektor sein muss, können wir dabei aggressiver vorgehen: Wir suchen t und u so, dass $b^T y' = b^T y + tb^T u$ maximal ist unter der Nebenbedingung

$$A^T y' = A^T y + tA^T u \leq c.$$

Die Richtung u berechnen wir dafür als Lösung des *reduzierten dualen Problems*

$$(9.5) \quad \left\{ \begin{array}{l} \max_{u \in \mathbb{R}^n} b^T u \\ \text{mit } A_J^T u \leq 0 \\ \quad \quad \quad u \leq 1. \end{array} \right.$$

Dies ist gerade das duale Problem zu (9.4). Da letzteres lösbar ist, hat (9.5) nach Satz 2.4 ebenfalls eine Lösung \bar{u} mit $b^T \bar{u} = \bar{\mu} > 0$. Daher gilt für alle $t > 0$

$$b^T y' = b^T y + t b^T \bar{u} = b^T y + \bar{\mu} > b^T y.$$

Nun müssen wir durch geeignete Wahl der Schrittlänge t noch erreichen, dass $y + t\bar{u}$ zulässig ist. Ist $[A^T \bar{u}]_j \leq 0$ für alle $j \notin J$, so gilt dies für jede Wahl von $t > 0$ (für aktive Indizes $j \in J$ folgt dies bereits aus der Zulässigkeit von \bar{u}); wegen $b^T y' = b^T y + t\bar{\mu} \rightarrow \infty$ für $t \rightarrow \infty$ muss dann aber (9.2) unbeschränkt und daher (9.1) unzulässig sein. Ansonsten existiert

$$t = \min_{j \notin J, [A^T \bar{u}]_j > 0} \frac{c_j - [A^T y]_j}{[A^T \bar{u}]_j},$$

für das nach Konstruktion $y' = y + t\bar{u}$ zulässig ist.

Damit ist unser primal-duales Verfahren im Prinzip schon vollständig spezifiziert; für die gleichzeitige Lösung der beiden reduzierten Hilfsprobleme bietet sich natürlich das (primale oder duale) Simplex-Verfahren an.

Algorithmus 9.1 : Primal-Duales-Simplex-Verfahren

Input : $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, zulässiger Vektor $y \in P(A^T, c)$

Output : Primal-duales Paar (\bar{x}, \bar{y}) oder Information „primales Problem unzulässig“

```

1 while nicht fertig do
2   Bestimme aktive Menge  $J := \{j \in \{1, \dots, n\} : [A^T y]_j = c_j\}$ 
3   Bestimme Lösung  $(x_J, z)$  von (9.4) und Lösung  $u$  von (9.5) mit Simplex-Verfahren
4   if  $b^T u = 0$  then return  $(x, y)$  (mit  $x_j = 0$  für  $j \notin J$ )
5   else if  $[A^T u]_j \leq 0$  für alle  $j \notin J$  then return Problem unbeschränkt
6   Wähle  $t = \min_{j \notin J, [A^T u]_j > 0} \frac{c_j - [A^T y]_j}{[A^T u]_j}$ 
7   Setze  $y = y + tu$ 

```

Für den dual zulässigen Startvektor kann man im Fall $c \geq 0$ den Vektor $y = 0$ wählen; ansonsten löst man wie in Kapitel 7.2 ein geeignet konstruiertes Hilfsproblem. Für die praktische Anwendung des (primalen) Simplex-Verfahrens in Schritt (3) kann man verwenden, dass eine Basislösung $(x_{J^k}^k, z^k)$ in Iteration k bereits einen zulässigen Basisvektor für Iteration $k + 1$ liefert: Ist nämlich $j \in J^k$ mit $x_j^k > 0$ (sonst ist x^k eine degenerierte Ecke, und Basis-komponenten mit $x_j^k = 0$ müssen ausgetauscht werden), so folgt aus der Komplementarität der reduzierten Probleme $[A^T u]_j = 0$ und damit

$$[A^T y^{k+1}]_j = [A^T y^k]_j + t[A^T u]_j = A^T y^k = c_j,$$

d. h. $j \in J^{k+1}$. Setzen wir also $x_j^{k+1} = x_j^k$ für $j \in J^k$ und $x_j^{k+1} = 0$ für $j \in J^{k+1} \setminus J^k$ (falls nichtleer), so ist (x^{k+1}, z^k) zulässig für das reduzierte Problem in Iteration $k + 1$ und $B^{k+1} = B^k$ weiterhin eine Basis. (Die Anzahl der Gleichungs-Nebenbedingungen ändert sich ja nicht.) In der ersten Iteration kann man für $b \geq 0$ mit dem Basisvektor $(x_J, z) = (0, b)$ starten.

