

Einführung in die  
Lineare und Kombinatorische Optimierung  
(Algorithmische Diskrete Mathematik I, kurz ADM I)

Skriptum zur Vorlesung im WS 2014/2015

Prof. Dr. Martin Grötschel  
Institut für Mathematik  
Technische Universität Berlin

finale Version vom 16. Februar 2015



## Vorwort

Bei dem vorliegenden Skript handelt es sich um die Ausarbeitung der vierstündigen Vorlesung „Einführung in die Lineare und Kombinatorische Optimierung“ (mit zugehörigen Übungen und Tutorien), die die grundlegende Vorlesung des dreisemestrigen Zyklus „Algorithmische Diskrete Mathematik“ bildet. Diese Vorlesung wurde von mir im Wintersemester 2014/15 zusammen mit Axel Werner an der TU Berlin gehalten, der auch an der Überarbeitung des vorliegenden Vorlesungsskripts beteiligt war.

Das erste Ziel dieser Vorlesung ist, das Verständnis für Fragestellungen der mathematischen Optimierung und deren Anwendungen zu wecken und das Vorgehen bei der mathematischen Modellierung von Optimierungsproblemen aus der Praxis kennenzulernen. Das zweite und wichtigere Ziel ist die Einführung in die Methoden zur Lösung derartiger Probleme. Die erste Vorlesung des Zyklus vermittelt Grundlagen der Theorie der Graphen und Netzwerke sowie der linearen, kombinatorischen und ganzzahligen Optimierung. Hierbei wird auf algorithmische Aspekte besonderer Wert gelegt.

Grundkenntnisse der linearen Algebra werden vorausgesetzt, die aus der diskreten Mathematik (vornehmlich Graphentheorie) benötigten Begriffe und Grundlagen werden in der Vorlesung vorgestellt.

In der Vorlesung werden insbesondere kombinatorische Optimierungsprobleme behandelt, die sich graphentheoretisch formulieren lassen, wobei vornehmlich Probleme untersucht werden, die mit Hilfe polynomialer Algorithmen gelöst werden können. Hierzu gehört natürlich eine Einführung in die Komplexitätstheorie (die Klassen  $\mathcal{P}$  und  $\mathcal{NP}$ ,  $\mathcal{NP}$ -Vollständigkeit). Es werden gleichfalls einige Heuristiken und Approximationsverfahren für „schwere“ Probleme vorgestellt sowie Methoden zu deren Analyse. Mit der Darstellung des Simplexalgorithmus und einiger seiner theoretischen Konsequenzen (Dualitätssatz, Sätze vom komplementären Schlupf) beginnt die Einführung in die lineare Optimierung, wobei auch bereits einige Aspekte der ganzzahligen Optimierung behandelt werden.

Es gibt kein einzelnes Buch, das den gesamten, in dieser Vorlesung abgehandelten Themenkreis abdeckt. Daher sind in die einzelnen Kapitel Literaturhinweise eingearbeitet worden. Hinweise auf aktuelle Lehrbücher, die als Begleittexte zur Vorlesung geeignet sind finden sich auf der zur Vorlesung gehörigen Webseite:

<http://www.zib.de/groetschel/teaching/WS1415/VL-WS1415.htm>

Die vorliegende Ausarbeitung ist ein Vorlesungsskript und kein Buch. Obwohl mit der gebotenen Sorgfalt geschrieben, war nicht genügend Zeit für das bei Lehrbüchern notwendige intensive Korrekturlesen und das Einarbeiten umfassender Literaturhinweise. Die daher vermutlich vorhandenen Fehler bitte ich zu entschuldigen (und mir wenn möglich mitzuteilen). Das Thema wird nicht erschöpfend behandelt. Das Manuskript enthält nur die wesentlichen Teile der Vorlesung. Insbesondere sind die Schilderungen komplexer Anwendungsfälle, der Schwierigkeiten bei der Modellierung praktischer Probleme, der Probleme bei der praktischen Umsetzung und die Darstellung der Erfolge, die in den letzten Jahren beim Einsatz der hier vorgestellten Methodik in der Industrie erzielt wurden, nicht in das Skript aufgenommen worden.

Martin Grötschel



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	Einführendes Beispiel . . . . .	1
1.2	Optimierungsprobleme . . . . .	6
<b>2</b>	<b>Grundlagen und Notation</b>	<b>11</b>
2.1	Graphen und Digraphen: Wichtige Definitionen und Bezeichnungen . . . .	11
2.1.1	Grundbegriffe der Graphentheorie . . . . .	11
2.1.2	Graphen . . . . .	11
2.1.3	Digraphen . . . . .	15
2.1.4	Ketten, Wege, Kreise, Bäume . . . . .	16
2.2	Lineare Algebra . . . . .	20
2.2.1	Grundmengen . . . . .	20
2.2.2	Vektoren und Matrizen . . . . .	21
2.2.3	Kombinationen von Vektoren, Hüllen, Unabhängigkeit . . . . .	25
2.3	Polyeder und lineare Programme . . . . .	26
<b>3</b>	<b>Diskrete Optimierungsprobleme</b>	<b>37</b>
3.1	Kombinatorische Optimierungsprobleme . . . . .	37
3.2	Klassische Fragestellungen der Graphentheorie . . . . .	38
3.3	Graphentheoretische Optimierungsprobleme: Beispiele . . . . .	42
<b>4</b>	<b>Komplexitätstheorie und Speicherung von Graphen</b>	<b>59</b>
4.1	Probleme, Komplexitätsmaße, Laufzeiten . . . . .	59
4.2	Die Klassen $\mathcal{P}$ und $\mathcal{NP}$ , $\mathcal{NP}$ -Vollständigkeit . . . . .	62
4.3	Datenstrukturen zur Speicherung von Graphen . . . . .	69
4.3.1	Kanten- und Bogenlisten . . . . .	69
4.3.2	Adjazenzmatrizen . . . . .	71
4.3.3	Adjazenzlisten . . . . .	71
<b>5</b>	<b>Bäume und Wege</b>	<b>77</b>
5.1	Graphentheoretische Charakterisierungen . . . . .	77
5.2	Optimale Bäume und Wälder . . . . .	81
5.3	Kürzeste Wege . . . . .	89
5.3.1	Ein Startknoten, nichtnegative Gewichte . . . . .	91
5.3.2	Ein Startknoten, beliebige Gewichte . . . . .	94
5.3.3	Kürzeste Wege zwischen allen Knotenpaaren . . . . .	100
5.3.4	Min-Max-Sätze und weitere Bemerkungen . . . . .	103

5.4	LP/IP-Aspekte von Bäumen und Wegen . . . . .	106
5.5	Exkurs: Greedy-Heuristiken für das Rucksackproblem und deren Analyse .	109
<b>6</b>	<b>Maximale Flüsse in Netzwerken</b>	<b>115</b>
6.1	Das Max-Flow-Min-Cut-Theorem . . . . .	116
6.2	Der Ford-Fulkerson-Algorithmus . . . . .	119
6.3	Einige Anwendungen . . . . .	125
<b>7</b>	<b>Flüsse mit minimalen Kosten</b>	<b>129</b>
7.1	Flüsse mit minimalen Kosten . . . . .	129
7.2	Transshipment-, Transport- u. Zuordnungsprobleme . . . . .	137
7.3	Der Netzwerk-Simplex-Algorithmus . . . . .	139
<b>8</b>	<b>Grundlagen der Polyedertheorie</b>	<b>151</b>
<b>9</b>	<b>Die Grundversion des Simplex-Algorithmus</b>	<b>159</b>
9.1	Basen, Basislösungen, Entartung . . . . .	160
9.2	Basisaustausch (Pivoting), Simplexkriterium . . . . .	164
9.3	Das Simplexverfahren . . . . .	170
9.4	Spalten- und Zeilenauswahlregeln . . . . .	181
9.5	Die Phase I . . . . .	184
<b>10</b>	<b>Fourier-Motzkin-Elimination und Projektion</b>	<b>191</b>
10.1	Fourier-Motzkin-Elimination . . . . .	191
10.2	Lineare Optimierung und Fourier-Motzkin-Elimination . . . . .	198
<b>11</b>	<b>Das Farkas-Lemma und Dualitätstheorie</b>	<b>201</b>
11.1	Verschiedene Versionen des Farkas-Lemmas . . . . .	201
11.2	Alternativ- und Transpositionssätze . . . . .	203
11.3	Trennsätze . . . . .	204
11.4	Der Dualitätssatz der linearen Programmierung . . . . .	207
<b>12</b>	<b>Varianten der Simplex-Methode</b>	<b>217</b>
12.1	Der revidierte Simplexalgorithmus . . . . .	217
12.2	Die Behandlung oberer Schranken . . . . .	219
12.3	Das duale Simplexverfahren . . . . .	226
12.4	Postoptimierung und parametrische Programme . . . . .	232
12.5	Zur Numerik des Simplexverfahrens . . . . .	239
<b>13</b>	<b>Ganzzahligkeit von Polyedern und Totale Unimodularität</b>	<b>249</b>

# 1 Einführung

## 1.1 Einführendes Beispiel

Ein Unternehmen produziert am Standort  $A$  ein Gut, das es mit der Bahn zu den Städten  $B$ ,  $C$  und  $D$  transportieren möchte. Genauer, es werden wöchentlich 6 Waggons benötigt, von denen 2 nach  $B$ , einer nach  $C$  sowie 3 nach  $D$  befördert werden müssen. Auf Anfrage des Unternehmens nennt die Bahn die maximalen wöchentlichen Beförderungskapazitäten (in Waggons) sowie die Kosten pro Waggon, siehe Tabelle 1.1.

Um sich die Situation zu veranschaulichen, macht der verantwortliche Planer eine Zeichnung der im Netz möglichen Verbindungen, ihrer Kapazitäten, sowie ihrer Kosten, siehe Abbildung 1.1. Diese Abbildung repräsentiert einen *gerichteten Graphen* (auch *Digraph* genannt)  $D = (V, A)$ , der aus *Knoten*  $V$  und *Bögen* (auch *gerichtete Kanten* genannt)  $A \subseteq V \times V$  besteht, wobei die Bögen die Knoten miteinander verbinden. Die Knoten entsprechen dabei den Städten  $A, B, C$  und  $D$ , die Bögen den möglichen Verbindungen zwischen diesen Städten. Zusätzlich zu dieser Information enthält der Graph in Abbildung 1.1 weitere Daten, die abstrakt als *Bogengewichte* bezeichnet werden und jedem Bogen gewisse von der jeweiligen Fragestellung abhängige Werte zuordnen. Konkret haben wir die Bogengewichte  $l: A \rightarrow \mathbb{R}_+$ ,  $u: A \rightarrow \mathbb{R}_+$  und  $c: A \rightarrow \mathbb{R}_+$ , die jeweils die Mindestzahl der Waggons pro Woche (hier immer 0), maximale Anzahl der Waggons pro Woche sowie Kosten pro Waggon auf jeder Verbindung angeben. (Die „Benamung“ der Variablen und Daten erfolgt entsprechend dem üblichen Vorgehen in der englischsprachigen Literatur; „ $V$ “ steht für „vertices“, „ $A$ “ für „arcs“, „ $l$ “ und „ $u$ “ für „lower bound“ und „upper bound“, „ $c$ “ für „cost“ usw.)

Der Planer möchte nun die kostengünstigste Möglichkeit bestimmen, die 6 Waggons zu ihren Zielorten zu transportieren. Dies ist ein typisches Problem der *kombinatorischen Optimierung*, welche sich (ganz allgemein formuliert) damit beschäftigt, aus einer endlichen Menge, deren Elemente „Lösungen“ genannt werden und die bewertet sind, eine

Verbindung	maximale Anzahl Waggons pro Woche	Kosten pro Waggon
$A$ nach $B$	4	5
$A$ nach $C$	3	1
$B$ nach $C$	2	2
$B$ nach $D$	3	2
$C$ nach $D$	4	1

Tabelle 1.1: Kapazitäten und Kosten der Verbindungen im Bahnnetz.

## 1 Einführung

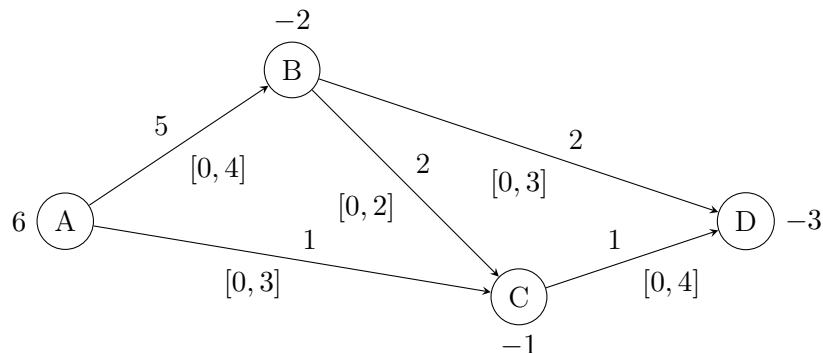


Abbildung 1.1: Beispielproblem als gerichteter Graph. Die Zahl an einem Knoten gibt an, wieviele Waggonen dort bereitgestellt werden (positiv) bzw. wieviele Waggonen angeliefert werden (negativ). Zahlen oberhalb eines Bogens geben die Transportkosten pro Waggon, Intervalle unterhalb eines Bogens die Kapazitäten der Verbindung an.

beste Lösung auszusuchen. Die in der kombinatorischen Optimierung auftretenden Probleme sind in der Regel „strukturiert“. In unserem Fall ist eine Grundstruktur durch einen Digraphen und die Funktionen  $l$ ,  $u$  und  $c$  gegeben. Die Menge der Lösungen, über die optimiert werden soll, besteht aus der Menge aller möglichen Transporte von  $A$  zu den Zielen  $B$ ,  $C$  und  $D$ . Kombinatorische Optimierungsprobleme können durch Enumeration aller Lösungen gelöst werden. Dies ist in diesem Beispiel möglich, aber im Allgemeinen keine gute Idee. In der Vorlesung geht es u. a. darum, mathematische Methoden zu entwickeln, um derartige Probleme (einigermaßen) effizient zu lösen.

In unserem Beispiel überlegt sich nun der Planer, dass er lediglich entscheiden muss, wieviele Waggonen auf jeder der 5 Verbindungen befördert werden sollen. Dazu führt er die Variablen  $f_{AB}$ ,  $f_{AC}$ ,  $f_{BC}$ ,  $f_{BD}$  und  $f_{CD}$  ein. „ $f$ “ steht für „flow“ oder „Fluss“. Er überlegt sich, dass jede der Variablen mindestens den Wert 0 haben muss und nicht größer als die von der Bahn vorgegebene Kapazitätsgrenze sein darf. Außerdem notiert er sich, dass jede Variable nur ganzzahlige Werte annehmen kann. Insgesamt gelangt er so zu den *Bedingungen*

$$\begin{aligned} f_{AB} &\in \mathbb{Z}, & 0 \leq f_{AB} \leq 4, \\ f_{AC} &\in \mathbb{Z}, & 0 \leq f_{AC} \leq 3, \\ f_{BC} &\in \mathbb{Z}, & 0 \leq f_{BC} \leq 2, \\ f_{BD} &\in \mathbb{Z}, & 0 \leq f_{BD} \leq 3, \\ f_{CD} &\in \mathbb{Z}, & 0 \leq f_{CD} \leq 4. \end{aligned}$$

Mit diesen Variablen können die Gesamtkosten eines Transportes leicht als

$$5f_{AB} + f_{AC} + 2f_{BC} + 2f_{BD} + f_{CD}$$

dargestellt werden, und das Ziel ist, diese Gesamtkosten so gering wie möglich zu halten. Deswegen spricht man hier auch von der *Zielfunktion* (englisch: objective function). Nun



muss der Planer noch die Bedingungen festlegen, die den gewünschten Transport von  $A$  nach  $B$ ,  $C$  und  $D$  beschreiben. Zunächst müssen 6 Waggons  $A$  verlassen, was sich als Gleichung

$$f_{AB} + f_{AC} = 6$$

schreiben lässt. Von den Waggons, die in  $B$  ankommen, sollen 2 dort verbleiben und der Rest weiter nach  $C$  und  $D$  fahren:

$$f_{AB} = 2 + f_{BC} + f_{BD}.$$

Analog können auch die Bedingungen in  $C$  und  $D$  formuliert werden.

Insgesamt hat der Planer nun folgende mathematische Formulierung vorliegen:

$$\min \quad 5f_{AB} + f_{AC} + 2f_{BC} + 2f_{BD} + f_{CD} \quad (1.1a)$$

$$f_{AB} + f_{AC} = 6 \quad (1.1b)$$

$$-f_{AB} + f_{BC} + f_{BD} = -2 \quad (1.1c)$$

$$-f_{AC} - f_{BC} + f_{CD} = -1 \quad (1.1d)$$

$$-f_{BD} - f_{CD} = -3 \quad (1.1e)$$

$$0 \leq f_{AB} \leq 4 \quad (1.1f)$$

$$0 \leq f_{AC} \leq 3 \quad (1.1g)$$

$$0 \leq f_{BC} \leq 2 \quad (1.1h)$$

$$0 \leq f_{BD} \leq 3 \quad (1.1i)$$

$$0 \leq f_{CD} \leq 4 \quad (1.1j)$$

$$f_{AB}, f_{AC}, f_{BC}, f_{BD}, f_{CD} \in \mathbb{Z} \quad (1.1k)$$

Ein Optimierungsproblem dieser Art, in dem alle Variablen ganzzahlig, alle Nebenbedingungen lineare Gleichungen oder Ungleichungen sind, und dessen Zielfunktion ebenfalls linear ist, heißt *Ganzzahliges Lineares Programm* oder als englische Abkürzung kurz *ILP* (oft auch nur *IP*, wenn aus dem Kontext klar ist, dass Nebenbedingungen und Zielfunktion linear sind). Sind alle Variablen kontinuierlich, so spricht man von einem *Linearen Programm* oder kurz *LP*. Zum Beispiel ist das Optimierungsproblem (1.1a)–(1.1j) ein LP.

Um nun eine optimale Transportvariante zu bestimmen, beschließt der Planer, die Ganzzahligkeitsbedingungen (1.1k) zunächst zu ignorieren, da dann nur ein lineares Gleichungssystem mit Variablenschranken übrig bleibt. Dem Planer fällt auf, dass die vier Gleichungen (1.1b) bis (1.1e) linear abhängig sind, weil sie sich zu 0 summieren. Man kann sich leicht überlegen, dass eine beliebige Gleichung gestrichen werden kann und die verbleibenden drei Gleichungen linear unabhängig sind. Wie wir aus der Linearen Algebra wissen, ist dann der Lösungsraum des Gleichungssystems ein 2-dimensionaler affiner Unterraum des  $\mathbb{R}^5$ . Mithilfe des Gauss-Algorithmus berechnet der Planer folgende Para-

## 1 Einführung

metrisierung dieses Unterraums:

$$\begin{pmatrix} f_{AB} \\ f_{BC} \\ f_{BD} \\ f_{AC} \\ f_{CD} \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \\ 3 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -1 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}. \quad (1.2)$$

Aus der Parametrisierung (1.2) und den Schranken (1.1f) bis (1.1j) leitet der Planer das folgende LP her:

$$\min -6s + t \quad (1.3a)$$

$$-s + t \geq -1 \quad (1.3b)$$

$$-s + t \leq 1 \quad (1.3c)$$

$$2 \leq s \leq 3 \quad (1.3d)$$

$$0 \leq t \leq 3. \quad (1.3e)$$

Dieses LP ist „äquivalent“ zu LP (1.1a)–(1.1j) in folgendem Sinn: Es gibt eine bijektive Abbildung zwischen den Lösungsräumen der beiden LPs, die mit den durch die Zielfunktionen gegebenen Ordnungen kompatibel ist. Mit anderen Worten: Jede Optimallösung von LP (1.3) entspricht genau einer Optimallösung von LP (1.1a)–(1.1j) und umgekehrt. Daher genügt es, das LP (1.3) zu lösen.

Da in LP (1.3) nur zwei Variablen vorkommen, kann man die zulässige Menge aller Paare  $(s, t)$ , die alle Bedingungen erfüllen, graphisch darstellen (siehe Abbildung 1.2). Aus dieser Abbildung kann man direkt die optimale Lösung ablesen: Die optimale Lösung ist derjenige Punkt der grauen Fläche, der am weitesten in der dem Zielfunktionsvektor entgegengesetzten Richtung liegt (weil minimiert wird). Im Beispiel ist dies der Punkt  $(3, 2)$ , der der Lösung  $f_{AB} = 3$ ,  $f_{AC} = 3$ ,  $f_{BC} = 0$ ,  $f_{BD} = 1$ ,  $f_{CD} = 2$  entspricht. Da alle Werte ganzzahlig sind, ist der Planer zufrieden und kommuniziert den entsprechenden Plan dem Bahnunternehmen.

Ist es immer so, dass bei ganzzahligen Problemdata ein lineares Programm eine Optimallösung besitzt, die ganzzahlig ist? Natürlich nicht! Und deswegen geht unsere Story weiter: Nach einigen Tagen bekommt der Planer von der Bahn den Bescheid, dass die Waggons so leider nicht befördert werden können, weil die Kapazität des Rangierbahnhofs in  $C$  nicht ausreicht. Die Bahn beschreibt die Kapazitätsbeschränkung des Rangierbahnhofs genauer und der Planer bekommt die zusätzliche Bedingung

$$2f_{AC} + f_{BC} \leq 5,$$

die er via (1.2) in die äquivalente Bedingung

$$s + t \leq 4 \quad (1.4)$$

übersetzt. Das resultierende LP (1.3) mit der zusätzlichen Ungleichung (1.4) hat nun keine ganzzahlige Optimallösung mehr (siehe Abbildung 1.2); die Optimallösung ist  $(2.5, 1.5)$  bzw.  $f_{AB} = 3.5$ ,  $f_{AC} = 2.5$ ,  $f_{BC} = 0$ ,  $f_{BD} = 1.5$ ,  $f_{CD} = 1.5$  im Originalproblem.