9.2 KOMBINATORISCHE PRIMAL-DUALE ALGORITHMEN

Das primal-duale Simplex-Verfahren ist natürlich nur dann von Vorteil, wenn sich die reduzierten Probleme (9.4) und (9.5) deutlich schneller lösen lassen als die vollen Probleme (9.1) und (9.2). Dies ist insbesondere dann der Fall, wenn für die reduzierten Probleme eine Lösung direkt konstruiert werden kann, was für kombinatorische Probleme oft möglich ist. So entspricht zum Beispiel das Finden eines erhöhenden Weges im Ford–Fulkerson-Verfahren der Lösung eines reduzierten dualen Problems (und der Betrag ε , um den der Fluss auf diesem Weg erhöht wird, dem Wert t in Schritt 6); mit Algorithmus 6.1 haben wir also bereits ein primal-duales Verfahren kennengelernt.¹

Anstatt das Ford–Fulkerson-Verfahren noch einmal unter diesem Gesichtspunkt herzuleiten, betrachten wir ein verwandtes (einfacheres) Problem: Das Finden kürzester Wege in gerichteten Graphen. Sei wieder $G = (V, K)$ ein gerichteter Graph mit Knotenmenge $V = \{v_1, \dots, v_m\}$, Kantenmenge $K = \{k_1, \dots, k_n\}$ und Kantengewichten $c_j \geq 0$ für alle $k_j \in K$. Zu gegebenen $s, t \in V$ suchen wir nun den kürzesten (s, t) -Weg, d. h. denjenigen Weg von s nach t , für den die Summe der Gewichte entlang der durchlaufenen Kanten minimal ist. Um dies als lineares Optimierungsproblem schreiben zu können, repräsentieren wir einen Weg w durch einen Vektor $x \in \mathbb{R}^n$ mittels der Definition

$$x_j = \begin{cases} 1 & k_j \in w, \\ 0 & k_j \notin w. \end{cases}$$

Die Länge eines Weges w ist dann gegeben durch

$$\sum_{k_j \in w} c_j = c^T x.$$

Die Bedingung, dass w ein (s, t) -Weg ist, lässt sich wie folgt ausdrücken: Für jeden Knoten $v_i \in V$ gilt

$$\sum_{k_j=(p,v_i) \in K} x_j - \sum_{k_j=(v_i,q) \in K} x_j = b_i := \begin{cases} -1 & \text{falls } v_i = s, \\ 1 & \text{falls } v_i = t, \\ 0 & \text{sonst.} \end{cases}$$

In Worten: Die einzigen Knoten, die in w genau einmal auftauchen, sind s und t – alle anderen Knoten kommen entweder gar nicht vor oder zweimal (einmal als Start- und einmal als Endknoten einer inneren Kante). Mit Hilfe der Inzidenzmatrix $A \in \mathbb{R}^{n \times m}$,

$$A_{ij} = \begin{cases} -1 & k_j = (v_i, q) \text{ für ein } q \in V, \\ 1 & k_j = (p, v_i) \text{ für ein } p \in V, \\ 0 & \text{sonst,} \end{cases}$$

¹Tatsächlich wurde Algorithmus 9.1 von Dantzig, Ford, und Fulkerson 1956 als Verallgemeinerung des Ford–Fulkerson-Verfahrens vorgestellt.

können wir diese Bedingungen kurz schreiben als $Ax = b$. Ersetzen wir die Bedingung $x \in \{0, 1\}^n$ durch $x \geq 0$, liegt also ein lineares Optimierungsproblem in Normalform (9.1) vor, auf das wir das primal-duale Simplex-Verfahren anwenden können. Allerdings hat die Matrix A keinen vollen Rang (da sich alle Zeilen zu Null addieren); wir streichen daher die Zeile, die zum Knoten s gehört; mit der Konvention $v_1 = s$ und $v_2 = t$ ist dies also die erste Zeile. Der Übersichtlichkeit halber passen wir die Nummerierung *nicht* an. Damit erhalten wir das duale Problem

$$(9.6) \quad \begin{cases} \max_{y \in \mathbb{R}^m} y_2 \\ \text{mit } y_r - y_i \leq c_j, & k_j = (v_i, v_r) \in K, \\ y_1 = 0, \end{cases}$$

wobei wir $y_1 = 0$ gesetzt haben, da die zugehörige Ungleichung im primalen Problem eliminiert wurde.

Da $c \geq 0$ vorausgesetzt wurde, haben wir mit $y = 0$ einen dual zulässigen Startvektor. Die zu y gehörende aktive Menge J ist dann gegeben durch alle Indizes j mit

$$c_j = [A^T y]_j = y_r - y_i \quad \text{für } k_j = (v_i, v_r).$$

Das reduzierte duale Problem ist dann

$$(9.7) \quad \begin{cases} \max_{u \in \mathbb{R}^m} u_2 \\ \text{mit } u_r - u_i \leq 0, & k_j = (v_i, v_r) \in K, j \in J, \\ u \leq 1, \\ u_1 = 0. \end{cases}$$

Der wesentliche Punkt ist nun, dass für dieses Problem eine Lösung direkt angegeben werden kann. Einen Knoten $v \in V$ nennen wir dafür *J-erreichbar*, wenn es einen (s, v) -Weg gibt, der nur Kanten k_j mit $j \in J$ enthält.

Lemma 9.1. Sei $u \in \mathbb{R}^m$ mit

$$u_i = \begin{cases} 0 & v_i \text{ ist } J\text{-erreichbar,} \\ 1 & \text{sonst,} \end{cases}$$

und bezeichne $\bar{\mu}$ den optimalen Wert in (9.7). Dann gilt:

- (i) u ist zulässig für (9.7);
- (ii) $\bar{\mu} = 0$ genau dann, wenn t ein J -erreichbarer Knoten ist;
- (iii) ist $u_2 \neq 0$, so ist $u_2 = 1$ und u Lösung von (9.7).