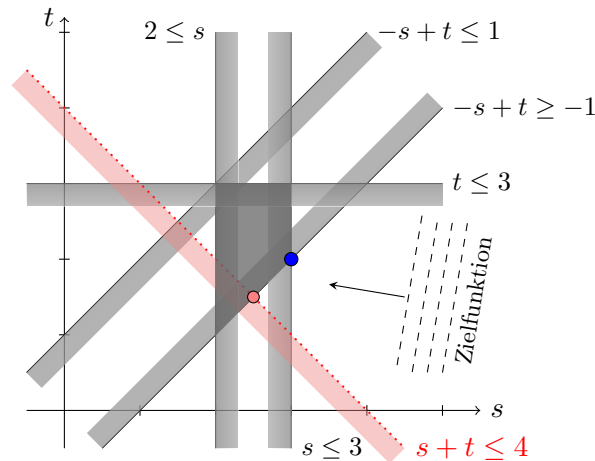


Abbildung 1.2: Graphische Darstellung des Optimierungsproblems (1.3). Die graue Menge ist die Menge der zulässigen Lösungen. Der gestrichelte Pfeil zeigt die Richtung, in der die Zielfunktion ansteigt. Die gepunktete Ungleichung entspricht der zusätzlichen Kapazitätsbedingung der Bahn.

*Schluss, Variante 1:* Der Planer ist nun mit seinem Schulwissen am Ende und fragt seinen Freund, einen Mathematiker, um Rat. Dieser empfiehlt ihm, sein Problem mit der Software SCIP (siehe <http://scip.zib.de>) zur Lösung ganzzahliger Optimierungsprobleme zu lösen. Damit findet der Planer die neue Lösung  $f_{AB} = 4$ ,  $f_{AC} = 2$ ,  $f_{BC} = 0$ ,  $f_{BD} = 2$ ,  $f_{CD} = 1$ , mit der die Bahn einverstanden ist.

*Schluss, Variante 2:* Der Planer schaut sich die zulässige Menge in Abbildung 1.2 genau an und stellt fest, dass nur die ganzzahligen Punkte  $(2, 1)$  und  $(2, 2)$  zulässig sind. Er wählt den billigeren der beiden und gelangt zu der neuen Lösung  $f_{AB} = 4$ ,  $f_{AC} = 2$ ,  $f_{BC} = 0$ ,  $f_{BD} = 2$ ,  $f_{CD} = 1$ , mit der die Bahn einverstanden ist.

Die vorliegende Einführung ist natürlich „nur“ ein didaktisches Beispiel. Wir haben gezeigt, wie man aus einer praktischen Fragestellung (die wir so vereinfacht haben, dass ihre Lösung graphisch gefunden werden kann), ein mathematisches Problem erstellt. Man nennt solch ein Vorgehen *mathematische Modellierung*. Es ist keineswegs so, dass jedem praktischen Problem ein eindeutiges mathematisches Modell entspricht. Es gibt viele unterschiedliche Modellierungsmethoden. Für welche man sich entscheidet, ist eine Frage des Geschmacks, der Vorbildung, oder der Algorithmen, die zur Lösung der mathematischen Modelle verfügbar sind. Ein einfaches Beispiel ist die Modellierung einer ja/nein-Entscheidung: Wir möchten ausdrücken, dass eine reelle Variable  $x$  nur die Werte 0 oder 1 annehmen darf ( $x = 0$  entspricht „nein“,  $x = 1$  entspricht „ja“). Das können wir z. B. auf die folgenden Weisen erreichen:

- $x \in \{0, 1\}$ ,
- $0 \leq x \leq 1$ ,  $x \in \mathbb{Z}$ ,
- $x = x^2$ .

## 1 Einführung

Welche Modellierung sinnvoll ist, kann man nicht grundsätzlich entscheiden. Die Wahl des Modells hängt vom gewählten Gesamtmodell und der geplanten algorithmischen Vorgehensweise ab.

### 1.2 Optimierungsprobleme

Wir wollen zunächst ganz informell, ohne auf technische Spitzfindigkeiten einzugehen, mathematische Optimierungsprobleme einführen. Sehr viele Probleme lassen sich wie folgt formulieren.

Gegeben seien eine Menge  $S$  und eine geordnete Menge  $(T, \leq)$ , d. h. zwischen je zwei Elementen  $s, t \in T$  gilt genau eine der folgenden Beziehungen  $s < t$ ,  $s > t$  oder  $s = t$ . Ferner sei eine Abbildung  $f : S \rightarrow T$  gegeben. Gesucht ist ein Element  $x^* \in S$  mit der Eigenschaft  $f(x^*) \geq f(x)$  für alle  $x \in S$  (Maximierungsproblem) oder  $f(x^*) \leq f(x)$  für alle  $x \in S$  (Minimierungsproblem). Es ist üblich, hierfür eine der folgenden Schreibweisen zu benutzen:

$$\begin{array}{ll} \max_{x \in S} f(x) & \text{oder} \quad \max\{f(x) \mid x \in S\}, \\ \min_{x \in S} f(x) & \text{oder} \quad \min\{f(x) \mid x \in S\}. \end{array} \quad (1.5)$$

In der Praxis treten als geordnete Mengen  $(T, \leq)$  meistens die reellen Zahlen  $\mathbb{R}$ , die rationalen Zahlen  $\mathbb{Q}$  oder die ganzen Zahlen  $\mathbb{Z}$  auf, alle mit der natürlichen Ordnung versehen (die wir deswegen auch gar nicht erst notieren).

Die Aufgabe (1.5) ist viel zu allgemein, um darüber etwas Interessantes sagen zu können. Wenn  $S$  durch die Auflistung aller Elemente gegeben ist, ist das Problem entweder sinnlos oder trivial (man rechnet ganz einfach  $f(x)$  für alle  $x \in S$  aus). Das heißt,  $S$  muss irgendwie (explizit oder implizit) strukturiert sein, so dass vernünftige Aussagen über  $S$  möglich sind, ohne dass man alle Elemente in  $S$  einzeln kennt. Das gleiche gilt für die Funktion  $f : S \rightarrow T$ . Ist sie nur punktweise durch  $x \mapsto f(x)$  gegeben, lohnt sich das Studium von (1.5) nicht. Erst wenn  $f$  durch hinreichend strukturierte „Formeln“ bzw. „Eigenschaften“ bestimmt ist, werden tieferliegende mathematische Einsichten möglich.

Die Optimierungsprobleme, die in der Praxis auftreten, haben fast alle irgendeine „vernünftige“ Struktur. Das muss nicht unbedingt heißen, dass die Probleme dadurch auf einfache Weise lösbar sind, aber immerhin ist es meistens möglich, sie in das zur Zeit bekannte und untersuchte Universum der verschiedenen Typen von Optimierungsproblemen einzureihen und zu klassifizieren.

Im Laufe des Studiums werden Ihnen noch sehr unterschiedliche Optimierungsaufgaben begegnen. Viele werden von einem der folgenden Typen sein.

**(1.6) Definition (Kontrollproblem).** *Gegeben sei ein Steuerungsprozess (z. B. die Bewegungsgleichung eines Autos), etwa der Form*

$$\dot{x}(t) = f(t, x(t), u(t)),$$

wobei  $u$  eine Steuerung ist (Benzinzufuhr). Ferner seien eine Anfangsbedingung

$$x(0) = x_0,$$

(z. B.: das Auto steht) sowie eine Endbedingung

$$x(T) = x_1$$

(z. B.: das Auto hat eine Geschwindigkeit von 50 km/h) gegeben. Gesucht ist eine Steuerung  $u$  für den Zeitraum  $[0, T]$ , so dass z. B.

$$\int_0^T |u|^2 dt$$

minimal ist (etwa minimaler Benzinverbrauch). △

**(1.7) Definition (Approximationsproblem).** Gegeben sei eine (numerisch schwierig auszuwertende) Funktion  $f$ , finde ein Polynom  $p$  vom Grad  $n$ , so dass

$$\|f - p\| \quad \text{oder} \quad \|f - p\|_\infty$$

minimal ist. △

**(1.8) Definition (Nichtlineares Optimierungsproblem).** Es seien  $f, g_i$  mit  $i = 1, \dots, m$  sowie  $h_j$  mit  $j = 1, \dots, p$  differenzierbare Funktionen von  $\mathbb{R}^n \rightarrow \mathbb{R}$ , dann heißt

$$\begin{aligned} \min f(x) \\ g_i(x) \leq 0 \quad i = 1, \dots, m \\ h_j(x) = 0 \quad j = 1, \dots, p \\ x \in \mathbb{R}^n \end{aligned}$$

ein nichtlineares Optimierungsproblem. Ist eine der Funktionen nicht differenzierbar, so spricht man von einem nichtdifferenzierbaren Optimierungsproblem. Im Allgemeinen wird davon ausgegangen, dass alle betrachteten Funktionen zumindest stetig sind. △

**(1.9) Definition (Konvexes Optimierungsproblem).** Eine Menge  $S \subseteq \mathbb{R}^n$  heißt konvex, falls gilt: Sind  $x, y \in S$  und ist  $\lambda \in \mathbb{R}$ ,  $0 \leq \lambda \leq 1$ , dann gilt  $\lambda x + (1 - \lambda)y \in S$ . Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt konvex, falls für alle  $\lambda \in \mathbb{R}$ ,  $0 \leq \lambda \leq 1$  und alle  $x, y \in \mathbb{R}^n$  gilt

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y).$$

Ist  $S \subseteq \mathbb{R}^n$  konvex (z. B. kann  $S$  wie folgt gegeben sein:  $S = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\}$ , wobei die  $g_i$  konvexe Funktionen sind), und ist  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  eine konvexe Funktion, dann ist

$$\min_{x \in S} f(x)$$

ein konvexes Minimierungsproblem. △

## 1 Einführung

**(1.10) Definition (Lineares Optimierungsproblem (Lineares Programm)).** Gegeben seien  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{(m,n)}$ ,  $b \in \mathbb{R}^m$ , dann heißt

$$\begin{aligned} \max \quad & c^T x \\ & Ax \leq b \\ & x \in \mathbb{R}^n \end{aligned}$$

lineares Optimierungsproblem. △

**(1.11) Definition (Lineares ganzzahliges Optimierungsproblem).** Gegeben seien  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{(m,n)}$ ,  $b \in \mathbb{R}^m$ , dann heißt

$$\begin{aligned} \max \quad & c^T x \\ & Ax \leq b \\ & x \in \mathbb{Z}^n \end{aligned}$$

lineares ganzzahliges (oder kurz: ganzzahliges) Optimierungsproblem. △

Selbst bei Optimierungsproblemen wie (1.6), ..., (1.11), die nicht sonderlich allgemein erscheinen mögen, kann es sein, dass (bei spezieller Wahl der Zielfunktion  $f$  und der Nebenbedingungen) die Aufgabenstellung (finde ein  $x^* \in S$ , so dass  $f(x^*)$  so groß (oder klein) wie möglich ist) keine vernünftige Antwort besitzt. Es mag sein, dass  $f$  über  $S$  unbeschränkt ist;  $f$  kann beschränkt sein über  $S$ , aber ein Maximum kann innerhalb  $S$  nicht erreicht werden, d. h., das „max“ müsste eigentlich durch „sup“ ersetzt werden.  $S$  kann leer sein, ohne dass dies a priori klar ist, etc. Der Leser möge sich Beispiele mit derartigen Eigenschaften überlegen! Bei der Formulierung von Problemen dieser Art muss man sich also Gedanken darüber machen, ob die betrachtete Fragestellung überhaupt eine sinnvolle Antwort erlaubt.

In unserer Vorlesung werden wir uns lediglich mit den Problemen (1.10) und (1.11) beschäftigen. Das lineare Optimierungsproblem (1.10) ist sicherlich das derzeit für die Praxis bedeutendste Problem, da sich außerordentlich viele und sehr unterschiedliche reale Probleme als lineare Programme formulieren lassen, bzw. durch die Lösung einer endlichen Anzahl von LPs gelöst werden können. Außerdem liegt eine sehr ausgefeilte Theorie vor. Mit den modernen Verfahren der linearen Optimierung können derartige Probleme mit Hunderttausenden (und manchmal sogar mehr) von Variablen und Ungleichungen fast „müheles“ gelöst werden. Dagegen ist Problem (1.11) viel schwieriger. Die Einschränkung der Lösungsmenge auf die zulässigen ganzzahligen Lösungen führt direkt zu einem Sprung im Schwierigkeitsgrad des Problems. Verschiedene spezielle lineare ganzzahlige Programme können in beliebiger Größenordnung gelöst werden. Bei wenig strukturierten allgemeinen Problemen des Typs (1.11) versagen dagegen auch die besten Lösungsverfahren manchmal bereits bei weniger als 100 Variablen und Nebenbedingungen.

Über Kontrolltheorie (Probleme des Typs (1.6)), Approximationstheorie (Probleme des Typs (1.7)), Nichtlineare Optimierung (Probleme des Typs (1.8)) und (1.9) werden

an der TU Berlin Spezialvorlesungen angeboten. Es ist anzumerken, dass sich sowohl die Theorie als auch die Algorithmen zur Lösung von Problemen des Typs (1.6) bis (1.9) ganz erheblich von denen zur Lösung von Problemen des Typs (1.10) und (1.11) unterscheiden.

Ziel dieser Vorlesung ist zunächst, das Verständnis für Fragestellung der Optimierung und deren Anwendungen zu wecken. Die Studierenden sollen darüber hinaus natürlich in die Optimierungstheorie eingeführt werden und einige Werkzeuge theoretischer und algorithmischer Natur zur Lösung von linearen, kombinatorischen und ganzzahligen Optimierungsproblemen kennenlernen. Damit wird grundlegendes Rüstzeug zur Behandlung (Modellierung, numerischen Lösung) von brennenden Fragen der heutigen Zeit bereitgestellt. Die wirklich wichtigen Fragen benötigen zur ihrer Lösung jedoch erheblich mehr Mathematik sowie weitergehendes algorithmisches und informationstechnisches Know-how. Und ohne enge Zusammenarbeit mit Fachleuten aus Anwendungsdisziplinen wird man die meisten Fragen nicht anwendungsadäquat beantworten können. Einige Themen von derzeit großer Bedeutung, bei denen der Einsatz von mathematischer Optimierung (in Kombination mit vielfältigen Analysetechniken und praktischen Erfahrungen aus anderen Disziplinen) wichtig ist, hat der Präsident von acatech (Deutsche Akademie der Technikwissenschaften), Reinhard F. Hüttel, zu Beginn seiner Rede zur Eröffnung der Festveranstaltung 2012 am 16. Oktober 2012 im Konzerthaus am Gendarmenmarkt in Berlin beschrieben:

*„Als wir im letzten Jahr an diesem Ort zur acatech-Festveranstaltung zusammenkamen, war die Energiewende bereits das, was man bürokratisch als „Beschlusslage“ bezeichnen würde. Gut ein Jahr nach diesem Beschluss bestimmen mehr Fragen als Antworten die Diskussion um unsere zukünftige Energieversorgung: Was wollen wir erreichen? - Den schnellstmöglichen Ausstieg aus der Kernenergie? Eine Verlangsamung des Klimawandels? Den raschen Ausbau erneuerbarer Energien? Möglichst hohe Energieeffizienz? Und wer sollte für welches Ziel und welche Maßnahme die Umsetzung verantworten? Und vor allem: Welche Ziele haben Priorität, und, um die dominierende Frage der letzten Tage und Wochen aufzugreifen, zu welchem Preis – im wörtlichen wie im übertragenen Sinne – wollen und können wir diese Ziele erreichen? Die Debatte der vergangenen Zeit hat gezeigt, dass es auf diese Fragen viele Antworten gibt. – In etwa so viele, wie es Interessensgruppen, Verbände, Unternehmen, Parteien und Parlamente gibt. Vielleicht lässt genau deshalb die Umsetzung der Energiewende auf sich warten. Auch die Wissenschaft hat keine endgültigen Antworten auf diese zentrale gesellschaftliche Frage. Aber: Wissenschaft kann helfen, Daten und Fakten zu sichten, ans Licht zu befördern und zu gewichten. Wissenschaft kann Handlungsoptionen darstellen und auch bewerten. Dies tun wir in unseren aktuellen Publikationen zu Fragen der Energie, der Mobilität, aber auch zu vielen anderen Themen. Das Entscheidungsprimat aber – und dies ist mir gerade in diesem Kontext sehr wichtig – lag und liegt bei der Politik.“*

Die mathematische Optimierung ist eine der wissenschaftlichen Disziplinen, die bei der Behandlung der skizzierten Problemfelder und Fragen eingesetzt werden muss. Aber sie

ist hier nur eine Hilfswissenschaft, denn die Formulierung von Gewichtungen und Zielen und die Bewertung von Ergebnissen unterliegen komplexen gesellschaftlichen Debatten. Wir sollten jedoch die Bedeutung der Mathematik hierbei nicht unterschätzen, sie hilft komplexe politische Entscheidungen (ein wenig) rationaler zu machen.

### Terminologie: Optimierung oder Programmierung

Sie werden in der Literatur die beiden Bezeichnungen nebeneinander finden und sich fragen, wieso das so ist. Hier ist eine kurze historische Erläuterung.

In seinem (für das Fach bedeutenden) Buch aus dem Jahre 1963 schreibt George Dantzig, der uns in der Vorlesung noch oft begegnen wird, auf der ersten Seite der Introduction (Chapter 1) Folgendes:

#### 1-2. THE PROGRAMMING PROBLEM

**Industrial production, the flow of resources in the economy, the exertion of military effort in a war theater—all are complexes of numerous interrelated activities. Differences may exist in the goals to be achieved, the particular processes involved, and the magnitude of effort. Nevertheless, it is possible to abstract the underlying essential similarities in the management of these seemingly disparate systems. To do this entails a look at the structure and state of the system, and at the objective to be fulfilled, in order to *construct a statement of the actions to be performed, their timing, and their quantity (called a “program” or “schedule”), which will permit the system to move from a given status toward the defined objective.***

Das Wort „program“ wurde in der Zeit, als die lineare Optimierung in den USA entstand (um 1950), von dem Personenkreis, der sich mit derartigen Fragestellungen beschäftigte, im Sinne des deutschen Wortes „Plan“ benutzt. Und „programming“ war dann nichts anderes als „einen Plan aufstellen“. „Linear Programming“ bezeichnete dann die Aufstellung eines Plans, bei dem die Restriktionen und die Zielfunktion linear sind. So ist dann die Bezeichnung „Lineare Programmierung“ in die deutsche und (in entsprechender Variation) auch in andere Sprachen aufgenommen worden. Wegen der im Deutschen fehlenden Konnotation war damit niemand richtig glücklich, und so kam parallel der Begriff „Lineare Optimierung“ auf. Mittlerweile war aber das Wort „programming“ auch in der Informatik in Verwendung gekommen und bezeichnete etwas anderes. Die Verwirrung (bei Laien) wuchs, und die Informatik-Bedeutung dominierte eindeutig. Mittlerweile hat sich daher auch in der englischsprachigen Welt die Bezeichnung „Linear Optimization“ durchgesetzt. Am besten sieht man das daran, dass die 1973 gegründete (weltweite) Organisation der mathematischen Optimierer, die Mathematical Programming Society (MPS), sich 2010 in Mathematical Optimization Society (MOS) umbenannt hat. Trotzdem heißt das von der MOS herausgegebene Journal, die vermutlich wichtigste Zeitschrift in der Optimierung, weiterhin *Mathematical Programming*. Das Durcheinander wird also bleiben – auch in diesem Manuskript. Aber man kann damit leben.



## 2 Grundlagen und Notation

### 2.1 Graphen und Digraphen: Wichtige Definitionen und Bezeichnungen

Bei der nachfolgenden Zusammenstellung von Begriffen und Bezeichnungen aus der Graphentheorie handelt es sich nicht um eine didaktische Einführung in das Gebiet der diskreten Mathematik. Dieses Kapitel ist lediglich als Nachschlagewerk gedacht, in dem die wichtigsten Begriffe und Bezeichnungen zusammengefasst und definiert sind.

#### 2.1.1 Grundbegriffe der Graphentheorie

Die Terminologie und Notation in der Graphentheorie ist leider sehr uneinheitlich. Wir wollen daher hier einen kleinen Katalog wichtiger graphentheoretischer Begriffe und Bezeichnungen zusammenstellen und zwar in der Form, wie sie (in der Regel) in meinen Vorlesungen benutzt werden. Definitionen werden durch Kursivdruck hervorgehoben. Nach einer Definition folgen gelegentlich (in Klammern) weitere Bezeichnungen, um auf alternative Namensgebungen in der Literatur hinzuweisen.

Es gibt sehr viele Bücher über Graphentheorie. Wenn man zum Beispiel in der Datenbank MATH des Zentralblattes für Mathematik nach Büchern sucht, die den Begriff „graph theory“ im Titel enthalten, erhält man ungefähr 350 Verweise. Bei über 50 Büchern taucht das Wort „Graphentheorie“ im Titel auf. Ich kenne natürlich nicht alle dieser Bücher. Zur Einführung in die mathematische Theorie empfehle ich u. a. Aigner (1984), Bollobás (1998), Diestel (2006)<sup>1</sup>, Bondy and Murty (2008) und West (2005). Stärker algorithmisch orientiert und anwendungsbezogen sind z. B. Ebert (1981), Jungnickel (2013), Walther and Nägler (1987) sowie Krumke and Noltemeier (2005).

Übersichtsartikel zu verschiedenen Themen der Graphentheorie sind in den Handbüchern Graham et al. (1995) und Gross and Yellen (2004) zu finden.

#### 2.1.2 Graphen

Ein *Graph*  $G$  ist ein Tripel  $(V, E, \Psi)$  bestehend aus einer nicht-leeren Menge  $V$ , einer Menge  $E$  und einer *Inzidenzfunktion*  $\Psi : E \rightarrow V^{(2)}$ . Hierbei bezeichnet  $V^{(2)}$  die Menge der ungeordneten Paare von (nicht notwendigerweise verschiedenen) Elementen von  $V$ . Ein Element aus  $V$  heißt *Knoten* (oder *Ecke* oder *Punkt* oder *Knotenpunkt*; englisch: *vertex* oder *node* oder *point*), ein Element aus  $E$  heißt *Kante* (englisch: *edge* oder *line*). Zu jeder Kante  $e \in E$  gibt es also Knoten  $u, v \in V$  mit  $\Psi(e) = uv = vu$ . (In der Literatur

<sup>1</sup> <http://www.math.uni-hamburg.de/home/diestel/books/graphentheorie/GraphentheorieIII.pdf>

werden auch die Symbole  $[u, v]$  oder  $\{u, v\}$  zur Bezeichnung des ungeordneten Paares  $uv$  benutzt. Wir lassen zur Bezeichnungsvereinfachung die Klammern weg, es sei denn, dies führt zu unklarer Notation. Zum Beispiel bezeichnen wir die Kante zwischen Knoten 1 und Knoten 23 nicht mit 123, wir schreiben dann  $\{1, 23\}$ .)

Die Anzahl der Knoten eines Graphen heißt *Ordnung* des Graphen. Ein Graph heißt *endlich*, wenn  $V$  und  $E$  endliche Mengen sind, andernfalls heißt  $G$  *unendlich*. Wir werden uns nur mit endlichen Graphen beschäftigen und daher ab jetzt statt „endlicher Graph“ einfach „Graph“ schreiben. Wie werden versuchen, die natürliche Zahl  $n$  für die Knotenzahl und die natürliche Zahl  $m$  für die Kantenzahl eines Graphen zu reservieren. (Das gelingt wegen der geringen Anzahl der Buchstaben unseres Alphabets nicht immer.)

Gilt  $\Psi(e) = uv$  für eine Kante  $e \in E$ , dann heißen die Knoten  $u, v \in V$  *Endknoten* von  $e$ , und wir sagen, dass  $u$  und  $v$  mit  $e$  *inzidieren* oder *auf  $e$  liegen*, dass  $e$  die Knoten  $u$  und  $v$  *verbindet*, und dass  $u$  und  $v$  *Nachbarn* bzw. *adjazent* sind. Wir sagen auch, dass zwei Kanten *inzident* sind, wenn sie einen gemeinsamen Endknoten haben. Eine Kante  $e$  mit  $\Psi(e) = uu$  heißt *Schlinge* (oder *Schleife*); Kanten  $e, f$  mit  $\Psi(e) = uv = \Psi(f)$  heißen *parallel*, man sagt in diesem Falle auch, dass die Knoten  $u$  und  $v$  durch eine *Mehrfachkante* verbunden sind. Graphen, die weder Mehrfachkanten noch Schlingen enthalten, heißen *einfach*. Der einfache Graph, der zu jedem in  $G$  adjazenten Knotenpaar  $u, v$  mit  $u \neq v$  genau eine  $u$  und  $v$  verbindende Kante enthält, heißt der  $G$  *unterliegende einfache Graph*. Mit  $\Gamma(v)$  bezeichnen wir die Menge der Nachbarn eines Knotens  $v$ . Falls  $v$  in einer Schlinge enthalten ist, ist  $v$  natürlich mit sich selbst benachbart.  $\Gamma(W) := \bigcup_{v \in W} \Gamma(v)$  ist die Menge der Nachbarn von  $W \subseteq V$ . Ein Knoten ohne Nachbarn heißt *isoliert*.

Die Benutzung der Inzidenzfunktion  $\Psi$  führt zu einem relativ aufwendigen Formalismus. Wir wollen daher die Notation etwas vereinfachen. Dabei entstehen zwar im Falle von nicht-einfachen Graphen gelegentlich Mehrdeutigkeiten, die aber i. A. auf offensichtliche Weise interpretiert werden können. Statt  $\Psi(e) = uv$  schreiben wir von nun an einfach  $e = uv$  (oder äquivalent  $e = vu$ ) und meinen damit die Kante  $e$  mit den Endknoten  $u$  und  $v$ . Das ist korrekt, solange es nur eine Kante zwischen  $u$  und  $v$  gibt. Gibt es mehrere Kanten mit den Endknoten  $u$  und  $v$ , und sprechen wir von der Kante  $uv$ , so soll das heißen, dass wir einfach eine der parallelen Kanten auswählen. Von jetzt an vergessen wir also die Inzidenzfunktion  $\Psi$  und benutzen die Abkürzung  $G = (V, E)$ , um einen Graphen zu bezeichnen. Manchmal schreiben wir auch, wenn erhöhte Präzision erforderlich ist,  $E_G$  oder  $E(G)$  bzw.  $V_G$  oder  $V(G)$  zur Bezeichnung der Kanten- bzw. Knotenmenge eines Graphen  $G$ .

Zwei Graphen  $G = (V, E)$  und  $H = (W, F)$  heißen *isomorph*, wenn es eine bijektive Abbildung  $\varphi : V \rightarrow W$  gibt, so dass  $uv \in E$  genau dann gilt, wenn  $\varphi(u)\varphi(v) \in F$  gilt. Isomorphe Graphen sind also – bis auf die Benennung der Knoten und Kanten – identisch.

Eine Menge  $F$  von Kanten heißt *Schnitt*, wenn es eine Knotenmenge  $W \subseteq V$  gibt, so dass  $F = \delta(W) := \{uv \in E \mid u \in W, v \in V \setminus W\}$  gilt; manchmal wird  $\delta(W)$  der *durch  $W$  induzierte Schnitt* genannt. Statt  $\delta(\{v\})$  schreiben wir kurz  $\delta(v)$ . Ein Schnitt, der keinen anderen nicht-leeren Schnitt als echte Teilmenge enthält, heißt *Cokreis* (oder *minimaler Schnitt*). Wollen wir betonen, dass ein Schnitt  $\delta(W)$  bezüglich zweier Knoten  $s, t \in V$  die Eigenschaft  $s \in W$  und  $t \in V \setminus W$  hat, so sagen wir,  $\delta(W)$  ist ein  *$s$  und  $t$  trennender Schnitt* oder kurz ein  *$[s, t]$ -Schnitt*.

## 2.1 Graphen und Digraphen: Wichtige Definitionen und Bezeichnungen

Generell benutzen wir die eckigen Klammern  $[\cdot, \cdot]$ , um anzudeuten, dass die Reihenfolge der Objekte in der Klammer ohne Bedeutung ist. Z. B. ist ein  $[s, t]$ -Schnitt natürlich auch ein  $[t, s]$ -Schnitt, da ja  $\delta(W) = \delta(V \setminus W)$  gilt.

Wir haben oben Bezeichnungen wie  $\Gamma(v)$  oder  $\delta(W)$  eingeführt unter der stillschweigenden Voraussetzung, dass man weiß, in Bezug auf welchen Graphen diese Mengen definiert sind. Sollten mehrere Graphen involviert sein, so werden wir, wenn Zweideutigkeiten auftreten können, die Graphennamen als Indizes verwenden, also z. B.  $\Gamma_G(v)$  oder  $\delta_G(V)$  schreiben. Analog wird bei allen anderen Symbolen verfahren.

Der *Grad* (oder die *Valenz*) eines Knotens  $v$  (Bezeichnung:  $\deg(v)$ ) ist die Anzahl der Kanten, mit denen er inzidiert, wobei Schlingen doppelt gezählt werden. Hat ein Graph keine Schlingen, so ist der Grad von  $v$  gleich  $|\delta(v)|$ . Ein Graph heißt *k-regulär*, wenn jeder Knoten den Grad  $k$  hat, oder kurz *regulär*, wenn der Grad  $k$  nicht hervorgehoben werden soll.

Sind  $W$  eine Knotenmenge und  $F$  eine Kantenmenge in  $G = (V, E)$ , dann bezeichnen wir mit  $E(W)$  die Menge aller Kanten von  $G$  mit beiden Endknoten in  $W$  und mit  $V(F)$  die Menge aller Knoten, die Endknoten mindestens einer Kante aus  $F$  sind.

Sind  $G = (V, E)$  und  $H = (W, F)$  zwei Graphen, so heißt der Graph  $(V \cup W, E \cup F)$  die *Vereinigung* von  $G$  und  $H$ , und  $(V \cap W, E \cap F)$  heißt der *Durchschnitt* von  $G$  und  $H$ .  $G$  und  $H$  heißen *disjunkt*, falls  $V \cap W = \emptyset$ , *kantendisjunkt*, falls  $E \cap F = \emptyset$ . Wir sprechen von einer *disjunkten* bzw. *kantendisjunkten Vereinigung* von zwei Graphen, wenn sie disjunkt bzw. kantendisjunkt sind.

Sind  $G = (V, E)$  und  $H = (W, F)$  Graphen, so dass  $W \subseteq V$  und  $F \subseteq E$  gilt, so heißt  $H$  *Untergraph* (oder *Teilgraph*) von  $G$ . Falls  $W \subseteq V$ , so bezeichnet  $G - W$  den Graphen, den man durch *Entfernen* aller Knoten in  $W$  und aller Kanten mit mindestens einem Endknoten in  $W$  gewinnt.  $G[W] := G - (V \setminus W)$  heißt der *von  $W$  induzierte Untergraph* von  $G$ . Es gilt also  $G[W] = (W, E(W))$ . Für  $F \subseteq E$  ist  $G - F := (V, E \setminus F)$  der Graph, den man durch *Entfernen* der Kantenmenge  $F$  erhält. Statt  $G - \{f\}$  schreiben wir  $G - f$ , analog schreiben wir  $G - w$  statt  $G - \{w\}$  für  $w \in V$ . Ein Untergraph  $H = (W, F)$  von  $G = (V, E)$  heißt *aufspannend*, falls  $V = W$  gilt.

Ist  $G = (V, E)$  ein Graph und  $W \subseteq V$  eine Knotenmenge, so bezeichnen wir mit  $G \cdot W$  den Graphen, der durch *Kontraktion der Knotenmenge  $W$*  entsteht. Das heißt, die Knotenmenge von  $G \cdot W$  besteht aus den Knoten  $V \setminus W$  und einem neuen Knoten  $w$ , der die Knotenmenge  $W$  ersetzt. Die Kantenmenge von  $G \cdot W$  enthält alle Kanten von  $G$ , die mit keinem Knoten aus  $W$  inzidieren, und alle Kanten, die genau einen Endknoten in  $W$  haben, aber dieser Endknoten wird durch  $w$  ersetzt (also können viele parallele Kanten entstehen). Keine der Kanten von  $G$ , die in  $E(W)$  liegen, gehört zu  $G \cdot W$ . Falls  $e = uv \in E$  und falls  $G$  keine zu  $e$  parallele Kante enthält, dann ist der Graph, der durch *Kontraktion der Kante  $e$*  entsteht (Bezeichnung  $G \cdot e$ ), der Graph  $G \cdot \{u, v\}$ . Falls  $G$  zu  $e$  parallele Kanten enthält, so erhält man  $G \cdot e$  aus  $G \cdot \{u, v\}$  durch Addition von so vielen Schlingen, die den neuen Knoten  $w$  enthalten, wie es Kanten in  $G$  parallel zu  $e$  gibt. Der Graph  $G \cdot F$ , den man durch *Kontraktion einer Kantenmenge  $F \subseteq E$*  erhält, ist der Graph, der durch sukzessive Kontraktion (in beliebiger Reihenfolge) der Kanten aus  $F$  gewonnen wird. Ist  $e$  eine Schlinge von  $G$ , so sind  $G \cdot e$  und  $G - e$  identisch.

Ein einfacher Graph heißt *vollständig*, wenn jedes Paar seiner Knoten durch eine Kante

verbunden ist. Offenbar gibt es – bis auf Isomorphie – nur einen vollständigen Graphen mit  $n$  Knoten. Dieser wird mit  $K_n$  bezeichnet. Ein Graph  $G$ , dessen Knotenmenge  $V$  in zwei disjunkte nicht-leere Teilmengen  $V_1, V_2$  mit  $V_1 \cup V_2 = V$  zerlegt werden kann, so dass keine zwei Knoten in  $V_1$  und keine zwei Knoten in  $V_2$  benachbart sind, heißt *bipartit* (oder *paar*). Die Knotenmengen  $V_1, V_2$  nennt man eine *Bipartition* (oder *2-Färbung*) von  $G$ . Falls  $G$  zu je zwei Knoten  $u \in V_1$  und  $v \in V_2$  genau eine Kante  $uv$  enthält, so nennt man  $G$  *vollständig bipartit*. Den – bis auf Isomorphie eindeutig bestimmten – vollständig bipartiten Graphen mit  $|V_1| = m, |V_2| = n$  bezeichnen wir mit  $K_{m,n}$ .

Ist  $G$  ein Graph, dann ist das *Komplement* von  $G$ , bezeichnet mit  $\overline{G}$ , der einfache Graph, der dieselbe Knotenmenge wie  $G$  hat und bei dem zwei Knoten genau dann durch eine Kante verbunden sind, wenn sie in  $G$  nicht benachbart sind. Ist  $G$  einfach, so gilt  $\overline{\overline{G}} = G$ . Der *Kantengraph* (englisch: *line graph*)  $L(G)$  eines Graphen  $G$  ist der einfache Graph, dessen Knotenmenge die Kantenmenge von  $G$  ist und bei dem zwei Knoten genau dann adjazent sind, wenn die zugehörigen Kanten in  $G$  einen gemeinsamen Endknoten haben.

Eine *Clique* in einem Graphen  $G$  ist eine Knotenmenge  $Q$ , so dass je zwei Knoten aus  $Q$  in  $G$  benachbart sind. Eine *stabile Menge* in einem Graphen  $G$  ist eine Knotenmenge  $S$ , so dass je zwei Knoten aus  $S$  in  $G$  nicht benachbart sind. Für stabile Mengen werden auch die Begriffe *unabhängige Knotenmenge* oder *Coclique* verwendet. Eine Knotenmenge  $K$  in  $G$  heißt *Knotenüberdeckung* (oder *Überdeckung von Kanten durch Knoten*), wenn jede Kante aus  $G$  mit mindestens einem Knoten in  $K$  inzidiert. Die größte Kardinalität (= Anzahl der Elemente) einer stabilen Menge (bzw. Clique) in einem Graphen bezeichnet man mit  $\alpha(G)$  (bzw.  $\omega(G)$ ); die kleinste Kardinalität einer Knotenüberdeckung mit  $\tau(G)$ .

Eine Kantenmenge  $M$  in  $G$  heißt *Matching* (oder *Paarung* oder *Korrespondenz* oder *unabhängige Kantenmenge*), wenn  $M$  keine Schlingen enthält und je zwei Kanten in  $M$  keinen gemeinsamen Endknoten besitzen.  $M$  heißt *perfekt*, wenn jeder Knoten von  $G$  Endknoten einer Kante des Matchings  $M$  ist. Ein perfektes Matching wird auch *1-Faktor* genannt. Eine Kantenmenge  $F$  in  $G$  heißt *k-Faktor* (oder *perfektes k-Matching*), wenn jeder Knoten von  $G$  in genau  $k$  Kanten aus  $F$  enthalten ist. Eine *Kantenüberdeckung* (oder *Überdeckung von Knoten durch Kanten*) ist eine Kantenmenge, so dass jeder Knoten aus  $G$  mit mindestens einer Kante dieser Menge inzidiert. Die größte Kardinalität eines Matchings in  $G$  bezeichnet man mit  $\nu(G)$ , die kleinste Kardinalität einer Kantenüberdeckung mit  $\rho(G)$ .

Eine Zerlegung der Knotenmenge eines Graphen in stabile Mengen, die so genannten *Farbklassen*, heißt *Knotenfärbung*; d. h. die Knoten werden so gefärbt, dass je zwei benachbarte Knoten eine unterschiedliche Farbe haben. Eine Zerlegung der Kantenmenge in Matchings heißt *Kantenfärbung*; die Kanten werden also so gefärbt, dass je zwei inzidente Kanten verschieden gefärbt sind. Eine Zerlegung der Knotenmenge von  $G$  in Cliques heißt *Cliquenüberdeckung* von  $G$ . Die minimale Anzahl von stabilen Mengen (bzw. Cliques) in einer Knotenfärbung (bzw. Cliquenüberdeckung) bezeichnet man mit  $\chi(G)$  (bzw.  $\overline{\chi}(G)$ ), die minimale Anzahl von Matchings in einer Kantenfärbung mit  $\gamma(G)$ . Die Zahl  $\gamma(G)$  heißt *chromatischer Index* (oder *Kantenfärbungszahl*),  $\chi(G)$  *Färbungszahl* (oder *Knotenfärbungszahl* oder *chromatische Zahl*).

Ein Graph  $G = (V, E)$  kann in die Ebene gezeichnet werden, indem man jeden Knoten durch einen Punkt repräsentiert und jede Kante durch eine Kurve (oder Linie oder Streckenstück), die die beiden Punkte verbindet, die die Endknoten der Kante repräsentieren. Ein Graph heißt *planar* (oder plättbar), falls er in die Ebene gezeichnet werden kann, so dass sich keine zwei Kanten (d. h. die sie repräsentierenden Kurven) schneiden – außer möglicherweise in ihren Endknoten. Eine solche Darstellung eines planaren Graphen  $G$  in der Ebene nennt man auch *Einbettung* von  $G$  in die Ebene.

### 2.1.3 Digraphen

Die Kanten eines Graphen haben keine Orientierung. In vielen Anwendungen spielen aber Richtungen eine Rolle. Zur Modellierung solcher Probleme führen wir gerichtete Graphen ein. Ein *Digraph* (oder *gerichteter Graph*)  $D = (V, A)$  besteht aus einer (endlichen) nicht-leeren *Knotenmenge*  $V$  und einer (endlichen) Menge  $A$  von *Bögen* (oder *gerichteten Kanten*; englisch: *arc*). Ein *Bogen*  $a$  ist ein geordnetes Paar von Knoten, also  $a = (u, v)$ ,  $u$  ist der *Anfangs-* oder *Startknoten*,  $v$  der *End-* oder *Zielknoten* von  $a$ ;  $u$  heißt *Vorgänger* von  $v$ ,  $v$  *Nachfolger* von  $u$ ,  $a$  *inzidiert mit*  $u$  und  $v$ . (Um exakt zu sein, müssten wir hier ebenfalls eine Inzidenzfunktion  $\Psi = (t, h) : A \rightarrow V \times V$  einführen. Für einen Bogen  $a \in A$  ist dann  $t(a)$  der Anfangsknoten (englisch: *tail*) und  $h(a)$  der Endknoten (englisch: *head*) von  $a$ . Aus den bereits oben genannten Gründen wollen wir jedoch die Inzidenzfunktion nur in Ausnahmefällen benutzen.) Wie bei Graphen gibt es auch hier *parallele Bögen* und *Schlingen*. Die Bögen  $(u, v)$  und  $(v, u)$  heißen *antiparallel*.

In manchen Anwendungsfällen treten auch „Graphen“ auf, die sowohl gerichtete als auch ungerichtete Kanten enthalten. Wir nennen solche Objekte *gemischte Graphen* und bezeichnen einen gemischten Graphen mit  $G = (V, E, A)$ , wobei  $V$  die Knotenmenge,  $E$  die Kantenmenge und  $A$  die Bogenmenge von  $G$  bezeichnet.

Falls  $D = (V, A)$  ein Digraph ist und  $W \subseteq V, B \subseteq A$ , dann bezeichnen wir mit  $A(W)$  die Menge der Bögen, deren Anfangs- und Endknoten in  $W$  liegen, und mit  $V(B)$  die Menge der Knoten, die als Anfangs- oder Endknoten mindestens eines Bogens in  $B$  auftreten. Unterdigraphen, induzierte Unterdigraphen, aufspannende Unterdigraphen, Vereinigung und Durchschnitt von Digraphen, das Entfernen von Bogen- oder Knotenmengen und die Kontraktion von Bogen- oder Knotenmengen sind genau wie bei Graphen definiert.

Ist  $D = (V, A)$  ein Digraph, dann heißt der Graph  $G = (V, E)$ , der für jeden Bogen  $(i, j) \in A$  eine Kante  $ij$  enthält, der  $D$  *unterliegende Graph*. Analog werden der  $D$  *unterliegende einfache Graph* und der *unterliegende einfache Digraph* definiert. Wir sagen, dass ein Digraph eine „ungerichtete“ Eigenschaft hat, wenn der ihm unterliegende Graph diese Eigenschaft hat (z. B.,  $D$  ist bipartit oder planar, wenn der  $D$  unterliegende Graph  $G$  bipartit oder planar ist). Geben wir jeder Kante  $ij$  eines Graphen  $G$  eine Orientierung, d. h., ersetzen wir  $ij$  durch einen der Bögen  $(i, j)$  oder  $(j, i)$ , so nennen wir den so entstehenden Digraphen  $D$  *Orientierung* von  $G$ .