Beweis. Zu (i): Angenommen, u wäre unzulässig. Dann gibt es eine Kante $k_j = (v_i, v_r)$ für $j \in J$ mit $u_i = 0$ und $u_r = 1$ (die mit dieser Wahl von u einzige Möglichkeit für $u_r - u_i > 0$). Also ist v_i , aber nicht v_r , J -erreichbar von s . Dies ist aber ein Widerspruch zu $j \in J$, denn jeden (s, v_i) -Weg in J kann man mit k_j zu einem (s, v_r) -Weg in J verlängern.

Zu (ii): Ist t nicht J -erreichbar, so ist u ein zulässiger Vektor mit $u_2 = 1 > 0$, d. h. $\bar{\mu} > 0$. Ist t aber J -erreichbar, so ergibt Addition der zu den entsprechenden Kanten gehörenden Ungleichungen in (9.7) für jeden zulässigen Vektor u die Bedingung

$$0 \geq u_2 - u_1 = u_2,$$

und damit ist $\bar{u} = 0$ eine Lösung mit $\bar{\mu} = \bar{u}_2 = 0$.

Zu (iii): Ist $\bar{\mu} > 0$, kann t nach (ii) nicht J -erreichbar sein. Also ist $u_2 = 1$, und wegen $0 < \bar{\mu} = \bar{u}_2 \leq 1$ wird in u das Maximum angenommen. \square

Ist $\bar{\mu} = 0$, so ist y Lösung von (9.6); ausserdem haben wir festgestellt, dass t ein J -erreichbarer Knoten ist. Es gibt also einen (s, t) -Weg w , der nur Kanten k_j mit $j \in J$ enthält. Definieren wir $x \in \mathbb{R}^n$ mit

$$x_j = \begin{cases} 1 & k_j \in w, \\ 0 & \text{sonst,} \end{cases}$$

so ist dieser Vektor zulässig und erfüllt wegen $x_j = 0$ für alle $j \notin J$ die Komplementaritätsbedingungen (9.3). Ansonsten wählen wir (falls das primale Problem zulässig ist, d. h. ein (s, t) -Weg existiert)

$$t = \min_{\substack{j \in J, k_j = (v_i, v_r) \\ u_i = 0, u_r = 1}} c_j - (y_r - y_i),$$

setzen $y = y + tu$, und wiederholen die Prozedur. Wie im Ford–Fulkerson-Verfahren argumentiert man, dass dieses Verfahren für ganzzahlige Kantengewichte nach endlich vielen Schritten terminiert.

Die zur Konstruktion von u notwendige Erreichbarkeitsbestimmung kann wieder über eine Breitensuche durchgeführt werden.

Algorithmus 9.2 : Primal-duales Verfahren für kürzeste Wege**Input** : $(V, K), c \in \mathbb{R}^n, y = 0$ **Output** : Kürzester Weg x oder Information „kein (s, t) -Weg“

```

1 while nicht fertig do
2   Bestimme aktive Menge  $J := \{j \in \{1, \dots, n\} : y_r - y_i = c_j, k_j = (v_i, v_r)\}$ 
3   Setze  $U = \{v_1\}, u = \mathbb{1}$ 
4   repeat // Markiere J-erreichbare Knoten
5     for  $v_i \in U$  do // Prüfe Knoten
6       for  $k_j = (v_i, v_r) \in K$  mit  $j \in J$  do // Prüfe alle J-erreichbaren Knoten
7         if  $r = 2$  then // t ist J-erreichbar
8           Setze  $x = 0$ 
9           repeat // Gehe (s,t)-Weg zurück
10            Setze  $x_j = 1, i = |\sigma_r|$ 
11            Finde  $j$  mit  $k_j = (v_i, v_r)$ 
12            until  $i = 1$ 
13            return  $x$  // y ist optimal, x kürzester Weg
14          else if  $u_r = 1$  then //  $v_r$  noch nicht besucht
15            Setze  $u_r = 0$  //  $v_r$  ist J-erreichbar
16            Setze  $\sigma_r = i$  // Vorgänger
17            Setze  $U := U \cup \{v_r\}$  //  $v_r$  in Warteschlange
18          Setze  $U := U \setminus \{v_i\}$  //  $v_i$  besucht
19 until  $U = \emptyset$  // alle J-erreichbaren Knoten markiert
20 if  $[A^T u]_j \leq 0$  für alle  $j \notin J$  then
21   | return kein  $(s, t)$ -Weg
22 else
23   | Wähle  $t = \min_{\substack{j \notin J, k_j = (v_i, v_r) \\ u_i = 0, u_r = 1}} c_j - (y_r - y_i)$ 
24   | Setze  $y = y + tu$ 

```

GANZZAHLIGE LINEARE OPTIMIERUNG

Zum Abschluss geben wir noch einen Ausblick auf Verfahren der ganzzahligen linearen Optimierung. Wir betrachten dafür das *ganzzahlige Optimierungsproblem* (englisch „integer linear program“)

10

$$(IP) \quad \begin{cases} \min_{x \in \mathbb{Z}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \end{cases}$$

mit $c \in \mathbb{Z}^n$, $A \in \mathbb{Z}^{m \times n}$ mit $\text{rang}(A) = m < n$, und $b \in \mathbb{Z}^m$. (Rationale Probleme können mit dem kleinsten gemeinsamen Vielfachen aller Nenner durchmultipliziert werden.) Für die bisher betrachteten Beispiele aus der Graphentheorie hatte die *LP-Relaxierung* (englisch: „linear programming relaxation“)

$$(LP) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \end{cases}$$

stets eine Lösung $\bar{x} \in \mathbb{Z}^n$; dies ist aber im Allgemeinen nicht der Fall. Ein Beispiel dafür ist das *Rucksackproblem*, bei dem möglichst viele von n Gegenständen mit Gewichten a_1, \dots, a_n in einen Behälter mit Kapazität $\beta \ll \sum_i a_i$ gepackt werden sollen. Führen wir die *Entscheidungsvariable* $x \in \{0, 1\}^n$ mit $x_i = 1$ genau dann, wenn Gegenstand i im Behälter ist, sowie den Gewichtsvektor $a = (a_1, \dots, a_n)^T$ ein, so führt dies auf das ganzzahlige Optimierungsproblem

$$\begin{cases} \max_{x \in \mathbb{Z}^n} \mathbb{1}^T x \\ \text{mit } a^T x \leq \beta, \\ 0 \leq x \leq 1. \end{cases}$$

Existiert hierfür keine Lösung mit $a^T x = \beta$ (z. B. für $a = (3, 4, 5)$ und $\beta = 10$), so wird die LP-Relaxierung nur nicht-ganzzahlige Lösungen haben, etwa $\bar{x} = (1, 1, 3/5)$; durch Runden

erhält man zwar einen zulässigen Vektor $(1, 1, 0)$, der aber nicht optimal ist (Wert 7 statt 9 für $(0, 1, 1)$). Gerundete Lösungen müssen nicht mal zulässig sein: Zum Beispiel hat das Problem

$$\begin{cases} \min_{x \in \mathbb{Z}^2} -x_1 - x_2 \\ \text{mit } 7x_1 - 5x_2 \leq 7, \\ -12x_1 + 15x_2 \leq 7, \\ x \geq 0, \end{cases}$$

die Lösung $\bar{x} = (2, 2)$ mit optimalem Funktionswert -4 . Die zugehörige LP-Relaxierung hat dagegen die Lösung $\hat{x} = (28/9, 133/45) \approx (3.11, 2.95)$ (mit optimalem Funktionswert $-273/45 \approx -6.6$) – und keiner der vier durch Runden von \hat{x} erzeugten ganzzahligen Punkte ist zulässig.

Das Problem ist also, dass der zulässige Bereich störende nicht-ganzzahlige Lösungen enthält. Wir müssen den zulässigen Bereich also so anpassen, dass die störenden Lösungen nicht mehr zulässig sind, alle ganzzahligen Vektoren aber zulässig bleiben. Idealerweise möchten wir dabei den zulässigen Bereich für (LP) auf ein Polyeder reduzieren, das nur ganzzahlige Ecken hat. Das entsprechende lineare Optimierungsproblem hat dann stets eine Lösung, die auf einer Ecke liegt und damit ganzzahlig ist.

10.1 SCHNITTEBENEN-VERFAHREN

Ein Ansatz, dies zu gewährleisten, ist die folgende Konstruktion. Wir bezeichnen den zulässigen Bereich für (IP) mit

$$IP^=(A, b) := P^=(A, b) \cap \mathbb{Z}^n = \{x \in \mathbb{Z}^n : Ax = b, x \geq 0\}.$$

Weiter sei

$$CIP^=(A, b) := \left\{ \sum_{i=1}^r \lambda_i x^i : r \in \mathbb{N}, x^1, \dots, x^r \in IP^=(A, b), \lambda_1, \dots, \lambda_r \geq 0 \text{ mit } \sum_{i=1}^r \lambda_i = 1 \right\}$$

die *konvexe Hülle* von $IP^=(A, b)$. Aus der konvexen Geometrie verwenden wir ohne Beweis folgende Eigenschaften.

Lemma 10.1. $CIP^=(A, b)$ ist ein Polyeder.¹

Lemma 10.2. $IP^=(A, b) \subset CIP^=(A, b) \subset P^=(A, b)$.

Lemma 10.3 (Carathéodory). Für alle $x \in CIP^=(A, b)$ existieren $x^1, \dots, x^{n+1} \in IP^=(A, b)$ und $\lambda_1, \dots, \lambda_{n+1} \geq 0$ mit $\sum_{i=1}^{n+1} \lambda_i = 1$.²

¹Siehe z. B. [Schrijver 1986, Theorem 16.1].

²Siehe z. B. [Hiriart-Urruty und Lemaréchal 2001, Theorem A.1.3.6].

Wir betrachten nun das lineare Optimierungsproblem

$$(CIP) \quad \begin{cases} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } x \in CIP^=(A, b), \end{cases}$$

und zeigen, dass sich unter den Lösungen von (CIP) auch eine Lösung von (IP) befindet.