Ein einfacher Digraph heißt *vollständig*, wenn je zwei Knoten  $u \neq v$  durch die beiden Bögen  $(u, v), (v, u)$  verbunden sind. Ein *Turnier* ist ein Digraph, der für je zwei Knoten  $u \neq v$  genau einen der Bögen  $(u, v)$  oder  $(v, u)$  enthält. (Der einem Turnier unterliegende Graph ist also ein vollständiger Graph; jedes Turnier ist die Orientierung eines

vollständigen Graphen.)

Für  $W \subseteq V$  sei  $\delta^+(W) := \{(i, j) \in A \mid i \in W, j \notin W\}$ ,  $\delta^-(W) := \delta^+(V \setminus W)$  und  $\delta(W) := \delta^+(W) \cup \delta^-(W)$ . Die Bogenmenge  $\delta^+(W)$  (bzw.  $\delta^-(W)$ ) heißt *Schnitt*. Ist  $s \in W$  und  $t \notin W$ , so heißt  $\delta^+(W)$  auch  $(s, t)$ -Schnitt. (Achtung: in einem Digraphen ist ein  $(s, t)$ -Schnitt kein  $(t, s)$ -Schnitt!)

Statt  $\delta^+(\{v\})$ ,  $\delta^-(\{v\})$ ,  $\delta(\{v\})$  schreiben wir  $\delta^+(v)$ ,  $\delta^-(v)$ ,  $\delta(v)$ . Der *Außengrad* (*Innengrad*) von  $v$  ist die Anzahl der Bögen mit Anfangsknoten (Endknoten)  $v$ . Die Summe von Außengrad und Innengrad ist der *Grad* von  $v$ . Ein Schnitt  $\delta^+(W)$ ,  $\emptyset \neq W \neq V$ , heißt *gerichteter Schnitt*, falls  $\delta^-(W) = \emptyset$ , d. h. falls  $\delta(W) = \delta^+(W)$ . Ist  $r \in W$ , so sagen wir auch, dass  $\delta^+(W)$  ein *Schnitt mit Wurzel*  $r$  ist.

### 2.1.4 Ketten, Wege, Kreise, Bäume

Das größte Durcheinander in der graphentheoretischen Terminologie herrscht bei den Begriffen Kette, Weg, Kreis und bei den damit zusammenhängenden Namen. Wir haben uns für folgende Bezeichnungen entschieden.

In einem Graphen oder Digraphen heißt eine endliche Folge

$$W = (v_0, e_1, v_1, e_2, v_2, \dots, e_k, v_k), \quad k \geq 0,$$

die mit einem Knoten beginnt und endet und in der Knoten und Kanten (Bögen) alternierend auftreten, so dass jede Kante (jeder Bogen)  $e_i$  mit den beiden Knoten  $v_{i-1}$  und  $v_i$  inzidiert, eine *Kette*. Der Knoten  $v_0$  heißt *Anfangsknoten*,  $v_k$  *Endknoten* der Kette; die Knoten  $v_1, \dots, v_{k-1}$  heißen *innere Knoten*;  $W$  wird auch  $[v_0, v_k]$ -*Kette* genannt. Die Zahl  $k$  heißt *Länge* der Kette (= Anzahl der Kanten bzw. Bögen in  $W$ , wobei einige Kanten/Bögen mehrfach auftreten können und somit mehrfach gezählt werden). Abbildung 2.1(b) zeigt eine Kette der Länge 13 im Graphen  $G$  aus Abbildung 2.1(a). Aus einem solchen Bild kann man in der Regel nicht entnehmen, in welcher Reihenfolge die Kanten durchlaufen werden.

Falls (in einem Digraphen) alle Bögen  $e_i$  der Kette  $W$  der Form  $(v_{i-1}, v_i)$  (also gleichgerichtet) sind, so nennt man  $W$  *gerichtete Kette* bzw.  $(v_0, v_k)$ -*Kette*. Ist  $W = (v_0, e_1, v_1, \dots, e_k, v_k)$  eine Kette und sind  $i, j$  Indizes mit  $0 \leq i < j \leq k$ , dann heißt die Kette  $(v_i, e_{i+1}, v_{i+1}, \dots, e_j, v_j)$  das  $[v_i, v_j]$ -*Segment* (bzw.  $(v_i, v_j)$ -*Segment*, wenn  $W$  gerichtet ist) von  $W$ . Jede (gerichtete) Kante, die zwei Knoten der Kette  $W$  miteinander verbindet, die aber nicht Element von  $W$  ist, heißt *Diagonale* (oder *Sehne*) von  $W$ .

Gibt es in einem Graphen keine parallelen Kanten, so ist eine Kette  $W$  bereits durch die Folge  $(v_0, \dots, v_k)$  ihrer Knoten eindeutig festgelegt. Desgleichen ist in einem Digraphen ohne parallele Bögen eine gerichtete Kette durch die Knotenfolge  $(v_0, \dots, v_k)$  bestimmt. Zur Bezeichnungsvereinfachung werden wir daher häufig von der Kette  $(v_0, \dots, v_k)$  in einem Graphen bzw. der gerichteten Kette  $(v_0, \dots, v_k)$  in einem Digraphen sprechen, obgleich bei parallelen Kanten (Bögen) die benutzten Kanten (Bögen) hiermit nicht eindeutig festgelegt sind. Diese geringfügige Ungenauigkeit sollte aber keine Schwierigkeiten bereiten. Gelegentlich interessiert man sich mehr für die Kanten (Bögen) einer Kette, insbesondere wenn diese ein Weg oder ein Kreis (siehe unten) ist. In solchen Fällen ist es zweckmäßiger, eine Kette als Kantenfolge  $(e_1, e_2, \dots, e_k)$  zu betrachten. Ist  $C$  die Menge

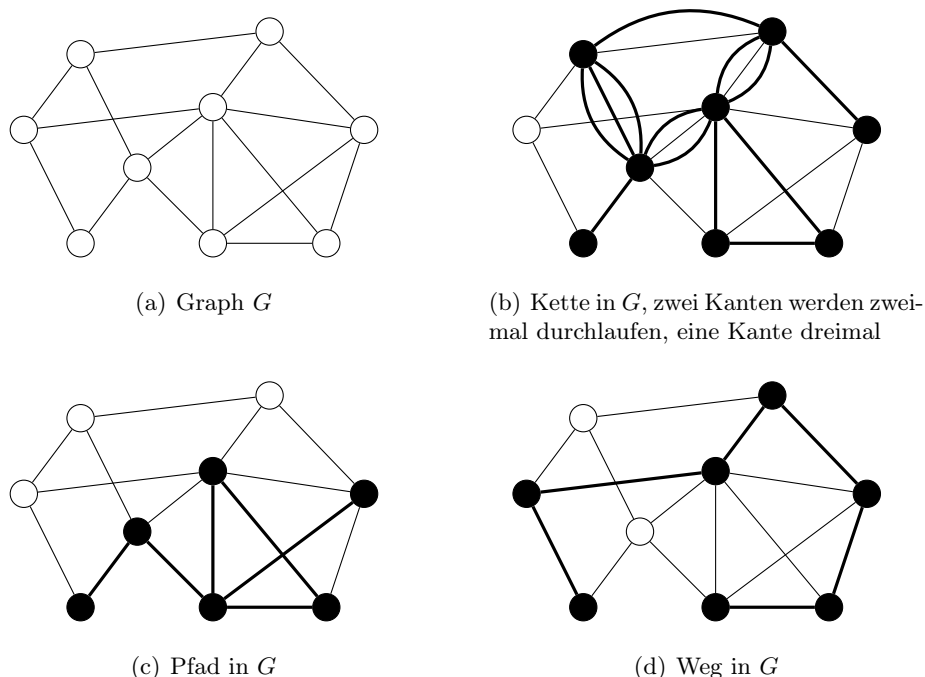


Abbildung 2.1: Kette, Pfad und Weg in einem Graphen  $G$ .

der Kanten (Bögen) eines Kreises oder eines Weges, so spricht man dann einfach vom Kreis oder Weg  $C$ , während  $V(C)$  die Menge der Knoten des Kreises oder Weges bezeichnet. Je nach behandeltem Themenkreis wird hier die am besten geeignete Notation benutzt.

Eine Kette, in der alle Knoten voneinander verschieden sind, heißt *Weg* (siehe Abbildung 2.1(d)). Eine Kette, in der alle Kanten oder Bögen verschieden sind, heißt *Pfad*. Ein Beispiel ist in Abb. 2.1(c) dargestellt. Ein Weg ist also ein Pfad, aber nicht jeder Pfad ist ein Weg. Ein Weg oder Pfad in einem Digraphen, der eine gerichtete Kette ist, heißt *gerichteter Weg* oder *gerichteter Pfad*. Wie bei Ketten sprechen wir von  $[u, v]$ -Wegen,  $(u, v)$ -Wegen etc.

Im Englischen heißt Kette *walk* oder *chain*. Im Deutschen benutzen z. B. Domschke (1997), Hässig (1979) und Berge and Ghouila-Houri (1969) ebenfalls das Wort Kette, dagegen schreiben Aigner (1984), Diestel (2006) und Wagner (1970) hierfür „Kantenzug“, während König (1936), Halin (1989) und Sachs (1970) „Kantenfolge“ benutzen; Ebert (1981) schließlich nennt unsere Ketten „ungerichtete Pfade“. Dieses Wirrwarr setzt sich bezüglich der Begriffe Pfad und Weg auf ähnliche Weise fort.

Eine Kette heißt *geschlossen*, falls ihre Länge nicht Null ist und falls ihr Anfangsknoten mit ihrem Endknoten übereinstimmt. Ein geschlossener (gerichteter) Pfad, bei dem der Anfangsknoten und alle inneren Knoten voneinander verschieden sind, heißt *Kreis* (*gerichteter Kreis*). Offensichtlich enthält jeder geschlossene Pfad einen Kreis, siehe Abb. 2.2.

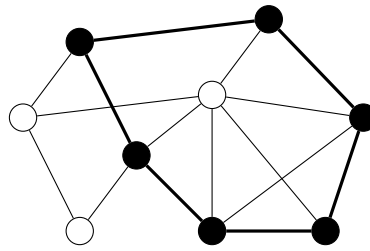


Abbildung 2.2: Ein Kreis in einem Graphen  $G$ .

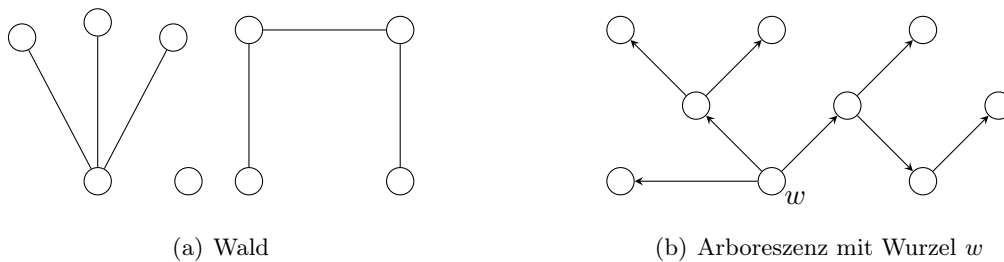


Abbildung 2.3: Ein Wald und eine Arboreszenz.

Ein (gerichteter) Pfad, der jede Kante (jeden Bogen) eines Graphen (Digraphen) genau einmal enthält, heißt (gerichteter) *Eulerpfad*. Ein geschlossener Eulerpfad heißt *Eulertour*. Ein *Eulergraph* (*Eulerdigraph*) ist ein Graph (Digraph), der eine (gerichtete) Eulertour enthält.

Ein (gerichteter) Kreis (Weg) der Länge  $|V|$  (bzw.  $|V| - 1$ ) heißt (gerichteter) *Hamiltonkreis* (*Hamiltonweg*). Ein Graph (Digraph), der einen (gerichteten) Hamiltonkreis enthält, heißt *hamiltonsch*. Manchmal sagen wir statt Hamiltonkreis einfach *Tour*.

Ein *Wald* ist ein Graph, der keinen Kreis enthält, siehe Abb. 2.3(a). Ein zusammenhängender Wald heißt *Baum*. Ein Baum in einem Graphen heißt *aufspannend*, wenn er alle Knoten des Graphen enthält. Ein *Branching*  $B$  ist ein Digraph, der ein Wald ist, so dass jeder Knoten aus  $B$  Zielknoten von höchstens einem Bogen von  $B$  ist. Ein zusammenhängendes Branching heißt *Arboreszenz*, siehe Abb. 2.3(b). Eine *aufspannende Arboreszenz* ist eine Arboreszenz in einem Digraphen  $D$ , die alle Knoten von  $D$  enthält. Eine Arboreszenz enthält einen besonderen Knoten, genannt *Wurzel*, von dem aus jeder andere Knoten auf genau einem gerichteten Weg erreicht werden kann. Arboreszenzen werden auch *Wurzelbäume* genannt. Ein Digraph, der keinen gerichteten Kreis enthält, heißt *azyklisch*.

Ein Graph heißt *zusammenhängend*, falls es zu jedem Paar von Knoten  $s, t$  einen  $[s, t]$ -Weg in  $G$  gibt. Ein Digraph  $D$  heißt *stark zusammenhängend*, falls es zu je zwei Knoten  $s, t$  von  $D$  sowohl einen gerichteten  $(s, t)$ -Weg als auch einen gerichteten  $(t, s)$ -Weg in  $D$  gibt. Die *Komponenten* (*starken Komponenten*) eines Graphen (Digraphen) sind die bezüglich Kanteninklusion (Bogeninklusion) maximalen zusammenhängenden Untergraphen von  $G$



## 2.1 Graphen und Digraphen: Wichtige Definitionen und Bezeichnungen

(maximalen stark zusammenhängenden Unterdigraphen von  $D$ ). Eine Komponente heißt *ungerade Komponente*, falls ihre Knotenzahl ungerade ist, andernfalls heißt sie *gerade Komponente*.

Sei  $G = (V, E)$  ein Graph. Eine Knotenmenge  $W \subseteq V$  heißt *trennend*, falls  $G - W$  unzusammenhängend ist. Für Graphen  $G = (V, E)$ , die keinen vollständigen Graphen der Ordnung  $|V|$  enthalten, setzen wir  $\kappa(G) := \min\{|W| \mid W \subseteq V \text{ ist trennend}\}$ . Die Zahl  $\kappa(G)$  heißt *Zusammenhangszahl* (oder *Knotenzusammenhangszahl*) von  $G$ . Für jeden Graphen  $G = (V, E)$ , der einen vollständigen Graphen der Ordnung  $|V|$  enthält, setzen wir  $\kappa(G) := |V| - 1$ . Falls  $\kappa(G) \geq k$ , so nennen wir  $G$  *k-fach knotenzusammenhängend* (kurz: *k-zusammenhängend*). Ein wichtiger Satz der Graphentheorie (Satz von Menger) besagt, dass  $G$  *k-fach zusammenhängend* genau dann ist, wenn jedes Paar  $s, t$ ,  $s \neq t$ , von Knoten durch mindestens  $k$  knotendisjunkte  $[s, t]$ -Wege miteinander verbunden ist. (Eine Menge von  $[s, t]$ -Wegen heißt *knotendisjunkt*, falls keine zwei Wege einen gemeinsamen inneren Knoten besitzen und die Menge der in den  $[s, t]$ -Wegen enthaltenen Kanten keine parallelen Kanten enthält.)

Eine Kantenmenge  $F$  eines Graphen  $G = (V, E)$  heißt *trennend*, falls  $G - F$  unzusammenhängend ist. Für Graphen  $G$ , die mehr als einen Knoten enthalten, setzen wir  $\lambda(G) := \min\{|F| \mid F \subseteq E \text{ trennend}\}$ . Die Zahl  $\lambda(G)$  heißt *Kantenzusammenhangszahl*. Für Graphen  $G$  mit nur einem Knoten setzen wir  $\lambda(G) = 0$ . Falls  $\lambda(G) \geq k$ , so nennen wir  $G$  *k-fach kantenzusammenhängend* (kurz: *k-kantenzusammenhängend*). Eine Version des Menger'schen Satzes besagt, dass  $G$  *k-kantenzusammenhängend* genau dann ist, wenn jedes Paar  $s, t$ ,  $s \neq t$ , von Knoten durch mindestens  $k$  kantendisjunkte  $[s, t]$ -Wege verbunden ist. Für Graphen  $G$  mit mindestens einem Knoten sind die Eigenschaften „ $G$  ist zusammenhängend“, „ $G$  ist 1-kantenzusammenhängend“ äquivalent.

Analoge Konzepte kann man in Digraphen definieren. Man benutzt hierbei den Zusatz „stark“, um den „gerichteten Zusammenhang“ zu kennzeichnen. Wir sagen, dass ein Digraph  $D = (V, A)$  *stark k-zusammenhängend* (bzw. *stark k-bogenzusammenhängend*) ist, falls jedes Knotenpaar  $s, t$ ,  $s \neq t$  durch mindestens  $k$  knotendisjunkte (bzw. bogendisjunkte)  $(s, t)$ -Wege verbunden ist.

Wir setzen  $\vec{\kappa}(D) := \max\{k \mid D \text{ stark } k\text{-zusammenhängend}\}$  und  $\vec{\lambda}(D) := \max\{k \mid D \text{ stark } k\text{-bogenzusammenhängend}\}$ ;  $\vec{\kappa}(D)$  heißt die *starke Zusammenhangszahl* von  $D$ ,  $\vec{\lambda}(D)$  die *starke Bogenzusammenhangszahl* von  $D$ .

Ein Kante  $e$  von  $G$  heißt *Brücke* (oder *Isthmus*), falls  $G - e$  mehr Komponenten als  $G$  hat. Ein Knoten  $v$  von  $G$  heißt *Trennungsknoten* (oder *Artikulation*), falls die Kantenmenge  $E$  von  $G$  so in zwei nicht-leere Teilmengen  $E_1$  und  $E_2$  zerlegt werden kann, dass  $V(E_1) \cap V(E_2) = \{v\}$  gilt. Ist  $G$  schlingenlos mit  $|V| \geq 2$ , dann ist  $v$  ein Trennungsknoten genau dann, wenn  $\{v\}$  eine trennende Knotenmenge ist, d. h. wenn  $G - v$  mehr Komponenten als  $G$  besitzt. Ein zusammenhängender Graph ohne Trennungsknoten wird *Block* genannt. Blöcke sind entweder isolierte Knoten, Schlingen oder Graphen mit 2 Knoten, die durch eine Kante oder mehrere parallele Kanten verbunden sind oder, falls  $|V| \geq 3$ , 2-zusammenhängende Graphen. Ein *Block eines Graphen* ist ein Untergraph, der ein Block und maximal bezüglich dieser Eigenschaft ist. Jeder Graph ist offenbar die Vereinigung seiner Blöcke.

Abbildung 2.4 zeigt einen 2-fach knotenzusammenhängenden Graphen (Block) sowie

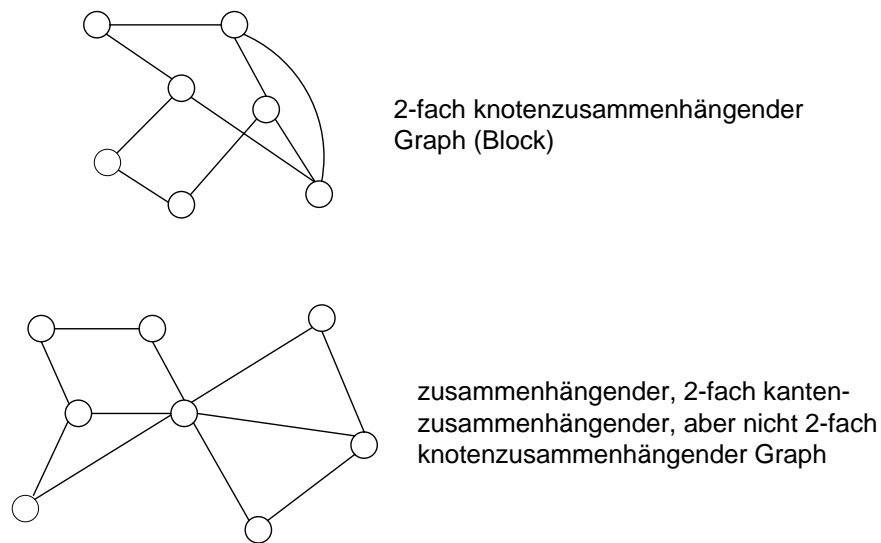


Abbildung 2.4: Ein Block und ein 2-fach kantenzusammenhängender Graph, der kein Block ist.

einen zusammenhängenden, 2-fach kantenzusammenhängenden, aber nicht 2-fach knotenzusammenhängenden Graphen.

## 2.2 Lineare Algebra

### 2.2.1 Grundmengen

Wir benutzen folgende Bezeichnungen:

$\mathbb{N} = \{1, 2, 3, \dots\}$  = Menge der *natürlichen Zahlen*,

$\mathbb{Z}$  = Menge der *ganzen Zahlen*,

$\mathbb{Q}$  = Menge der *rationalen Zahlen*,

$\mathbb{R}$  = Menge der *reellen Zahlen*.