Satz 10.4. Falls $IP^=(A, b)$ nichtleer ist, so gilt:

- (i) (CIP) ist unbeschränkt genau dann, wenn (IP) unbeschränkt ist;
- (ii) ist (CIP) beschränkt, so hat (CIP) eine Lösung, die auch Lösung von (IP) ist;
- (iii) jede Lösung von (IP) ist auch Lösung von (CIP).

Beweis. Bezeichne mit $\bar{x}^{IP} \in \mathbb{Z}^n$ eine Lösung von (IP) und mit $\bar{\mu}^{IP}$ den optimalen Funktionswert $\bar{\mu}^{IP}$ (der $-\infty$ sein kann, falls (IP) unzulässig ist); ebenso schreiben wir $\bar{x}^{CIP} \in \mathbb{R}^n$ für eine Lösung von (CIP) und $\bar{\mu}^{CIP}$ für den optimalen Wert.

Zu (i): Wegen $IP^=(A, b) \subset CIP^=(A, b)$ ist stets $\bar{\mu}^{CIP} \leq \bar{\mu}^{IP}$. Aus $\bar{\mu}^{IP} = -\infty$ folgt daher sofort $\bar{\mu}^{CIP} = -\infty$. Sei umgekehrt $\bar{\mu}^{CIP} = -\infty$. Dann existiert eine Folge $\{x^k\}_{k \in \mathbb{N}} \subset CIP^=(A, b)$ mit $c^T x^k \rightarrow -\infty$ für $k \rightarrow \infty$. Nach Definition der konvexen Hülle und Lemma 10.3 hat jedes x^k eine Darstellung der Form

$$x^k = \sum_{i=1}^{n+1} \lambda_{k,i} x^{k,i}, \quad x^{k,i} \in IP^=(A, b), \quad \lambda_{k,i} \geq 0 \text{ mit } \sum_{i=1}^{n+1} \lambda_{k,i} = 1.$$

Also gilt

$$\sum_{i=1}^{n+1} \lambda_{k,i} c^T x^{k,i} = c^T x^k \rightarrow -\infty \quad \text{für } k \rightarrow \infty.$$

Das ist aber nur möglich, wenn $c^T x^{k_i, j} \rightarrow -\infty$ für ein $1 \leq j \leq n+1$ und eine Teilfolge $\{k_i\}_{i \in \mathbb{N}}$ gilt. (Lemma 10.3 garantiert dabei, dass die Anzahl der Summanden für $k \rightarrow \infty$ beschränkt bleibt und daher mindestens ein Summand gegen $-\infty$ tendieren muss.) Wegen $x^{k,i} \in IP^=(A, b)$ muss dafür auch $\bar{\mu}^{IP} = -\infty$ sein.

Zu (ii): Angenommen, \bar{x}^{CIP} ist Lösung von (CIP) aber keine Lösung von (IP). Dann gilt $c^T \bar{x}^{CIP} < c^T x$ für alle $x \in IP^=(A, b)$ (der optimale Funktionswert für (IP) kann ja nicht kleiner sein, und bei Gleichheit wäre \bar{x}^{CIP} eine Lösung im Gegensatz zur Annahme). Wieder können wir $\bar{x}^{CIP} \in CIP^=(A, b)$ schreiben als

$$\bar{x}^{CIP} = \sum_{i=1}^{n+1} \lambda_i x^i, \quad x^i \in IP^=(A, b), \quad \lambda_i \geq 0 \text{ mit } \sum_{i=1}^{n+1} \lambda_i = 1.$$

Daraus folgt aber

$$c^T \bar{x}^{CIP} = \sum_{i=1}^{n+1} \lambda_i c^T x^i > \sum_{i=1}^{n+1} \lambda_i c^T \bar{x}^{CIP} = c^T \bar{x}^{CIP},$$

was einen Widerspruch darstellt. Also muss \bar{x}^{CIP} Lösung von (IP) sein.

Zu (iii): Sei umgekehrt \bar{x}^{IP} Lösung von (IP) und $x \in CIP^=(A, b)$ beliebig. Dann ist

$$x = \sum_{i=1}^{n+1} \lambda_i x^i, \quad x^i \in IP^=(A, b), \quad \lambda_i \geq 0 \text{ mit } \sum_{i=1}^{n+1} \lambda_i = 1$$

und damit

$$c^T x = \sum_{i=1}^{n+1} \lambda_i c^T x^i \geq \sum_{i=1}^{n+1} \lambda_i c^T \bar{x}^{IP} = c^T \bar{x}^{IP},$$

d. h. $\bar{x}^{IP} \in IP^=(A, b) \subset CIP^=(A, b)$ ist Lösung von (CIP). □

Im Prinzip können wir also ein Simplex-Verfahren auf (CIP) anwenden und damit auch eine Lösung von (IP) bestimmen. Allerdings ist eine Polyederdarstellung der konvexen Hülle $CIP^=(A, b)$ im Allgemeinen schwer zu finden. Die Idee von *Schnittebenen-Verfahren* ist nun, $CIP^=(A, b)$ von aussen anzunähern, indem durch sukzessives Hinzufügen von Nebenbedingungen – den *Schnittebenen* – diejenigen Teile von $P^=(A, b)$ unzulässig gemacht werden, die nicht Teil von $IP^=(A, b)$ sind. Die einzelnen Verfahren unterscheiden sich dabei in der Art, wie diese Bedingungen konstruiert werden. Wir betrachten hier nur die heute am weitesten verbreitete Methode, die auf Gomory und Chvátal zurückgeht.³ Dazu lösen wir in einem ersten Schritt die LP-Relaxation (LP) mit Hilfe des Simplex-Verfahrens und erhalten eine Lösung \bar{x} mit zugehöriger Basis B und Nichtbasis N, d. h. es gilt

$$A_B \bar{x}_B = b, \quad \bar{x}_N = 0.$$

Ist $\bar{x}_B \in \mathbb{Z}^m$, so ist \bar{x} bereits eine Lösung von (IP), und wir sind fertig. Ansonsten sei $x \in P^=(A, b)$ ein beliebiger zulässiger Vektor. Wegen der Invertierbarkeit von A_B ist die Bedingung $A_B x_B + A_N x_N = Ax = b$ äquivalent mit

$$(10.1) \quad x_B + (A_B^{-1} A_N) x_N = A_B^{-1} b = \bar{x}_B.$$

Wegen $\bar{x}_B \notin \mathbb{Z}^m$ existiert nun ein $j \in B$ mit $\bar{x}_j \notin \mathbb{Z}$. Die Idee ist jetzt, die entsprechende Gleichung in (10.1) zu runden und als neue ganzzahlige *Ungleichung* dem Problem hinzuzufügen. Dafür definieren wir die *Gauß-Klammer* für $\alpha \in \mathbb{R}$ durch

$$\lfloor \alpha \rfloor := \max\{m \in \mathbb{Z} : m \leq \alpha\},$$

³Publiziert in [Gomory 1958] und [Chvátal 1973].

und für $x \in \mathbb{R}^n$ durch

$$\lfloor x \rfloor := (\lfloor x_1 \rfloor, \dots, \lfloor x_n \rfloor)^T.$$

Dann gilt wegen $\lfloor x \rfloor \leq x$ für alle $x \in P^=(A, b)$

$$x_j + \lfloor [A_B^{-1} A_N]_j x_N \rfloor \leq x_j + [A_B^{-1} A_N]_j x_N = \bar{x}_j.$$

Für $x \in IP^=(A, b)$ ist die linke Seite stets ganzzahlig, und daher muss sogar gelten

$$(10.2) \quad x_j + \lfloor [A_B^{-1} A_N]_j x_N \rfloor \leq \lfloor \bar{x}_j \rfloor.$$

Dagegen ist

$$\bar{x}_j + \lfloor [A_B^{-1} A_N]_j \bar{x}_N \rfloor = \bar{x}_j > \lfloor \bar{x}_j \rfloor,$$

denn $\bar{x}_N = 0$ und \bar{x}_j war als nicht ganzzahlig vorausgesetzt. Damit ist (10.2) unsere gesuchte neue Ungleichung. Wir setzen daher $a \in \mathbb{Z}^n$ als

$$a_N := \lfloor [A_B^{-1} A_N]_j \rfloor, \quad a_j := 1, \quad a_{B \setminus \{j\}} = 0$$

sowie

$$\beta := \lfloor \bar{x}_j \rfloor \in \mathbb{Z},$$

und wiederholen die beschriebene Prozedur für das modifizierte Problem

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ \quad x \geq 0, \\ \quad a^T x \leq \beta, \end{array} \right.$$

welches dieselbe Form hat wie Problem (8.2) und daher nach Einführen einer – ebenfalls ganzzahligen – Schlupfvariablen effizient mit dem dualen Simplex-Verfahren gelöst werden kann. Wählt man den Index j (falls \bar{x} mehr als eine nicht-ganzzahlige Komponente besitzt) nach einer lexikographischen Regel, kann man zeigen, dass dieses Verfahren nach endlich vielen Schritten terminiert.

10.2 BRANCH-AND-BOUND-VERFAHREN

Ein alternativer Ansatz ist, $P^=(A, b)$ in kleine Quader zu unterteilen, deren Ecken dann nach Konstruktion ganzzahlig sind. Ein Simplex-Verfahren für die auf diesen Quader eingeschränkte LP-Relaxation liefert damit stets eine ganzzahlige Lösung. Die gesuchte Lösung von (IP) ist

dann diejenige unter allen Teilproblem-Lösungen, die den kleinsten Funktionswert liefert. Allerdings wäre der Aufwand für eine naive Aufzählung und Lösung der Teilprobleme selbst für moderate Dimensionen astronomisch, so dass man schlauer vorgehen muss. Die Grundidee von *Branch-and-Bound-Verfahren* (die auch für allgemeinere diskrete Optimierungsprobleme anwendbar ist) besteht darin, hierarchisch vorzugehen und frühzeitig zu überprüfen, ob sich eine weitere Unterteilung lohnt.