Mit  $M_+$  bezeichnen wir die Menge der nichtnegativen Zahlen in  $M$  für  $M \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ . Wir betrachten  $\mathbb{Q}$  und  $\mathbb{R}$  als Körper mit der üblichen Addition und Multiplikation und der kanonischen Ordnung „ $\leq$ “. Desgleichen betrachten wir  $\mathbb{N}$  und  $\mathbb{Z}$  als mit den üblichen Rechenarten versehen. Wir werden uns fast immer in diesen Zahlenuniversen bewegen, da diese die für die Praxis relevanten sind. Manche Sätze gelten jedoch nur, wenn wir uns auf  $\mathbb{Q}$  oder  $\mathbb{R}$  beschränken. Um hier eine saubere Trennung zu haben, treffen wir die folgende Konvention. Wenn wir das Symbol

$\mathbb{K}$

benutzen, so heißt dies immer, dass  $\mathbb{K}$  einer der angeordneten Körper  $\mathbb{R}$  oder  $\mathbb{Q}$  ist. Sollte ein Satz nur für  $\mathbb{R}$  oder nur für  $\mathbb{Q}$  gelten, so treffen wir die jeweils notwendige Einschränkung.

Für diejenigen, die sich für möglichst allgemeine Sätze interessieren, sei an dieser Stelle folgendes vermerkt. Jeder der nachfolgend angegebenen Sätze bleibt ein wahrer Satz, wenn wir als Grundkörper  $\mathbb{K}$  einen archimedisch angeordneten Körper wählen. Ein bekannter Satz besagt, dass jeder archimedisch angeordnete Körper isomorph zu einem Unterkörper von  $\mathbb{R}$  ist, der  $\mathbb{Q}$  enthält. Unsere Sätze bleiben also richtig, wenn wir statt  $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$  irgendeinen archimedisch angeordneten Körper  $\mathbb{K}$  mit  $\mathbb{Q} \subseteq \mathbb{K} \subseteq \mathbb{R}$  wählen.

Wir können in fast allen Sätzen (insbesondere bei denen, die keine Ganzzahligkeitsbedingungen haben) auch die Voraussetzung „archimedisch“ fallen lassen, d. h. fast alle Sätze gelten auch für angeordnete Körper. Vieles, was wir im  $\mathbb{K}^n$  beweisen, ist auch in beliebigen metrischen Räumen oder Räumen mit anderen als euklidischen Skalarprodukten richtig. Diejenigen, die Spaß an derartigen Verallgemeinerungen haben, sind eingeladen, die entsprechenden Beweise in die allgemeinere Sprache zu übertragen.

In dieser Vorlesung interessieren wir uns für so allgemeine Strukturen nicht. Wir verbleiben in den (für die Praxis besonders wichtigen) Räumen, die über den reellen oder rationalen Zahlen errichtet werden. Also, nochmals, wenn immer wir das Symbol  $\mathbb{K}$  im weiteren gebrauchen, gilt

$$\mathbb{K} \in \{\mathbb{R}, \mathbb{Q}\},$$

und  $\mathbb{K}$  ist ein Körper mit den üblichen Rechenoperationen und Strukturen. Natürlich ist  $\mathbb{K}_+ = \{x \in \mathbb{K} \mid x \geq 0\}$ .

Die Teilmengenbeziehung zwischen zwei Mengen  $M$  und  $N$  bezeichnen wir wie üblich mit  $M \subseteq N$ . Gilt  $M \subseteq N$  und  $M \neq N$ , so schreiben wir  $M \subset N$ .  $M \setminus N$  bezeichnet die mengentheoretische Differenz  $\{x \in M \mid x \notin N\}$ .

## 2.2.2 Vektoren und Matrizen

Ist  $R$  eine beliebige Menge,  $n \in \mathbb{N}$ , so bezeichnen wir mit

$$R^n$$

die Menge aller  $n$ -Tupel oder Vektoren der Länge  $n$  mit Komponenten aus  $R$ . (Aus technischen Gründen ist es gelegentlich nützlich, Vektoren  $x \in R^0$ , also Vektoren ohne Komponenten, zu benutzen. Wenn wir dies tun, werden wir es explizit erwähnen, andernfalls setzen wir immer  $n \geq 1$  voraus.) Wir betrachten Vektoren  $x = (x_i)_{i=1, \dots, n} \in R^n$  immer als *Spaltenvektoren*, d. h.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Wollen wir mit Zeilenvektoren rechnen, so schreiben wir  $x^T$  (lies:  $x$  transponiert). Die Menge  $\mathbb{K}^n$  ist bekanntlich ein  $n$ -dimensionaler Vektorraum über  $\mathbb{K}$ . Mit

$$x^T y := \sum_{i=1}^n x_i y_i$$

## 2 Grundlagen und Notation

bezeichnen wir das *innere Produkt* zweier Vektoren  $x, y \in \mathbb{K}^n$ . Wir nennen  $x$  und  $y$  *senkrecht* (*orthogonal*), falls  $x^T y = 0$  gilt. Der  $\mathbb{K}^n$  ist für uns immer (wenn nichts anderes gesagt wird) mit der *euklidischen Norm*

$$\|x\| := \sqrt{x^T x}$$

ausgestattet.

Für Mengen  $S, T \subseteq \mathbb{K}^n$  und  $\alpha \in \mathbb{K}$  benutzen wir die folgenden Standardbezeichnungen für Mengenoperationen

$$\begin{aligned} S + T &:= \{x + y \in \mathbb{K}^n \mid x \in S, y \in T\}, \\ S - T &:= \{x - y \in \mathbb{K}^n \mid x \in S, y \in T\}, \\ \alpha S &:= \{\alpha x \in \mathbb{K}^n \mid x \in S\}. \end{aligned}$$

Einige Vektoren aus  $\mathbb{K}^n$  werden häufig auftreten, weswegen wir sie mit besonderen Symbolen bezeichnen. Mit  $e_j$  bezeichnen wir den Vektor aus  $\mathbb{K}^n$ , dessen  $j$ -te Komponente 1 und dessen übrige Komponenten 0 sind. Mit  $0$  bezeichnen wir den Nullvektor, mit  $\mathbb{1}$  den Vektor, dessen Komponenten alle 1 sind. Also

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad 0 = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix}.$$

Welche Dimension die Vektoren  $e_j$ ,  $0$ ,  $\mathbb{1}$  haben, ergibt sich jeweils aus dem Zusammenhang.

Für eine Menge  $R$  und  $m, n \in \mathbb{N}$  bezeichnet

$$R^{(m,n)} \text{ oder } R^{m \times n}$$

die Menge der  $(m, n)$ -*Matrizen* ( $m$  Zeilen,  $n$  Spalten) mit Einträgen aus  $R$ . (Aus technischen Gründen werden wir gelegentlich auch  $n = 0$  oder  $m = 0$  zulassen, d.h. wir werden auch Matrizen mit  $m$  Zeilen und ohne Spalten bzw.  $n$  Spalten und ohne Zeilen betrachten. Dieser Fall wird jedoch immer explizit erwähnt, somit ist in der Regel  $n \geq 1$  und  $m \geq 1$  vorausgesetzt.) Ist  $A \in R^{(m,n)}$ , so schreiben wir

$$A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$$

und meinen damit, dass  $A$  die folgende Form hat

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

Wenn nichts anderes gesagt wird, hat  $A$  die Zeilenindexmenge  $M = \{1, \dots, m\}$  und die Spaltenindexmenge  $N = \{1, \dots, n\}$ . Die  $j$ -te Spalte von  $A$  ist ein  $m$ -Vektor, den wir mit  $A_{.j}$  bezeichnen,

$$A_{.j} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Die  $i$ -te Zeile von  $A$  ist ein Zeilenvektor der Länge  $n$ , den wir mit  $A_{i.}$  bezeichnen, d. h.

$$A_{i.} = (a_{i1}, a_{i2}, \dots, a_{in}).$$

Wir werden in dieser Vorlesung sehr häufig Untermatrizen von Matrizen konstruieren, sie umsortieren und mit ihnen rechnen müssen. Um dies ohne Zweideutigkeiten durchführen zu können, führen wir zusätzlich eine etwas exaktere als die oben eingeführte Standardbezeichnungsweise ein.

Wir werden – wann immer es aus technischen Gründen notwendig erscheint – die Zeilen- und Spaltenindexmengen einer  $(m, n)$ -Matrix  $A$  nicht als Mengen sondern als Vektoren auffassen. Wir sprechen dann vom

$$\begin{aligned} &\text{vollen Zeilenindexvektor } M = (1, 2, \dots, m), \\ &\text{vollen Spaltenindexvektor } N = (1, 2, \dots, n) \end{aligned}$$

von  $A$ . Ein *Zeilenindexvektor* von  $A$  ist ein Vektor mit höchstens  $m$  Komponenten, der aus  $M$  durch Weglassen einiger Komponenten von  $M$  und Permutation der übrigen Komponenten entsteht. Analog entsteht ein *Spaltenindexvektor* durch Weglassen von Komponenten von  $N$  und Permutation der übrigen Komponenten. Sind also

$$\begin{aligned} I &= (i_1, i_2, \dots, i_p) && \text{ein Zeilenindexvektor von } A \text{ und} \\ J &= (j_1, j_2, \dots, j_q) && \text{ein Spaltenindexvektor von } A, \end{aligned}$$

so gilt immer  $i_s, i_t \in \{1, \dots, m\}$  und  $i_s \neq i_t$  für  $1 \leq s < t \leq p$ , und analog gilt  $j_s, j_t \in \{1, \dots, n\}$  und  $j_s \neq j_t$  für  $1 \leq s < t \leq q$ . Wir setzen

$$A_{IJ} := \begin{pmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \dots & a_{i_1 j_q} \\ a_{i_2 j_1} & a_{i_2 j_2} & \dots & a_{i_2 j_q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_p j_1} & a_{i_p j_2} & \dots & a_{i_p j_q} \end{pmatrix}$$

und nennen  $A_{IJ}$  *Untermatrix von  $A$* .  $A_{IJ}$  ist also eine  $(p, q)$ -Matrix, die aus  $A$  dadurch entsteht, dass man die Zeilen, die zu Indizes gehören, die nicht in  $I$  enthalten sind, und die Spalten, die zu Indizes gehören, die nicht in  $J$  enthalten sind, streicht und dann die so entstehende Matrix umsortiert.

Ist  $I = (i)$  und  $J = (j)$ , so erhalten wir zwei Darstellungsweisen für Zeilen bzw. Spalten von  $A$ :

$$\begin{aligned} A_{IN} &= A_{i.}, \\ A_{MJ} &= A_{.j}. \end{aligned}$$

## 2 Grundlagen und Notation

Aus Gründen der Notationsvereinfachung werden wir auch die folgende (etwas unsaubere) Schreibweise benutzen. Ist  $M$  der volle Zeilenindexvektor von  $A$  und  $I$  ein Zeilenindexvektor, so schreiben wir auch

$$I \subseteq M,$$

obwohl  $I$  und  $M$  keine Mengen sind, und für  $i \in \{1, \dots, m\}$  benutzen wir

$$i \in I \quad \text{bzw.} \quad i \notin I,$$

um festzustellen, dass  $i$  als Komponente von  $I$  auftritt bzw. nicht auftritt. Analog verfahren wir bezüglich der Spaltenindizes.

Gelegentlich spielt die tatsächliche Anordnung der Zeilen und Spalten keine Rolle. Wenn also z. B.  $I \subseteq \{1, \dots, n\}$  und  $J \subseteq \{1, \dots, m\}$  gilt, dann werden wir auch einfach schreiben

$$A_{IJ},$$

obwohl diese Matrix dann nur bis auf Zeilen- und Spaltenpermutationen definiert ist. Wir werden versuchen, diese Bezeichnungen immer so zu verwenden, dass keine Zweideutigkeiten auftreten. Deshalb treffen wir ab jetzt die Verabredung, dass wir – falls wir Mengen (und nicht Indexvektoren)  $I$  und  $J$  benutzen – die Elemente von  $I = \{i_1, \dots, i_p\}$  und von  $J = \{j_1, \dots, j_q\}$  kanonisch anordnen, d. h. die Indizierung der Elemente von  $I$  und  $J$  sei so gewählt, dass  $i_1 < i_2 < \dots < i_p$  und  $j_1 < j_2 < \dots < j_q$  gilt. Bei dieser Verabredung ist dann  $A_{IJ}$  die Matrix die aus  $A$  durch Streichen der Zeilen  $i$ ,  $i \notin I$ , und der Spalten  $j$ ,  $j \notin J$ , entsteht.

Ist  $I \subseteq \{1, \dots, m\}$  und  $J \subseteq \{1, \dots, n\}$  oder sind  $I$  und  $J$  Zeilen- bzw. Spaltenindexvektoren, dann schreiben wir auch

$$\begin{aligned} A_I & \quad \text{statt } A_{IN}, \\ A_{.J} & \quad \text{statt } A_{MJ}. \end{aligned}$$

$A_I$  entsteht also aus  $A$  durch Streichen der Zeilen  $i$ ,  $i \notin I$ ,  $A_{.J}$  durch Streichen der Spalten  $j$ ,  $j \notin J$ .

Sind  $A, B \in \mathbb{K}^{(m,n)}$ ,  $C \in \mathbb{K}^{(n,s)}$ ,  $\alpha \in \mathbb{K}$ , so sind

$$\begin{aligned} \text{die Summe} & \quad A + B, \\ \text{das Produkt} & \quad \alpha A, \\ \text{das Matrixprodukt} & \quad AC \end{aligned}$$

wie in der linearen Algebra üblich definiert.

Für einige häufig auftretende Matrizen haben wir spezielle Symbole reserviert. Mit  $0$  bezeichnen wir die Nullmatrix (alle Matrixelemente sind Null), wobei sich die Dimension der Nullmatrix jeweils aus dem Zusammenhang ergibt. (Das Symbol  $0$  kann also sowohl eine Zahl, einen Vektor als auch eine Matrix bezeichnen). Mit  $I$  bezeichnen wir die Einheitsmatrix. Diese Matrix ist quadratisch, die Hauptdiagonalelemente von  $I$  sind Eins, alle übrigen Null. Wollen wir die Dimension von  $I$  betonen, so schreiben wir auch  $I_n$  und meinen damit die  $(n, n)$ -Einheitsmatrix. Diejenige  $(m, n)$ -Matrix, bei der alle Elemente

Eins sind, bezeichnen wir mit  $E$ . Wir schreiben auch  $E_{m,n}$  bzw.  $E_n$ , um die Dimension zu spezifizieren ( $E_n$  ist eine  $(n,n)$ -Matrix). Ist  $x$  ein  $n$ -Vektor, so bezeichnet  $\text{diag}(x)$  diejenige  $(n,n)$ -Matrix  $A = (a_{ij})$  mit  $a_{ii} = x_i$  ( $i = 1, \dots, n$ ) und  $a_{ij} = 0$  ( $i \neq j$ ).

Wir halten an dieser Stelle noch einmal Folgendes fest: Wenn wir von einer Matrix  $A$  sprechen, ohne anzugeben, welche Dimension sie hat und aus welchem Bereich sie ist, dann nehmen wir implizit an, dass  $A \in \mathbb{K}^{(m,n)}$  gilt. Analog gilt immer  $x \in \mathbb{K}^n$ , wenn sich nicht aus dem Zusammenhang anderes ergibt.

### 2.2.3 Kombinationen von Vektoren, Hüllen, Unabhängigkeit

Ein Vektor  $x \in \mathbb{K}^n$  heißt *Linearkombination* der Vektoren  $x_1, \dots, x_k \in \mathbb{K}^n$ , falls es einen Vektor  $\lambda = (\lambda_1, \dots, \lambda_k)^T \in \mathbb{K}^k$  gibt mit

$$x = \sum_{i=1}^k \lambda_i x_i.$$

Gilt zusätzlich

$$\left\{ \begin{array}{l} \lambda \geq 0 \\ \lambda^T \mathbf{1} = 1 \\ \lambda \geq 0 \text{ und } \lambda^T \mathbf{1} = 1 \end{array} \right\} \text{ so heißt } x \left\{ \begin{array}{l} \textit{konische} \\ \textit{affine} \\ \textit{konvexe} \end{array} \right\} \textit{Kombination}$$

der Vektoren  $x_1, \dots, x_k$ . Diese Kombinationen heißen *echt*, falls weder  $\lambda = 0$  noch  $\lambda = e_j$  für ein  $j \in \{1, \dots, k\}$  gilt.

Für eine nichtleere Teilmenge  $S \subseteq \mathbb{K}^n$  heißt

$$\left\{ \begin{array}{l} \text{lin}(S) \\ \text{cone}(S) \\ \text{aff}(S) \\ \text{conv}(S) \end{array} \right\} \text{ die } \left\{ \begin{array}{l} \textit{lineare} \\ \textit{konische} \\ \textit{affine} \\ \textit{konvexe} \end{array} \right\} \textit{Hülle von } S,$$

d. h. die Menge aller Vektoren, die als lineare (konische, affine oder konvexe) Kombination von endlich vielen Vektoren aus  $S$  dargestellt werden können. Wir setzen außerdem

$$\begin{aligned} \text{lin}(\emptyset) &:= \text{cone}(\emptyset) := \{0\}, \\ \text{aff}(\emptyset) &:= \text{conv}(\emptyset) := \emptyset. \end{aligned}$$

Ist  $A$  eine  $(m,n)$ -Matrix, so schreiben wir auch

$$\text{lin}(A), \text{cone}(A), \text{aff}(A), \text{conv}(A)$$

und meinen damit die lineare, konische, affine bzw. konvexe Hülle der Spaltenvektoren  $A_{.1}, A_{.2}, \dots, A_{.n}$  von  $A$ . Eine Teilmenge  $S \subseteq \mathbb{K}^n$  heißt

$$\left\{ \begin{array}{l} \textit{linearer Raum} \\ \textit{Kegel} \\ \textit{affiner Raum} \\ \textit{konvexe Menge} \end{array} \right\} \text{ falls } \left\{ \begin{array}{l} S = \text{lin}(S) \\ S = \text{cone}(S) \\ S = \text{aff}(S) \\ S = \text{conv}(S) \end{array} \right\}.$$

Die hier benutzten Begriffe sind üblicher Standard, wobei für „linearen Raum“ in der linearen Algebra in der Regel *Vektorraum* oder *Untervektorraum* benutzt wird. Der Begriff „Kegel“ wird jedoch – u. a. in verschiedenen Zweigen der Geometrie – allgemeiner verwendet. „Unser Kegel“ ist in der Geometrie ein „abgeschlossener konvexer Kegel“.

Eine nichtleere endliche Teilmenge  $S \subseteq \mathbb{K}^n$  heißt *linear* (bzw. *affin*) *unabhängig*, falls kein Element von  $S$  als echte Linearkombination (bzw. Affinkombination) von Elementen von  $S$  dargestellt werden kann. Die leere Menge ist affin, jedoch nicht linear unabhängig. Jede Menge  $S \subseteq \mathbb{K}^n$ , die nicht linear bzw. affin unabhängig ist, heißt *linear* bzw. *affin abhängig*. Aus der linearen Algebra wissen wir, dass eine linear (bzw. affin) unabhängige Teilmenge des  $\mathbb{K}^n$  höchstens  $n$  (bzw.  $n+1$ ) Elemente enthält. Für eine Teilmenge  $S \subseteq \mathbb{K}^n$  heißt die Kardinalität einer größten linear (bzw. affin) unabhängigen Teilmenge von  $S$  der *Rang* (bzw. *affine Rang*) von  $S$ . Wir schreiben dafür  $\text{rang}(S)$  bzw.  $\text{arang}(S)$ . Die *Dimension* einer Teilmenge  $S \subseteq \mathbb{K}^n$ , Bezeichnung:  $\dim(S)$ , ist die Kardinalität einer größten affin unabhängigen Teilmenge von  $S$  minus 1, d. h.  $\dim(S) = \text{arang}(S) - 1$ .

Der *Rang einer Matrix*  $A$ , bezeichnet mit  $\text{rang}(A)$ , ist der Rang ihrer Spaltenvektoren. Aus der linearen Algebra wissen wir, dass  $\text{rang}(A)$  mit dem Rang der Zeilenvektoren von  $A$  übereinstimmt. Gilt für eine  $(m, n)$ -Matrix  $A$   $\text{rang}(A) = \min\{m, n\}$ , so sagen wir, dass  $A$  *vollen Rang* hat. Eine  $(n, n)$ -Matrix mit vollem Rang ist *regulär*, d. h. sie besitzt eine (eindeutig bestimmte) inverse Matrix (geschrieben  $A^{-1}$ ) mit der Eigenschaft  $AA^{-1} = I$ .

## 2.3 Polyeder und lineare Programme

Ein wichtiger Aspekt der Vorlesung ist die Behandlung von Problemen der linearen Programmierung bzw. die Modellierung von kombinatorischen Optimierungsproblemen als (ganzzahlige) lineare Programme. In diesem Abschnitt stellen wir einige grundlegende Begriffe der zugehörigen Theorie bereit, stellen dar, in welchen Formen lineare Programme vorkommen, und wie man sie ineinander transformiert. Schließlich beweisen wir als Einführung den schwachen Dualitätssatz.

### (2.1) Definition.

- a) Eine Teilmenge  $G \subseteq \mathbb{K}^n$  heißt *Hyperebene*, falls es einen Vektor  $a \in \mathbb{K}^n \setminus \{0\}$  und  $\alpha \in \mathbb{K}$  gibt mit

$$G = \{x \in \mathbb{K}^n \mid a^T x = \alpha\}.$$

Der Vektor  $a$  heißt *Normalenvektor* zu  $G$ .

- b) Eine Teilmenge  $H \subseteq \mathbb{K}^n$  heißt *Halbraum*, falls es einen Vektor  $a \in \mathbb{K}^n \setminus \{0\}$  und  $\alpha \in \mathbb{K}$  gibt mit

$$H = \{x \in \mathbb{K}^n \mid a^T x \leq \alpha\}.$$

Wir nennen  $a$  den *Normalenvektor* zu  $H$ . Die Hyperebene  $G = \{x \in \mathbb{K}^n \mid a^T x = \alpha\}$  heißt die zum Halbraum  $H$  gehörende *Hyperebene* oder die  $H$  berandende Hyperebene, und  $H$  heißt der zu  $G$  gehörende *Halbraum*.



- c) Eine Teilmenge  $P \subseteq \mathbb{K}^n$  heißt Polyeder, falls es ein  $m \in \mathbb{Z}_+$ , eine Matrix  $A \in \mathbb{K}^{(m,n)}$  und einen Vektor  $b \in \mathbb{K}^m$  gibt mit

$$P = \{x \in \mathbb{K}^n \mid Ax \leq b\}.$$

Um zu betonen, dass  $P$  durch  $A$  und  $b$  definiert ist, schreiben wir auch

$$P = P(A, b) := \{x \in \mathbb{K}^n \mid Ax \leq b\}.$$

- d) Ein Polyeder  $P$  heißt Polytop, wenn es beschränkt ist, d. h. wenn es ein  $B \in \mathbb{K}$ ,  $B > 0$  gibt mit  $P \subseteq \{x \in \mathbb{K}^n \mid \|x\| \leq B\}$ . △

Polyeder können wir natürlich auch in folgender Form schreiben

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

Halbräume sind offensichtlich Polyeder. Aber auch die leere Menge ist ein Polyeder, denn  $\emptyset = \{x \mid 0^T x \leq -1\}$ , und der gesamte Raum ist ein Polyeder, denn  $\mathbb{K}^n = \{x \mid 0^T x \leq 0\}$ . Sind alle Zeilenvektoren  $A_i$  von  $A$  vom Nullvektor verschieden, so sind die bei der obigen Durchschnittsbildung beteiligten Mengen Halbräume. Ist ein Zeilenvektor von  $A$  der Nullvektor, sagen wir  $A_1 = 0^T$ , so ist  $\{x \in \mathbb{K}^n \mid A_1 \cdot x \leq b_1\}$  entweder leer (falls  $b_1 < 0$ ) oder der gesamte Raum  $\mathbb{K}^n$  (falls  $b_1 \geq 0$ ). Das heißt, entweder ist  $P(A, b)$  leer oder die Mengen  $\{x \mid A_i \cdot x \leq b_i\}$  mit  $A_i = 0$  können bei der obigen Durchschnittsbildung weggelassen werden. Daraus folgt:

Jedes Polyeder  $P \neq \mathbb{K}^n$  ist Durchschnitt von endlich vielen Halbräumen.

Gilt  $P = P(A, b)$ , so nennen wir das Ungleichungssystem  $Ax \leq b$  ein  $P$  definierendes System (von linearen Ungleichungen). Sind  $\alpha > 0$  und  $1 \leq i < j \leq m$ , so gilt offensichtlich

$$P(A, b) = P(A, b) \cap \{x \mid \alpha A_i \cdot x \leq \alpha b_i\} \cap \{x \mid (A_i + A_j) \cdot x \leq b_i + b_j\}.$$

Daraus folgt, dass  $A$  und  $b$  zwar  $P = P(A, b)$  eindeutig bestimmen, dass aber  $P$  unendlich viele Darstellungen der Form  $P(D, d)$  hat.

**(2.2) Beispiel.** Wir betrachten das Ungleichungssystem

$$2x_1 \leq 5 \tag{1}$$

$$-2x_2 \leq 1 \tag{2}$$

$$-x_1 - x_2 \leq -1 \tag{3}$$

$$2x_1 + 9x_2 \leq 23 \tag{4}$$

$$6x_1 - 2x_2 \leq 13 \tag{5}$$

## 2 Grundlagen und Notation

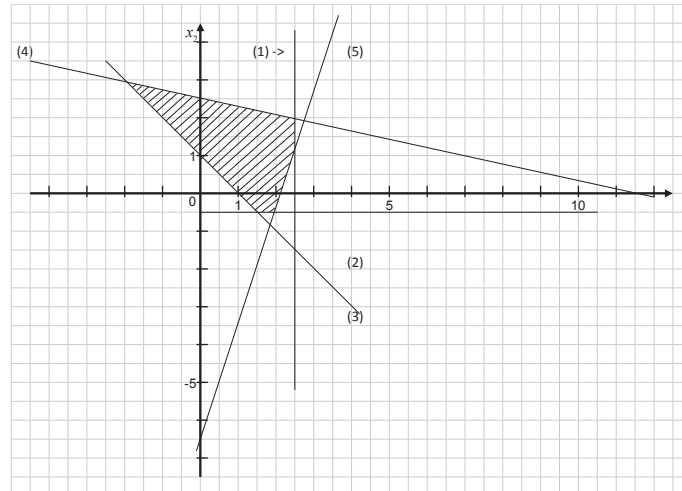


Abbildung 2.5: Darstellung des Polyeders aus Beispiel (2.2).

Hieraus erhalten wir die folgende Matrix  $A$  und den Vektor  $b$ :

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

Das Polyeder  $P = P(A, b)$  ist die Lösungsmenge des obigen Ungleichungssystems (1)–(5) und ist in Abbildung 2.5 graphisch dargestellt.  $\triangle$

Die Mengen zulässiger Lösungen linearer Programme treten nicht immer in der Form  $Ax \leq b$  auf. Häufig gibt es auch Gleichungen und vorzeichenbeschränkte Variable. Vorzeichenbeschränkungen sind natürlich auch lineare Ungleichungssysteme, und ein Gleichungssystem  $Dx = d$  kann in der Form von zwei Ungleichungssystemen  $Dx \leq d$  und  $-Dx \leq -d$  geschrieben werden. Allgemeiner gilt:

**(2.3) Bemerkung.** Die Lösungsmenge des Systems

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

ist ein Polyeder.  $\triangle$

**Beweis.** Setze  $n := p + q$  und

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

Dann ist  $P(A, b)$  die Lösungsmenge des vorgegebenen Gleichungs- und Ungleichungssystems.  $\square$

Ein spezieller Polyedertyp wird uns häufig begegnen, weswegen wir für ihn eine besondere Bezeichnung wählen wollen. Für  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$  setzen wir

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

Nicht alle Polyeder können in der Form  $P^=(A, b)$  dargestellt werden, z. B. nicht  $P = \{x \in \mathbb{K} \mid x \leq 1\}$ .

Wir werden später viele Sätze über Polyeder  $P$  beweisen, deren Aussagen *darstellungsabhängig* sind, d. h. die Art und Weise, wie  $P$  gegeben ist, geht explizit in die Satzaussage ein. So werden sich z. B. die Charakterisierungen gewisser Polyedereigenschaften von  $P(A, b)$  (zumindest formal) von den entsprechenden Charakterisierungen von  $P^=(A, b)$  unterscheiden. Darstellungsabhängige Sätze wollen wir jedoch nur einmal beweisen (normalerweise für Darstellungen, bei denen die Resultate besonders einprägsam oder einfach sind), deshalb werden wir uns nun Transformationsregeln überlegen, die angeben, wie man von einer Darstellungsweise zu einer anderen und wieder zurück kommt.

#### (2.4) Transformationen.

Regel I: *Einführung von Schlupfvariablen*

Gegeben seien  $a \in \mathbb{K}^n$ ,  $\alpha \in \mathbb{K}$ . Wir schreiben die Ungleichung

$$a^T x \leq \alpha \tag{2.5}$$

in der Form einer Gleichung und einer Vorzeichenbeschränkung

$$a^T x + y = \alpha, \quad y \geq 0. \tag{2.6}$$

$y$  ist eine neue Variable, genannt *Schlupfvariable*. Es gilt:

$$x \text{ erfüllt (2.5)} \implies \begin{pmatrix} x \\ y \end{pmatrix} \text{ erfüllt (2.6) für } y = \alpha - a^T x,$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \text{ erfüllt (2.6)} \implies x \text{ erfüllt (2.6)}.$$

Allgemein: Ein Ungleichungssystem  $Ax \leq b$  kann durch Einführung eines Schlupfvariablenvektors  $y$  transformiert werden in ein Gleichungssystem mit Vorzeichenbedingung  $Ax + y = b$ ,  $y \geq 0$ . Zwischen den Lösungsmengen dieser beiden Systeme bestehen die oben angegebenen Beziehungen.  $P(A, b)$  und  $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{K}^{n+m} \mid Ax + Iy = b, y \geq 0 \right\}$  sind jedoch zwei durchaus verschiedene Polyeder in verschiedenen Vektorräumen.

## 2 Grundlagen und Notation

Regel II: *Einführung von vorzeichenbeschränkten Variablen*

Ist  $x$  eine (eindimensionale) nicht vorzeichenbeschränkte Variable, so können wir zwei vorzeichenbeschränkte Variablen  $x^+$  und  $x^-$  einführen, um  $x$  darzustellen. Wir setzen

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

Mit den Regeln I und II aus (2.4) können wir z. B. jedem Polyeder  $P(A, b) \subseteq \mathbb{K}^n$  ein Polyeder  $P^=(D, d) \subseteq \mathbb{K}^{2n+m}$  wie folgt zuordnen. Wir setzen

$$D := (A, -A, I_m), \quad d := b,$$

d. h. es gilt

$$P^=(D, d) = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{K}^{2n+m} \mid Ax - Ay + z = b, x, y, z \geq 0 \right\}.$$

Es ist üblich, die oben definierten Polyeder  $P(A, b)$  und  $P^=(D, d)$  *äquivalent* zu nennen. Hierbei hat „Äquivalenz“ folgende Bedeutung. Für  $x \in \mathbb{K}^n$  sei

$$\begin{aligned} x^+ &:= (x_1^+, \dots, x_n^+)^T \quad \text{mit} \quad x_i^+ = \max\{0, x_i\}, \\ x^- &:= (x_1^-, \dots, x_n^-)^T \quad \text{mit} \quad x_i^- = \max\{0, -x_i\}. \end{aligned}$$

Dann gilt:

$$\begin{aligned} x \in P(A, b) &\implies \begin{pmatrix} x^+ \\ x^- \\ z \end{pmatrix} \in P^=(D, d) \quad \text{für} \quad z = b - Ax \\ \begin{pmatrix} u \\ v \\ w \end{pmatrix} \in P^=(D, d) &\implies x := u - v \in P(A, b). \end{aligned}$$

In der folgenden Tabelle 2.1 haben wir alle möglichen Transformationen aufgelistet. Sie soll als „Nachschlagewerk“ dienen.

Besonders einfache Polyeder, auf die sich jedoch fast alle Aussagen der Polyedertheorie zurückführen lassen, sind polyedrische Kegel.

**(2.7) Definition.** Ein Kegel  $C \subseteq \mathbb{K}^n$  heißt *polyedrisch genau* dann, wenn  $C$  ein Polyeder ist. △

**(2.8) Bemerkung.** Ein Kegel  $C \subseteq \mathbb{K}^n$  ist genau dann polyedrisch, wenn es eine Matrix  $A \in \mathbb{K}^{(m,n)}$  gibt mit

$$C = P(A, 0). \quad \triangle$$

<p>Transformation</p> <p>nach</p> <p>von</p>	$By \leq d$	$By = d$ $y \geq 0$	$By \leq d$ $y \geq 0$
$Ax = b$ <p>hierbei ist</p> $x_i^+ = \max\{0, x_i\}$ $x_i^- = \max\{0, -x_i\}$	$\begin{pmatrix} A \\ -A \end{pmatrix} y \leq \begin{pmatrix} b \\ -b \end{pmatrix}$ $y = x$	$(A, -A)y = b$ $y \geq 0$ $y = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}$	$\begin{pmatrix} A & -A \\ -A & A \end{pmatrix} y \leq \begin{pmatrix} b \\ -b \end{pmatrix}$ $y \geq 0$ $y = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}$
$Ax \leq b$	$Ay \leq b$ $y = x$	$(A, -A, I)y = b$ $y \geq 0$ $y = \begin{pmatrix} x^+ \\ x^- \\ b - Ax \end{pmatrix}$	$(A, -A)y \leq b$ $y \geq 0$ $y = \begin{pmatrix} x^+ \\ x^- \end{pmatrix}$
$Ax = b$ $x \geq 0$	$\begin{pmatrix} A \\ -A \\ -I \end{pmatrix} y \leq \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}$ $y = x$	$Ay = b$ $y \geq 0$ $y = x$	$\begin{pmatrix} A \\ -A \end{pmatrix} y \leq \begin{pmatrix} b \\ -b \end{pmatrix}$ $y \geq 0$ $y = x$
$Ax \leq b$ $x \geq 0$	$\begin{pmatrix} A \\ -I \end{pmatrix} y \leq \begin{pmatrix} b \\ 0 \end{pmatrix}$ $y = x$	$(A, I)y = b$ $y \geq 0$ $y = \begin{pmatrix} x \\ b - Ax \end{pmatrix}$	$Ay \leq b$ $y \geq 0$ $y = x$

Tabelle 2.1: Transformationen zwischen Polyedern.

## 2 Grundlagen und Notation

**Beweis.** Gilt  $C = P(A, 0)$ , so ist  $C$  ein Polyeder und offensichtlich ein Kegel.

Sei  $C$  ein polyedrischer Kegel, dann existieren nach Definition (2.1) c) eine Matrix  $A$  und ein Vektor  $b$  mit  $C = P(A, b)$ . Da jeder Kegel den Nullvektor enthält, gilt  $0 = A0 \leq b$ . Angenommen, es existiert ein  $\bar{x} \in C$  mit  $A\bar{x} \not\leq 0$ , d. h. es existiert eine Zeile von  $A$ , sagen wir  $A_i$ , mit  $t := A_i \bar{x} > 0$ . Da  $C$  ein Kegel ist, gilt  $\lambda \bar{x} \in C$  für alle  $\lambda \in \mathbb{K}_+$ . Jedoch für  $\bar{\lambda} := \frac{b_i}{t} + 1$  gilt einerseits  $\bar{\lambda} \bar{x} \in C$  und andererseits  $A_i(\bar{\lambda} \bar{x}) = \bar{\lambda} t > b_i$ , ein Widerspruch. Daraus folgt, für alle  $x \in C$  gilt  $Ax \leq 0$ . Hieraus ergibt sich  $C = P(A, 0)$ .  $\square$

Eine wichtige Eigenschaft von linearen Programmen ist die Tatsache, dass man effizient erkennen kann, ob ein gegebener Punkt tatsächlich der gesuchte Optimalpunkt ist oder nicht. Dahinter steckt die sogenannte *Dualitätstheorie*, die wir anhand unseres Einführungsbeispiels aus Kapitel 1.1 erläutern wollen. Durch Anwenden der Transformationsregeln können wir das LP (1.3) in der Form

$$\min -6s + t \tag{2.9a}$$

$$-s + t \geq -1 \tag{2.9b}$$

$$s - t \geq -1 \tag{2.9c}$$

$$s \geq 2 \tag{2.9d}$$

$$-s \geq -3 \tag{2.9e}$$

$$t \geq 0 \tag{2.9f}$$

$$-t \geq -3. \tag{2.9g}$$

schreiben. Wir wollen nun zeigen, dass der Punkt  $x^* = (3, 2)$  mit Zielfunktionswert  $-16$  minimal ist.

Ein Weg, dies zu beweisen, besteht darin, untere Schranken für das Minimum zu finden. Wenn man sogar in der Lage ist, eine untere Schranke zu bestimmen, die mit dem Zielfunktionswert einer Lösung des Ungleichungssystems übereinstimmt, ist man fertig: Der Zielfunktionswert irgendeiner Lösung ist immer eine obere Schranke, und wenn untere und obere Schranke gleich sind, ist der optimale Wert gefunden.

Um eine untere Schranke für die Zielfunktion  $-6s + t$  zu finden, können wir uns zunächst die Schranken der Variablen anschauen. Wegen Ungleichung (2.9e) haben wir  $-6s \geq -18$ . Zusammen mit Ungleichung (2.9f) erhalten wir  $-6s + t \geq -18$ . Um zu dieser unteren Schranke zu gelangen, haben wir positive Vielfache der Variablenschranken so addiert, das wir auf der linken Seite die Zielfunktion erhalten. (Da unsere Ungleichungen alle in der Form „ $\geq$ “ geschrieben sind, dürfen wir nur positiv skalieren, denn zwei Ungleichungen  $a_1x_1 + a_2x_2 \leq \alpha$  und  $b_1x_1 + b_2x_2 \geq \beta$  kann man nicht addieren.) Natürlich können wir dies mit allen Nebenbedingungsungleichungen machen, um so potenziell bessere untere Schranken zu bekommen. Gibt jede beliebige Skalierung und Addition eine untere Schranke? Machen wir noch einen Versuch. Addieren wir die Ungleichungen (2.9b) und (2.9e), so erhalten wir  $-2s + t \geq -4$ . Das hilft uns nicht weiter, denn die linke Seite der Ungleichung kann nicht zum Abschätzen der Zielfunktion benutzt werden: *Damit die rechte Seite der neuen Ungleichung eine untere Schranke für die Zielfunktion liefert, muss auf der linken Seite jeder Koeffizient höchstens so groß sein wie der entsprechende Koeffizient der Zielfunktion.*

Diese Erkenntnisse liefern uns ein neues mathematisches Problem. Wir suchen nichtnegative Multiplikatoren (Skalierungsfaktoren) der Ungleichungen (2.9b)–(2.9g) mit gewissen Eigenschaften. Multiplizieren wir die Ungleichungen mit  $y_1, \dots, y_6$ , so darf die Summe  $-y_1 + y_2 + y_3 - y_4$  nicht den Wert des ersten Koeffizienten der Zielfunktion (also  $-6$ ) überschreiten. Analog darf die Summe  $y_1 - y_2 + y_5 - y_6$  nicht den Wert 1 überschreiten. Und die  $y_i$  sollen nicht negativ sein. Ferner soll die rechte Seite der Summenungleichung, also  $-y_1 - y_2 + 2y_3 - 3y_4 - 3y_6$  so groß wie möglich werden. Daraus folgt, dass wir die folgende Aufgabe lösen müssen:

$$\max -y_1 - y_2 + 2y_3 - 3y_4 \quad - 3y_6 \tag{2.10a}$$

$$-y_1 + y_2 + y_3 - y_4 \leq -6 \tag{2.10b}$$

$$y_1 - y_2 + y_5 - y_6 \leq 1 \tag{2.10c}$$

$$y_1, y_2, y_3, y_4, y_5, y_6 \geq 0. \tag{2.10d}$$

Auch (2.10) ist ein lineares Programm. Aus unseren Überlegungen folgt, dass der Maximalwert von (2.10) höchstens so groß ist wie der Minimalwert von (1.3), da jede Lösung von (2.10) eine untere Schranke für (1.3) liefert. Betrachten wir z. B. den Punkt

$$y^* = (1, 0, 0, 5, 0, 0)^T.$$

Durch Einsetzen in die Ungleichungen von (2.10) sieht man dass  $y^*$  alle Ungleichungen erfüllt. Addieren wir also zu Ungleichung (2.9b) das 5-fache der Ungleichung (2.9f), so erhalten wir

$$-6s + t \geq -16.$$

Damit wissen wir, dass der Minimalwert von (1.3) mindestens  $-16$  ist. Der Punkt  $x^* = (3, 2)$  liefert gerade diesen Wert, er ist also optimal. Und außerdem ist  $y^*$  eine Optimallösung von (2.10).

Die hier dargestellten Ideen bilden die Grundgedanken der Dualitätstheorie der linearen Programmierung, und sie sind außerordentlich nützlich bei der Entwicklung von Algorithmen und in Beweisen. In der Tat haben wir bereits ein kleines Resultat erzielt, das wir wie folgt formal zusammenfassen wollen.

**(2.11) Satz.** *Es seien  $c \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$ , und  $A$  sei eine reelle  $(m, n)$ -Matrix. Betrachten wir die Aufgaben*

$$\min \{c^T x \mid Ax \geq b, x \geq 0\} \tag{P}$$

$$\text{und } \max \{y^T b \mid y^T A \leq c^T, y \geq 0\}, \tag{D}$$

*dann gilt Folgendes: Seien  $x_0 \in \mathbb{R}^n$  und  $y_0 \in \mathbb{R}^m$  Punkte mit  $Ax_0 \geq b$ ,  $x_0 \geq 0$  und  $y_0^T A \leq c^T$ ,  $y_0 \geq 0$ , dann gilt*

$$y_0^T b \leq c^T x_0. \tag{\triangle}$$

**Beweis.** Durch einfaches Einsetzen:

$$y_0^T b \leq y_0^T (Ax_0) = (y_0^T A)x_0 \leq c^T x_0. \tag{\square}$$

## Literaturverzeichnis

Satz (2.11) wird *schwacher Dualitätssatz* genannt, (D) heißt das zu (P) *duale LP* und (P) wird in diesem Zusammenhang als *primales LP* bezeichnet. Für Optimallösungen  $x^*$  und  $y^*$  von (P) bzw. (D) gilt nach (2.11)  $y^{*T}b \leq c^T x^*$ , das duale Problem liefert also stets eine untere Schranke für das primale Problem. Wir werden später zeigen, dass in diesem Falle sogar immer  $y^{*T}b = c^T x^*$  gilt. Diese *starke Dualität*, also das Übereinstimmen der Optimalwerte von (P) und (D) ist allerdings nicht ganz so einfach zu beweisen wie Satz (2.11).