Wir gehen dazu rekursiv vor. Als Start lösen wir die LP-Relaxation (LP) mit dem (primalen) Simplex-Verfahren. Ist (LP) unzulässig, so muss dies natürlich auch für (IP) gelten. Ist (LP), unbeschränkt, so muss (IP) dies allerdings nicht sein; allerdings kann man für ganzzahlige A und b zeigen, dass (IP) entweder unbeschränkt oder unzulässig sein muss.⁴ Wir können also annehmen, eine Lösung \bar{x} mit Funktionswert $\bar{\mu} = c^T \bar{x}$ gefunden zu haben. Ist diese bereits ganzzahlig, so sind wir fertig; ansonsten existiert mindestens eine Komponente $\bar{x}_j \notin \mathbb{Z}$. Wir zerlegen nun den zulässigen Bereich durch Hinzunahme von Ungleichungen in zwei Teilbereiche, von denen *keiner* die nicht-ganzzahlige Lösung enthält: Wir betrachten

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \\ x_j \leq \lfloor \bar{x}_j \rfloor, \end{array} \right. \quad \text{und} \quad \left\{ \begin{array}{l} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \\ x_j \geq \lfloor \bar{x}_j \rfloor + 1, \end{array} \right.$$

und lösen diese (mit dem dualen Simplexverfahren, da sie sich von (LP) nur durch Zunahme einer Ungleichung unterscheiden). Dadurch erhalten wir zwei weitere Lösungen \bar{x}^- mit Funktionswert $\bar{\mu}^-$ bzw. \bar{x}^+ mit Funktionswert $\bar{\mu}^+$. Ist eine von beiden Lösungen ganzzahlig, so merken wir uns diese als möglichen Kandidaten. (Im Gegensatz zum Schnittebenen-Verfahren sind ganzzahlige Teillösungen *keine* Lösungen von (IP)!) Ansonsten unterteilen wir weiter: Ist zum Beispiel $\bar{x}_i^- \notin \mathbb{Z}$, so betrachten wir

$$\left\{ \begin{array}{l} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \\ x_j \leq \lfloor \bar{x}_j \rfloor, \\ x_i \leq \lfloor \bar{x}_i^- \rfloor, \end{array} \right. \quad \text{und} \quad \left\{ \begin{array}{l} \min_{x \in \mathbb{R}^n} c^T x \\ \text{mit } Ax = b, \\ x \geq 0, \\ x_j \leq \lfloor \bar{x}_j \rfloor, \\ x_i \geq \lfloor \bar{x}_i^- \rfloor + 1, \end{array} \right.$$

und fahren fort wie beschrieben. Dadurch erzeugen wir einen binären Baum, deren Knoten lineare Optimierungsprobleme bzw. deren Lösungen darstellen – daher der Ausdruck *branch*, „verzweigen“.

Der Clou des Verfahrens liegt im *bound*-Teil, der unnötige Verzweigungen vermeidet. Da bei jeder Verzweigung lediglich Restriktionen hinzugefügt werden, kann der optimale Funktionswert der Teilprobleme höchstens größer werden als der des Ausgangsproblems. Hat

⁴siehe z. B. [Byrd u. a. 1987]

daher ein Teilproblem – nicht notwendigerweise mit ganzzahliger Lösung – einen optimalen Funktionswert $\bar{\mu} \geq \bar{\mu}^{\text{opt}}$ (dem bisher besten Funktionswert aller ganzzahligen Teillösungen), so muss das auch für alle weiteren Unterteilungen gelten und wir müssen diesen Zweig gar nicht weiterverfolgen. Ebenso muss ein unzulässiges Problem nicht weiter unterteilt werden. (Unbeschränkte Teilprobleme können dagegen nicht auftreten, denn dafür müsste ja (LP) bereits unbeschränkt gewesen sein.)

Das gesamte Verfahren ist in Algorithmus 10.1 zusammengefasst; dabei bezeichne $\bar{\mu}$ den Wert eines Teilproblems, wobei $\bar{\mu} = \infty$ für unzulässige Probleme und $\bar{\mu} = -\infty$ für unbeschränkte Probleme zugelassen ist; ansonsten liefert das Simplex-Verfahren (primal für $P=(A, b)$, dual sonst) eine Teillösung \bar{x} .

Algorithmus 10.1 : Branch-and-Bound-Verfahren

Input : $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$, $c \in \mathbb{Z}^n$

Output : Lösung $x \in \mathbb{Z}^n$ oder Information „Problem nicht lösbar“

```

1 Setze  $U = \{P=(A, b)\}$ ,  $\mu = \infty$ 
2 repeat
3   Wähle  $P \in U$ 
4   Wende (duales) Simplex-Verfahren an auf  $\min_{x \in P} c^T x$ 
5   if  $\bar{\mu} = -\infty$  then return Problem nicht lösbar
6   if  $\bar{\mu} < \mu$  then
7     if  $\bar{x} \in \mathbb{Z}^n$  then                                     // bound
8       | Setze  $\mu = \bar{\mu}$ ,  $x = \bar{x}$ 
9     else                                                 // branch
10      | Wähle  $j$  mit  $\bar{x}_j \notin \mathbb{Z}$ 
11      | Setze  $U = U \cup \{P \cap \{x : x_j \leq \lfloor \bar{x}_j \rfloor\}, P \cap \{x : x_j \geq \lfloor \bar{x}_j \rfloor + 1\}\}$ 
12    Setze  $U = U \setminus P$ 
13 until  $U = \emptyset$ 
14 if  $\mu < \infty$  then
15   | return  $x$ 
16 else
17   | return Problem unzulässig

```

Dabei bleiben noch einige Details offen. Am wichtigsten ist eine effiziente Datenstruktur für U , denn naives Abspeichern aller Problemdaten erfordert inakzeptablen Speicheraufwand. Auch die Reihenfolge, in der die Probleme abgearbeitet werden, beeinflusst die Effizienz. Hier hat sich eine *Tiefensuche* bewährt, in der U als FIFO-Speicher („*first in, first out*“) realisiert wird. Dies hat gleich mehrere Vorteile: Zum einen werden dabei so lange wie möglich dem gleichen Problem weitere Ungleichungen hinzugefügt, was die Anwendung des dualen Simplex-Verfahrens erleichtert. Ausserdem erhöht dies die Chance, ganzzahlige Lösungen zu erhalten, wodurch mehr Teilprobleme anhand ihres Optimalwerts ausgeschlossen werden können. Genauso wichtig ist die Wahl von j , falls mehr als eine Komponente nicht ganzzahlig

ist. Dafür existieren eine Reihe von Heuristiken, auf die hier nicht näher eingegangen werden soll. In diesem Kontext gibt es auch noch die Möglichkeit, in bestimmten Fällen statt einem „branch“ eine Schnittebene einzuführen; die Kombination wird als *Branch-and-Cut-Verfahren* bezeichnet und ist in nahezu jeder Software für ganzzahlige Optimierungsprobleme implementiert.

LITERATUR

- Y. Boykov und V. Kolmogorov (2004). *An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision*. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9), S. 1124–1137. DOI: [10.1109/TPAMI.2004.60](https://doi.org/10.1109/TPAMI.2004.60).
- R. H. Byrd, A. J. Goldman und M. Heller (1987). *Recognizing unbounded integer programs*. Oper. Res. 35.1, S. 140–142. DOI: [10.1287/opre.35.1.140](https://doi.org/10.1287/opre.35.1.140).
- V. Chvátal (1973). *Edmonds polytopes and a hierarchy of combinatorial problems*. Discrete Math. 4, S. 305–337. DOI: [10.1016/0012-365X\(73\)90167-2](https://doi.org/10.1016/0012-365X(73)90167-2).
- L. R. Ford Jr. und D. R. Fulkerson (1956). *Maximal flow through a network*. Canad. J. Math. 8, S. 399–404. DOI: [10.4153/CJM-1956-045-5](https://doi.org/10.4153/CJM-1956-045-5).
- C. Geiger und C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, Berlin. DOI: [10.1007/978-3-642-56004-0](https://doi.org/10.1007/978-3-642-56004-0).
- G. H. Golub und C. F. Van Loan (2013). *Matrix Computations*. 4. Aufl. Johns Hopkins University Press, Baltimore, MD.
- R. E. Gomory (1958). *Outline of an algorithm for integer solutions to linear programs*. Bull. Amer. Math. Soc. 64, S. 275–278. DOI: [10.1090/S0002-9904-1958-10224-4](https://doi.org/10.1090/S0002-9904-1958-10224-4).
- P. Gritzmann (2014). *Grundlagen der Mathematischen Optimierung*. Springer, Berlin. DOI: [10.1007/978-3-8348-2011-2](https://doi.org/10.1007/978-3-8348-2011-2).
- M. Grötschel (2010). „Lineare und Ganzzahlige Programmierung (ADM II)“. Vorlesungsskript, Institut für Mathematik, Technische Universität Berlin. URL: <http://www3.math.tu-berlin.de/Vorlesungen/WS09/LinOpt/index.de.html>.
- J.-B. Hiriart-Urruty und C. Lemaréchal (2001). *Fundamentals of convex analysis*. Springer, Berlin. DOI: [10.1007/978-3-642-56468-0](https://doi.org/10.1007/978-3-642-56468-0).
- C. Kanzow (2010). „Operations Research“. Vorlesungsskript, Institut für Mathematik, Universität Würzburg. URL: http://www.mathematik.uni-wuerzburg.de/~kanzow/OR_10.html.
- J. F. Nash (1950). *Equilibrium points in n-person games*. Proc. Natl. Acad. Sci. USA 36, S. 48–49. DOI: [10.1073/pnas.36.1.48](https://doi.org/10.1073/pnas.36.1.48).
- A. Schrijver (1986). *Theory of linear and integer programming*. John Wiley & Sons, Ltd., Chichester.

- A. Schrijver (2002). *On the history of the transportation and maximum flow problems*. *Mathematical Programming* 91.3, S. 437–445. DOI: [10.1007/s101070100259](https://doi.org/10.1007/s101070100259).
- R. Schultz (2013). „Optimierung 1“. Vorlesungsskript, Fakultät für Mathematik, Universität Duisburg-Essen.
- S. Ulbrich (2013). „Optimierung I“. Vorlesungsskript, Institut für Mathematik, Technische Universität Darmstadt. URL: <https://www3.mathematik.tu-darmstadt.de/index.php?id=84&evsver=1007&evsid=23&>.
- Z. Wang, B. Xu und H.-J. Zhou (2014). *Social cycling and conditional responses in the Rock-Paper-Scissors game*. arXiv: [1404.5199](https://arxiv.org/abs/1404.5199).