Die schwache Dualität überträgt sich auch auf ganzzahlige lineare Programme, für die jedoch im Allgemeinen kein starker Dualitätssatz gilt. Genauer gilt folgender Zusammenhang:

$$\begin{aligned} \min \{c^T x \mid Ax \geq b, x \geq 0, x \in \mathbb{Z}\} \\ &\geq \min \{c^T x \mid Ax \geq b, x \geq 0\} \\ &= \max \{y^T b \mid y^T A \leq c^T, y \geq 0\} \\ &\geq \max \{y^T b \mid y^T A \leq c^T, y \geq 0, y \in \mathbb{Z}\}. \end{aligned}$$

## Literaturverzeichnis

Zur linearen Algebra existieren unzählige Bücher. Deswegen geben wir hierzu keine Literaturliste an, sondern listen hier ausschließlich Literatur zur Graphentheorie und zur linearen Programmierung.

M. Aigner. *Graphentheorie: Eine Entwicklung aus dem 4-Farben-Problem*. Teubner Verlag, Studienbücher: Mathematik, Stuttgart, 1984. ISBN 3-519-02068-8.

Berge and Ghouila-Houri. *Programme, Spiele, Transportnetze*. Teubner Verlag, Leipzig, 1969.

B. Bollobás. *Modern Graph Theory*. Springer Verlag, New York, 1998. ISBN 0-387-98488-7.

J. A. Bondy and U. S. R. Murty. *Graph Theory*. Springer, Berlin, 2008.

V. Chvátal. *Linear Programming*. Freeman, New York, 1983.

G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, 1998.

R. Diestel. *Graphentheorie*. Springer-Verlag, Heidelberg, 3. auflage edition, 2006. ISBN 3-540-21391-0.

W. Domschke. *Logistik: Rundreisen und Touren*. Oldenbourg-Verlag, München – Wien, 4., erweiterte Auflage, 1997.

J. Ebert. Effiziente Graphenalgorithmen. *Studentexte: Informatik*, 1981.



- R. L. Graham, M. Grötschel, and L. Lovász, editors. *Handbook of Combinatorics, Volume I*. Elsevier (North-Holland); The MIT Press, Cambridge, Massachusetts, 1995. ISBN 0-444-82346-8/v.1 (Elsevier); ISBN 0-262-07170-3/v.1 (MIT).
- P. Gritzmann. *Grundlagen der mathematischen Optimierung*. Springer Spektrum, 2013.
- J. L. Gross and J. Yellen. *Handbook of Graph Theory*. CRC Press, Boca Raton, 2004. ISBN 1-58488-090-2.
- R. Halin. *Graphentheorie*. Akademie-Verlag Berlin, 2. Auflage, 1989.
- K. Hässig. *Graphentheoretische Methoden des Operations Research*. Teubner-Verlag, Stuttgart, 1979.
- D. Jungnickel. *Graphs, Networks and Algorithms*. Springer, 4. edition, 2013.
- D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig, 1936. Mehrfach auf deutsch und in englischer Übersetzung nachgedruckt.
- S. O. Krumke and H. Noltemeier. *Graphentheoretische Konzepte und Algorithmen*. Teubner, Wiesbaden, 2005. ISBN 3-519-00526-3.
- J. Matousek and B. Gärtner. *Using and Understanding Linear Programming*. Springer, 2007.
- M. Padberg. *Linear Optimization and Extensions*. Springer, 2001.
- H. Sachs. *Einführung in die Theorie der endlichen Graphen*. Teubner, Leipzig, 1970, und Hanser, München, 1971, 1970.
- A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, New York, 1998.
- R. J. Vanderbei. *Linear Programming – Foundations and Extensions*. Springer, 4th edition, 2014.
- K. Wagner. *Graphentheorie*. BI Wissenschaftsverlag, Mannheim, 1970.
- H. Walther and G. Nägler. *Graphen, Algorithmen, Programme*. VEB Fachbuchverlag, Leipzig, 1987.
- D. B. West. *Introduction to Graph Theory*. Prentice Hall, Upper Saddle River, third edition, 2005. ISBN 0-13-014400-2.



# 3 Diskrete Optimierungsprobleme

Dieses Kapitel enthält eine Liste von algorithmischen Fragestellungen der Graphentheorie. Wir werden – neben historisch interessanten Aufgaben – insbesondere Optimierungsprobleme aufführen, die ein weites Anwendungsspektrum besitzen.

## 3.1 Kombinatorische Optimierungsprobleme

Bevor wir auf graphentheoretische Optimierungsprobleme eingehen, führen wir kombinatorische Optimierungsprobleme in allgemeiner Form ein.

**(3.1) Definition (Allgemeines kombinatorisches Optimierungsproblem).** *Gegeben seien eine endliche Menge  $\mathcal{I}$  und eine Funktion  $f : \mathcal{I} \rightarrow \mathbb{R}$ , die jedem Element von  $\mathcal{I}$  einen „Wert“ zuordnet. Gesucht ist ein Element  $I^* \in \mathcal{I}$ , so daß  $f(I^*)$  so groß (oder klein) wie möglich ist.  $\triangle$*

Eine Problemformulierung dieser Art ist relativ sinnlos, da über ein Problem, das wie oben gegeben ist, kaum vernünftige mathematische Aussagen gemacht werden können. Algorithmisch ist (3.1) auf triviale Weise lösbar: man durchlaufe alle Elemente  $I$  von  $\mathcal{I}$ , werte die Funktion  $f(I)$  aus und wähle das Element  $I^*$  mit dem größten (oder kleinsten) Wert  $f(I^*)$  aus. Falls die Elemente  $I \in \mathcal{I}$  algorithmisch bestimmbar und  $f(I)$  auswertbar ist, hat der eben beschriebene Enumerationsalgorithmus eine sogenannte lineare Laufzeit, da jedes Element von  $\mathcal{I}$  nur einmal betrachtet wird.

Die üblicherweise auftretenden kombinatorischen Optimierungsprobleme sind jedoch auf andere, wesentlich strukturiertere Weise gegeben. Die Menge  $\mathcal{I}$  ist nicht durch explizite Angabe aller Elemente spezifiziert sondern implizit durch die Angabe von Eigenschaften, die die Elemente von  $\mathcal{I}$  haben sollen. Ebenso ist die Funktion  $f$  nicht punktweise sondern durch „Formeln“ definiert.

In dieser Vorlesung wollen wir uns hauptsächlich auf den folgenden Problemtyp konzentrieren.

**(3.2) Definition (Komb. Optimierungsproblem mit linearer Zielfunktion).** *Gegeben seien eine endliche Menge  $E$  (genannt Grundmenge), eine Teilmenge  $\mathcal{I}$  der Potenzmenge  $2^E$  von  $E$  (die Elemente von  $\mathcal{I}$  heißen zulässige Mengen oder zulässige Lösungen) und eine Funktion  $c : E \rightarrow \mathbb{R}$ . Für jede Menge  $F \subseteq E$  definieren wir ihren „Wert“ durch*

$$c(F) := \sum_{e \in F} c(e),$$

*und wir suchen eine Menge  $I^* \in \mathcal{I}$ , so dass  $c(I^*)$  so groß (oder klein) wie möglich ist.  $\triangle$*

### 3 Diskrete Optimierungsprobleme

Zur Notationsvereinfachung werden wir in Zukunft einfach *kombinatorisches Optimierungsproblem* sagen, wenn wir ein Problem des Typs (3.2) meinen. Da ein derartiges Problem durch die Grundmenge  $E$ , die zulässigen Lösungen  $\mathcal{I}$  und die Zielfunktion  $c$  definiert ist, werden wir kurz von einem kombinatorischen Optimierungsproblem  $(E, \mathcal{I}, c)$  sprechen.

Die Zielfunktion haben wir durch Formulierung (3.2) bereits sehr speziell strukturiert. Aber Problem (3.2) ist algorithmisch immer noch irrelevant, falls wir eine explizite Angabe von  $\mathcal{I}$  unterstellen. Wir werden nachfolgend (und im Verlaufe der Vorlesung noch sehr viel mehr) Beispiele des Typs (3.2) kennenlernen. Fast alle der dort auftretenden zulässigen Mengen lassen sich auf folgende Weise charakterisieren:

$$\mathcal{I} = \{I \subseteq E \mid I \text{ hat Eigenschaft } \Pi\}.$$

Wir werden uns damit beschäftigen, welche Charakteristika die Eigenschaft  $\Pi$  haben muss, damit die zugehörigen Probleme  $(E, \mathcal{I}, c)$  auf einfache Weise gelöst werden können. Nehmen wir an, dass  $E$  insgesamt  $n$  Elemente enthält, dann führt natürlich jede Eigenschaft  $\Pi$ , die impliziert, dass  $\mathcal{I}$  (relativ zu  $n$ ) nur sehr wenige Elemente enthält, dazu, dass  $(E, \mathcal{I}, c)$  einfach lösbar ist, falls man die Elemente von  $\mathcal{I}$  explizit angeben kann. Typischerweise haben jedoch die interessanten kombinatorischen Optimierungsprobleme eine Anzahl von Lösungen, die exponentiell in  $n$  ist, etwa  $n!$  oder  $2^n$ . Eine vollständige Enumeration der Elemente solcher Mengen ist offenbar auch auf den größten Rechnern (für z. B.  $n \geq 40$ ) nicht in „vernünftiger Zeit“ durchführbar. Das Ziel der kombinatorischen Optimierung besteht – kurz und vereinfachend gesagt – darin, Algorithmen zu entwerfen, die (erheblich) schneller als die Enumeration aller Lösungen sind.

## 3.2 Klassische Fragestellungen der Graphentheorie

Nachfolgend werden eine Reihe von graphentheoretischen Problemen skizziert, die die Entwicklung der Graphentheorie nachhaltig beeinflusst haben.

**(3.3) Euler und das Königsberger Brückenproblem.** Fast jedes Buch über Graphentheorie (Geben Sie einfach einmal “Königsberg bridges” in Google ein.) enthält einen Stadtplan von Königsberg und erläutert, wie Euler die Königsberger Karte zu dem Graphen aus Abbildung 3.1 „abstrahiert“ hat. Euler hat die Frage untersucht, ob es in diesem „Königsberger Brückengraphen“ einen geschlossenen Pfad gibt, der alle Kanten genau einmal enthält. Heute nennen wir einen solchen Pfad *Eulertour*. Er hat das Problem nicht nur für den Graphen aus Abbildung 3.1 gelöst, sondern für alle Graphen: Ein Graph enthält eine Eulertour genau dann, wenn er zusammenhängend ist und jeder Knoten einen geraden Grad hat. Diesen Satz hat Euler 1736 bewiesen und damit die Graphentheorie begründet. Hinweise zur Geschichte des Königsberger Brückenproblems und zu verwandten Optimierungsproblemen finden Sie u. a. in Grötschel and Yuan (2012).  $\triangle$

**(3.4) Das Haus vom Nikolaus.** Jeder kennt die Aufgabe aus dem Kindergarten: Zeichne das Haus des Nikolaus, siehe Abbildung 3.2, in einem Zug! Was hat diese Fragestellung mit dem Königsberger Brückenproblem zu tun?  $\triangle$

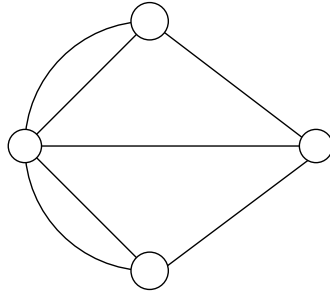


Abbildung 3.1: Das Königsberger Brückenproblem.

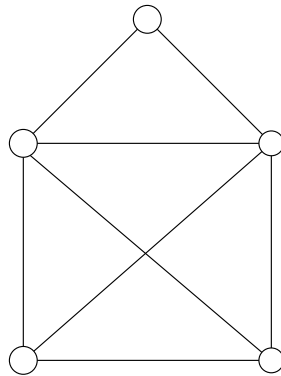
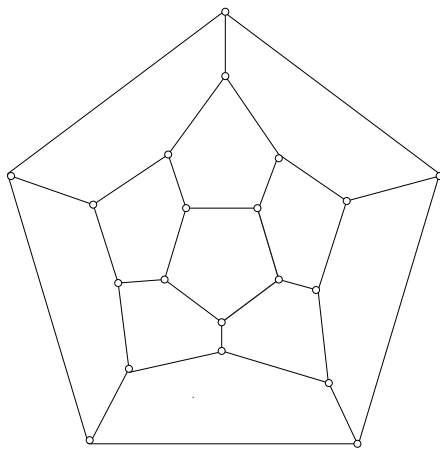


Abbildung 3.2: Das Haus vom Nikolaus.

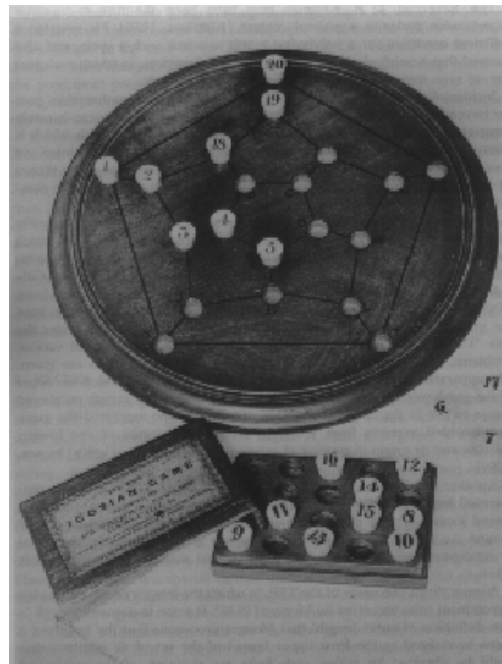
**(3.5) Hamiltonsche Kreise.** Der irische Mathematiker Sir William Hamilton (z. B. durch die „Erfindung“ der Quaternionen bekannt) hat sich Ende der 50er Jahre des 19. Jahrhunderts mit Wege-Problemen beschäftigt und sich besonders dafür interessiert, wie man auf dem Dodekaedergraphen, siehe Abbildung 3.3(a), Kreise findet, die alle Knoten durchlaufen (heute *hamiltonsche Kreise* genannt) und die noch gewissen Zusatzanforderungen genügen. Er fand diese Aufgabe so spannend, dass er sie als Spiel vermarktet hat (offenbar nicht sonderlich erfolgreich). Ein Exemplar dieses Spiels mit dem Namen “The Icosian Game” befindet sich noch in der Bibliothek des Trinity College in Dublin, Irland, siehe Abbildung 3.3(b).

Die Aufgabe, in einem Graphen, einen Hamiltonkreis zu finden, sieht so ähnlich aus wie das Problem, eine Eulertour zu bestimmen. Sie ist aber viel schwieriger. Das hamiltonische Graphen-Problem hat sich später zum Travelling-Salesman-Problem „entwickelt“. Historische Bemerkungen hierzu findet man zum Beispiel in Applegate et al. (2006) und Cook (2012). △

**(3.6) Färbung von Landkarten.** Nach Aigner (1984), der die Entwicklung der Graphentheorie anhand der vielfältigen Versuche, das 4-Farben-Problem zu lösen, darstellt, begann die mathematische Beschäftigung mit dem Färbungsproblem im Jahre 1852 mit einem Brief von Augustus de Morgan an William Hamilton:



(a)



(b)

Abbildung 3.3: (a) Graph mit einem Hamilton-Kreis. (b) The Icosian Game.

„Ein Student fragte mich heute, ob es stimmt, dass die Länder jeder Karte stets mit höchstens 4 Farben gefärbt werden können, unter der Maßgabe, dass angrenzende Länder verschiedene Farben erhalten.“

Der Urheber der Frage war *Francis Guthrie*.

Aus einer Landkarte kann man einen Graphen machen, indem jedes Land durch einen Knoten repräsentiert wird und je zwei Knoten genau dann durch eine Kante verbunden werden, wenn die zugehörigen Länder benachbart sind. Abbildung 3.4(a) zeigt die Karte der deutschen Bundesländer. Der „Bundesländergraph“ in Abbildung 3.4(b) hat daher je einen Knoten für die 16 Länder und einen weiteren Knoten für die „Außenwelt“. Dieser Knoten ist mit allen Bundesländerknoten verbunden, die an das Ausland oder das Meer (wie etwa Niedersachsen) grenzen.

„Landkartengraphen“ kann man nach Konstruktion in die Ebene so zeichnen, dass sich je zwei Kanten (genauer: die Linien, die die Kanten in der Ebene repräsentieren) nicht schneiden (außer natürlich in ihren Endpunkten, wenn sie einen gemeinsamen Knoten besitzen). Landkartengraphen sind also planar. Das 4-Farben-Problem (in etwas allgemeinerer Form) lautet dann: „Kann man die Knoten eines planaren Graphen so färben, dass je zwei benachbarte Knoten verschiedene Farben besitzen?“

Der Weg zur Lösung des 4-Farben-Problems war sehr lang, siehe hierzu Aigner (1984). Die erste vollständige Lösung (unter Zuhilfenahme von Computerprogrammen) wurde 1976/1977 von K. Appel und W. Haken vorgelegt. Die Dokumentation eines transparen-

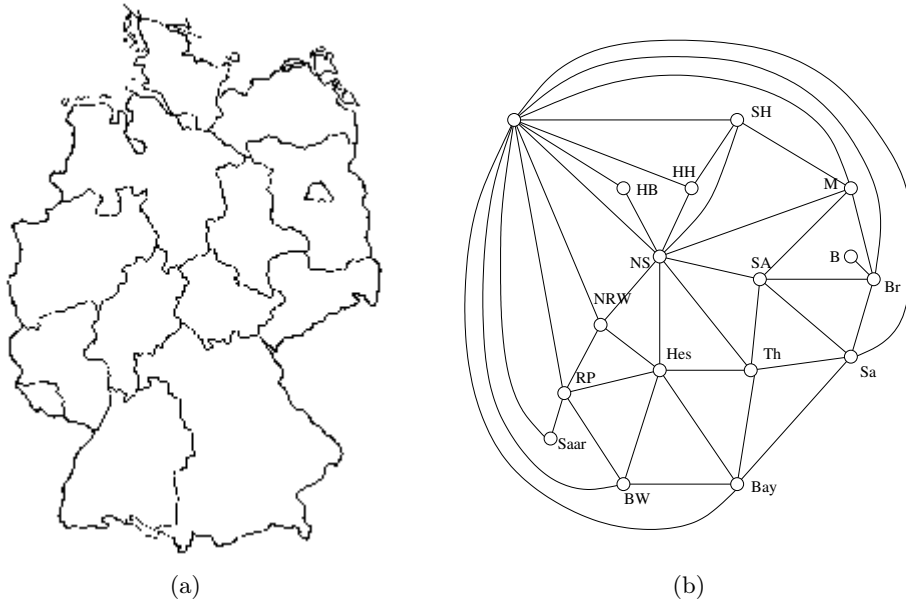


Abbildung 3.4: Abbildung (a) ist eine Karte von Deutschland mit seinen Bundesländern. In (b) repräsentiert jede Landeshauptstadt ihr zugehöriges Bundesland. Die Nachbarschaft eines jeden Knoten besteht aus den entsprechenden Nachbarländern.

ten Beweises von N. Robertson, D.P. Sanders, P. Seymour und R. Thomas, der weiterhin auf der Überprüfung vieler Einzelfälle durch Computerprogramme beruht, ist auf der Homepage von Robin Thomas zu finden:

<http://www.math.gatech.edu/~thomas/FC/fourcolor.html>

△

**(3.7) Planarität.** Durch das 4-Farben-Problem gelangte die Frage, wann kann man einen Graphen so in die Ebene einbetten, dass sich je zwei Kanten nicht überschneiden, in den Fokus der Forschung. Natürlich wurde sofort verallgemeinert: „Finde eine ‚gute‘ Charakterisierung dafür, dass ein Graph in die Ebene, auf dem Torus, in die projektive Ebene, auf Henkelflächen etc. überschneidungsfrei einbettbar ist.“

Kuratowski gelang 1930 ein entscheidender Durchbruch. Es ist einfach zu sehen, dass weder der vollständige Graph  $K_5$  noch der vollständige Graph  $K_{3,3}$  planar sind. Kuratowski bewies, dass jeder nicht-planare Graph einen der Graphen  $K_5$  oder  $K_{3,3}$  „enthält“. Das heißt, ist  $G$  nicht planar, so kann man aus  $G$  durch Entfernen und durch Kontraktion von Kanten entweder den  $K_5$  oder den  $K_{3,3}$  erzeugen. Dies ist auch heute noch ein keineswegs triviales Ergebnis. △

### 3.3 Graphentheoretische Optimierungsprobleme: Einige Beispiele

In diesem Abschnitt wollen wir mehrere Beispiele von kombinatorischen Optimierungsproblemen, die sich mit Hilfe von Graphentheorie formulieren lassen, und einige ihrer Anwendungen auflisten. Diese Sammlung ist nicht im geringsten vollständig, sondern umfasst nur einige in der Literatur häufig diskutierte oder besonders anwendungsnahe Probleme. Wir benutzen dabei gelegentlich englische Namen, die mittlerweile auch im Deutschen zu Standardbezeichnungen geworden sind. Fast alle der nachfolgend aufgeführten „Probleme“ bestehen aus mehreren eng miteinander verwandten Problemtypen. Wir gehen bei unserer Auflistung so vor, dass wir meistens zunächst die graphentheoretische Formulierung geben und dann einige Anwendungen skizzieren.

**(3.8) Kürzeste Wege.** Gegeben seien ein Digraph  $D = (V, A)$  und zwei verschiedene Knoten  $u, v \in V$ , stelle fest, ob es einen gerichteten Weg von  $u$  nach  $v$  gibt. Falls das so ist, und falls „Entfernungen“  $c_{ij} \geq 0$  für alle  $(i, j) \in A$  bekannt sind, bestimme einen kürzesten gerichteten Weg von  $u$  nach  $v$  (d. h. einen  $(u, v)$ -Weg  $P$ , so dass  $c(P)$  minimal ist). Dieses Problem wird üblicherweise *Problem des kürzesten Weges* (*shortest path problem*) genannt. Zwei interessante Varianten sind die folgenden: Finde einen kürzesten  $(u, v)$ -Weg gerader bzw. ungerader Länge (d. h. mit gerader bzw. ungerader Bogenzahl).

Das Problem des kürzesten Weges gibt es auch in einer ungerichteten Version. Hier sucht man in einem Graphen  $G = (V, E)$  mit Entfernungen  $c_e \geq 0$  für alle  $e \in E$  bei gegebenen Knoten  $u, v \in V$ ,  $u \neq v$ , einen kürzesten  $[u, v]$ -Weg. Analog kann man nach einem kürzesten Weg gerader oder ungerader Länge fragen.

Natürlich kann man in allen bisher angesprochenen Problemen, das Wort „kürzester“ durch „längster“ ersetzen und erhält dadurch *Probleme der längsten Wege* verschiedener Arten. Hätten wir beim Problem des kürzesten Weges nicht die Beschränkung  $c_{ij} \geq 0$  für die Zielfunktionskoeffizienten, wären die beiden Problemtypen offensichtlich äquivalent. Aber so sind sie es nicht! Ein Spezialfall (Zielfunktion  $c_e = 1$  für alle  $e \in E$ ) des Problems des längsten Weges ist das Problem zu entscheiden, ob ein Graph einen hamiltonschen Weg von  $u$  nach  $v$  enthält.  $\triangle$

Anwendungen dieses Problems und seiner Varianten sind offensichtlich. Alle Routenplaner, die im Internet zur Fahrstreckenplanung angeboten werden oder zur Unterstützung von Autofahrern in Navigationssysteme eingebaut sind, basieren auf Algorithmen zur Bestimmung kürzester Wege. Die Route jeder im Internet verschickten Nachricht wird ebenfalls durch (mehrfachen) Aufruf eines Kürzeste-Wege-Algorithmus ermittelt. Eine Anwendung aus der Wirtschafts- und Sozialgeographie, die nicht unbedingt im Gesichtsfeld von Mathematikern liegt, sei hier kurz erwähnt. Bei Fragen der Raumordnung und Landesplanung werden sehr umfangreiche Erreichbarkeitsanalysen angestellt, um Einzugsbereiche (bzgl. Straßen-, Nahverkehrs- und Bahnanbindung) festzustellen. Auf diese Weise werden Mittel- und Oberzentren des ländlichen Raumes ermittelt und Versorgungsgrade der Bevölkerung in Bezug auf Ärzte, Krankenhäuser, Schulen etc. bestimmt. Ebenso erfolgen Untersuchungen bezüglich des Arbeitsplatzangebots. Alle diese



Analysen basieren auf einer genauen Ermittlung der Straßen-, Bus- und Bahnentfernungen (in Kilometern oder Zeiteinheiten) und Algorithmen zur Bestimmung kürzester Wege in den „Verbindungsnetzwerken“.

**(3.9) Das Zuordnungsproblem (assignment problem).** Gegeben sei ein bipartiter Graph  $G = (V, E)$  mit Kantengewichten  $c_e \in \mathbb{R}$  für alle  $e \in E$ , gesucht ist ein Matching in  $G$  maximalen Gewichts. Man nennt dieses Problem das *Matchingproblem in bipartiten Graphen* oder kurz *bipartites Matchingproblem*. Haben die beiden Knotenmengen in der Bipartition von  $V$  gleiche Kardinalität und sucht man ein perfektes Matching minimalen Gewichts, so spricht man von einem *Zuordnungsproblem*. Es gibt noch eine weitere Formulierung des Zuordnungsproblems. Gegeben sei ein Digraph  $D = (V, A)$ , der auch Schlingen haben darf, mit Bogengewichten (meistens wird unterstellt, dass  $D$  vollständig ist und Schlingen hat), gesucht ist eine Bogenmenge minimalen Gewichts, so dass jeder Knoten von  $D$  genau einmal Anfangs- und genau einmal Endknoten eines Bogens aus  $B$  ist. ( $B$  ist also eine Menge knotendisjunkter gerichteter Kreise, so dass jeder Knoten auf genau einem Kreis liegt. Eine Schlinge wird hierbei als ein gerichteter Kreis aufgefasst.) Wir wollen dieses Problem *gerichtetes Zuordnungsproblem* nennen.  $\triangle$

Das Zuordnungsproblem hat folgende „Anwendung“. Gegeben seien  $n$  Männer und  $n$  Frauen, für  $1 \leq i, j \leq n$  sei  $c_{ij}$  ein „Antipathiekoeffizient“. Gesucht ist eine Zuordnung von Männern zu Frauen (Heirat), so dass die Summe der Antipathiekoeffizienten minimal ist. Dieses Problem wird häufig *Heiratsproblem* genannt.

Das Matchingproblem in bipartiten Graphen kann man folgendermaßen interpretieren. Ein Betrieb habe  $m$  offene Stellen und  $n$  Bewerber für diese Positionen. Durch Tests hat man herausgefunden, welche Eignung Bewerber  $i$  für die Stelle  $j$  hat. Diese „Kompetenz“ sei mit  $c_{ij}$  bezeichnet. Gesucht wird eine Zuordnung von Bewerbern zu Positionen, so dass die „Gesamtkompetenz“ maximal wird.

Das Zuordnungsproblem und das Matchingproblem in bipartiten Graphen sind offenbar sehr ähnlich, die Beziehungen zwischen dem Zuordnungsproblem und seiner gerichteten Version sind dagegen nicht ganz so offensichtlich. Dennoch sind diese drei Probleme in folgendem Sinne „äquivalent“: man kann sie auf sehr einfache Weise ineinander transformieren, d. h. mit einem schnellen Algorithmus zur Lösung des einen Problems kann man die beiden anderen Probleme lösen, ohne komplizierte Transformationsalgorithmen einzuschalten.

Transformationstechniken, die einen Problemtyp in einen anderen überführen, sind außerordentlich wichtig und zwar sowohl aus theoretischer als auch aus praktischer Sicht. In der Theorie werden sie dazu benutzt, Probleme nach ihrem Schwierigkeitsgrad zu klassifizieren (siehe Kapitel 4), in der Praxis ermöglichen sie die Benutzung eines einzigen Algorithmus zur Lösung der verschiedensten Probleme und ersparen daher erhebliche Codierungs- und Testkosten. Anhand der drei vorgenannten Probleme wollen wir nun derartige Transformationstechniken demonstrieren.

**Bipartites Matchingproblem  $\longrightarrow$  Zuordnungsproblem.** Angenommen wir haben ein Matchingproblem in einem bipartiten Graphen und wollen es mit einem Algorithmus

### 3 Diskrete Optimierungsprobleme

für Zuordnungsprobleme lösen. Das Matchingproblem ist gegeben durch einen bipartiten Graphen  $G = (V, E)$  mit Bipartition  $V_1, V_2$  und Kantengewichten  $c_e \in \mathbb{R}$  für alle  $e \in E$ . O. B. d. A. können wir annehmen, dass  $m = |V_1| \leq |V_2| = n$  gilt. Zur Menge  $V_1$  fügen wir  $n - m$  neue Knoten  $W$  (künstliche Knoten) hinzu. Wir setzen  $V'_1 := V_1 \cup W$ . Für je zwei Knoten  $i \in V'_1$  und  $j \in V_2$ , die nicht in  $G$  benachbart sind, fügen wir eine neue (künstliche) Kante  $ij$  hinzu. Die Menge der so hinzugefügten Kanten nennen wir  $E'$ , und den Graphen  $(V'_1 \cup V_2, E \cup E')$  bezeichnen wir mit  $G'$ .  $G'$  ist der vollständige bipartite Graph  $K_{n,n}$ . Wir definieren neue Kantengewichte  $c'_e$  wie folgt:

$$c'_e := \begin{cases} 0 & \text{falls } e \in E' \\ 0 & \text{falls } e \in E \text{ und } c_e \leq 0 \\ -c_e & \text{falls } e \in E \text{ und } c_e > 0 \end{cases}$$

Lösen wir das Zuordnungsproblem bezüglich  $G'$  mit den Gewichten  $c'_e$ ,  $e \in E \cup E'$ , so erhalten wir ein perfektes Matching  $M'$  minimalen Gewichts bezüglich  $c'$ . Es ist nun einfach zu sehen, dass

$$M := \{e \in M' \mid c'_e < 0\}$$

ein Matching in  $G$  ist, das maximal bezüglich der Gewichtsfunktion  $c$  ist.

**Zuordnungsproblem  $\rightarrow$  gerichtetes Zuordnungsproblem.** Wir zeigen nun, dass man das Zuordnungsproblem mit einem Algorithmus für das gerichtete Zuordnungsproblem lösen kann. Gegeben sei also ein bipartiter Graph  $G = (V, E)$  mit Bipartition  $V_1, V_2$  und Kantengewichten  $c_e$ . Es gelte  $V_1 = \{u_1, u_2, \dots, u_n\}$ ,  $V_2 = \{v_1, v_2, \dots, v_n\}$ . Wir definieren einen Digraphen  $D = (W, A)$  mit  $W = \{w_1, \dots, w_n\}$ . Zwei Knoten  $w_i, w_j \in W$  sind genau dann durch einen Bogen  $(w_i, w_j)$  verbunden, wenn  $u_i v_j \in E$  gilt. Das Gewicht  $c'((w_i, w_j))$  des Bogens  $(w_i, w_j)$  sei das Gewicht  $c(u_i v_j)$  der Kante  $u_i v_j$ . Ist  $B$  eine minimale Lösung des gerichteten Zuordnungsproblems bezüglich  $D$  und  $c'$ , so ist

$$M := \{u_i v_j \in E \mid (w_i, w_j) \in B\}$$

offenbar ein minimales perfektes Matching in  $G$  bezüglich der Gewichtsfunktion  $c$ . Es ist ebenfalls sofort klar, dass das gerichtete Zuordnungsproblem bezüglich  $D$  eine Lösung genau dann hat, wenn  $G$  ein perfektes Matching enthält.

**Gerichtetes Zuordnungsproblem  $\rightarrow$  bipartites Matchingproblem.** Schließlich wollen wir noch vorführen, dass man das gerichtete Zuordnungsproblem auf das Matchingproblem in bipartiten Graphen zurückführen kann. Gegeben sei also ein Digraph  $D = (W, A)$  mit  $W = \{w_1, \dots, w_n\}$  und Bogengewichten  $c((w_i, w_j))$  für alle  $(w_i, w_j) \in A$ . Wir definieren einen bipartiten Graphen  $G = (V, E)$  mit Bipartition  $V_1 = \{u_1, \dots, u_n\}$ ,  $V_2 = \{v_1, \dots, v_n\}$  und Kantenmenge  $E := \{u_i v_j \mid (w_i, w_j) \in A\}$ . Es seien

$$z := n (\max\{c((w_i, w_j)) \mid (w_i, w_j) \in A\}) + 1$$

und

$$c'(u_i v_j) := -c((w_i, w_j)) + z.$$

### 3.3 Graphentheoretische Optimierungsprobleme: Beispiele

Nach Konstruktion gilt, dass jedes Matching in  $G$  mit  $k$  Kanten ein geringeres Gewicht hat als ein Matching mit  $k+1$  Kanten,  $k = 0, \dots, n-1$ . Daraus folgt, dass es eine Lösung des gerichteten Zuordnungsproblems bezüglich  $D$  genau dann gibt, wenn jedes maximale Matching  $M$  bezüglich  $G$  und  $c'$  perfekt ist. Ist dies so, dann ist

$$B := \{(w_i, w_j) \in A \mid u_i v_j \in M\}$$

eine minimale Lösung des gerichteten Zuordnungsproblems mit Gewicht  $c(B) = -c'(M) + nz$ .

**(3.10) Das Matchingproblem.** Die Grundversion dieses Problems ist die folgende. Gegeben sei ein Graph  $G = (V, E)$  mit Kantengewichten  $c_e$  für alle  $e \in E$ . Ist ein Matching  $M$  von  $G$  maximalen Gewichts  $c(M)$  gesucht, so heißt dieses Problem *Matchingproblem*. Sucht man ein perfektes Matching minimalen Gewichts, so wird es *perfektes Matchingproblem* genannt.

Diese Probleme können wie folgt verallgemeinert werden. Gegeben seien zusätzlich nichtnegative ganze Zahlen  $b_v$  für alle  $v \in V$  (genannt *Gradbeschränkungen*) und  $u_e$  für alle  $e \in E$  (genannt *Kantenkapazitäten*). Ein (*perfektes*) *b-Matching* ist eine Zuordnung  $x_e$  von nichtnegativen ganzen Zahlen zu den Kanten  $e \in E$ , so dass für jeden Knoten  $v \in V$  die Summe der Zahlen  $x_e$  über die Kanten  $e \in E$ , die mit  $v$  inzidieren, höchstens (exakt)  $b_v$  ist. Das *unkapazitierte (perfekte) b-Matchingproblem* ist die Aufgabe ein (*perfektes*) *b-Matching*  $(x_e)_{e \in E}$  zu finden, so dass  $\sum_{e \in E} c_e x_e$  maximal (minimal) ist. Sollen die ganzzahligen Kantenwerte  $x_e$  für alle  $e \in E$  zusätzlich noch die Kapazitätsschranken  $0 \leq x_e \leq u_e$  erfüllen, so spricht man von einem (*perfekten*) *u-kapazitierten b-Matchingproblem*.  $\triangle$

An dieser Stelle wollen wir noch eine – nicht offensichtliche – Problemtransformation vorführen. Und zwar wollen wir zeigen, dass die Aufgabe, in einem ungerichteten Graphen  $G = (V, E)$  mit Kantengewichten  $c_e \geq 0$  für alle  $e \in E$  einen kürzesten Weg ungerader Länge zwischen zwei Knoten  $u, v \in V$  zu bestimmen, mit einem Algorithmus für das perfekte Matchingproblem gelöst werden kann. Und zwar konstruieren wir aus  $G$  einen neuen Graphen  $G'$  wie folgt. Nehmen wir an, dass  $V = \{v_1, \dots, v_n\}$  gilt. Die Graphen  $G_1 = (U, E_1)$  mit  $U := \{u_1, \dots, u_n\}$  und  $G_2 = (W, E_2)$  mit  $W := \{w_1, \dots, w_n\}$  seien knotendisjunkte isomorphe Bilder (also Kopien) von  $G$ , so dass die Abbildungen  $v_i \mapsto u_i$  und  $v_i \mapsto w_i$ ,  $i = 1, \dots, n$  Isomorphismen sind. Aus  $G_2$  entfernen wir die Bilder der Knoten  $u$  und  $v$ , dann verbinden wir die übrigen Knoten  $w_i \in W$  mit ihren isomorphen Bildern  $u_i \in U$  durch eine Kante  $u_i w_i$ . Diese neuen Kanten  $u_i w_i$  erhalten das Gewicht  $c(u_i w_i) = 0$ . Die Kanten aus  $G_1$  und  $G_2 - \{u, v\}$ , die ja Bilder von Kanten aus  $G$  sind, erhalten das Gewicht ihrer Urbildkanten. Der Graph  $G'$  entsteht also aus der Vereinigung von  $G_1$  mit  $G_2 - \{u, v\}$  unter Hinzufügung der Kanten  $u_i w_i$ , siehe Abbildung 3.5. Man überlegt sich leicht, dass jedes perfekte Matching in  $G'$  einer Kantenmenge in  $G$  entspricht, die einen ungeraden  $[u, v]$ -Weg in  $G$  enthält und dass jedes minimale perfekte Matching in  $G'$  einen minimalen ungeraden  $[u, v]$ -Weg bestimmt.

**Hausaufgabe.** Finden Sie eine ähnliche Konstruktion, die das Problem, einen kürzesten  $[u, v]$ -Weg gerader Länge zu bestimmen, auf ein perfektes Matchingproblem zurückführt!

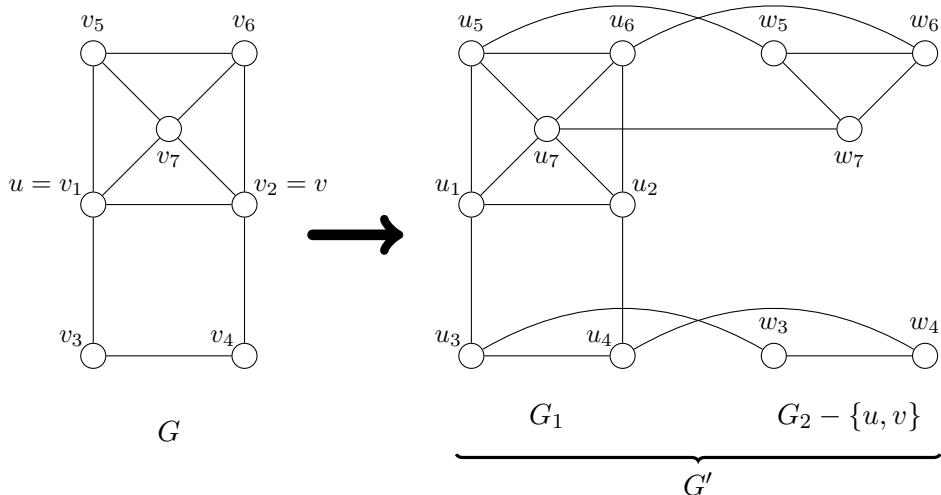


Abbildung 3.5: Hier entspricht z. B. dem ungeraden  $[u, v]$ -Weg  $(u, v_7, v_6, v)$  das perfekte Matching  $M = \{u_1u_7, w_7w_6, u_6u_2, u_3w_3, u_4w_4, u_5w_5\}$  und umgekehrt.

**(3.11) Wälder, Bäume, Branchings, Arboreszenzen.** Gegeben sei ein Graph  $G = (V, E)$  mit Kantengewichten  $c_e \in \mathbb{R}$  für alle  $e \in E$ . Die Aufgabe, einen Wald  $W \subseteq E$  zu finden, so dass  $c(W)$  maximal ist, heißt *Problem des maximalen Waldes*.

Die Aufgabe einen Baum  $T \subseteq E$  zu finden, der  $G$  aufspannt und dessen Gewicht  $c(T)$  minimal ist, heißt *Problem des minimalen aufspannenden Baumes (minimum spanning tree problem)*. Diese beiden Probleme haben auch eine gerichtete Version.

Gegeben sei ein Digraph  $D = (V, A)$  mit Bogengewichten  $c_a \in \mathbb{R}$  für alle  $a \in A$ . Die Aufgabe, ein Branching  $B \subseteq A$  maximalen Gewichts zu finden, heißt *maximales Branching-Problem*, die Aufgabe, eine Arboreszenz (mit vorgegebener Wurzel  $r$ ) von  $D$  minimalen Gewichts zu finden, heißt *minimales Arboreszenz-Problem ( $r$ -Arboreszenz-Problem)*.  $\triangle$

Die im folgenden Punkt zusammengefassten Probleme gehören zu den am meisten untersuchten und anwendungsreichsten Problemen.

**(3.12) Routenplanung.** Gegeben seien  $n$  Städte und Entfernungen  $c_{ij}$  zwischen diesen, gesucht ist eine Rundreise (*Tour*), die durch alle Städte genau einmal führt und minimale Länge hat. Haben die Entfernungen die Eigenschaft, dass  $c_{ij} = c_{ji}$  gilt,  $1 \leq i < j \leq n$ , so nennt man dieses Problem *symmetrisches Travelling-Salesman-Problem (TSP)*, andernfalls heißt es *asymmetrisches TSP*. Graphentheoretisch lässt sich das TSP wie folgt formulieren. Gegeben sei ein vollständiger Graph (oder Digraph)  $G$  mit Kantengewichten (oder Bogengewichten), gesucht ist ein (gerichteter) hamiltonscher Kreis minimaler Länge. Beim TSP geht man durch jeden Knoten genau einmal, beim (gerichteten) *Chinesischen Postbotenproblem (Chinese postman problem)* durch jede Kante (jeden Bogen) mindestens einmal, d. h. in einem Graphen (Digraphen) mit Kantengewichten (Bogen-

### 3.3 Graphentheoretische Optimierungsprobleme: Beispiele

gewichten) wird eine Kette (gerichtete Kette) gesucht, die jede Kante (jeden Bogen) mindestens einmal enthält und minimale Länge hat.

Zu diesen beiden Standardproblemen gibt es hunderte von Mischungen und Varianten. Z. B., man sucht eine Kette, die durch einige vorgegebene Knoten und Kanten mindestens einmal geht und minimale Länge hat; man legt verschiedene Ausgangspunkte (oder Depots) fest, zu denen man nach einer gewissen Streckenlänge wieder zurückkehren muss, etc. Eine relativ allgemeine Formulierung ist die folgende. Gegeben ist ein gemischter Graph mit Knotenmenge  $V$ , Kantenmenge  $E$  und Bogenmenge  $A$ . Ferner sind eine Menge von Depots  $W \subseteq V$ , von denen aus Reisen gestartet werden müssen, eine Menge  $U \subseteq V$  von Knoten, die mindestens einmal besucht werden müssen, und eine Menge  $B \subseteq E \cup A$  von Kanten und Bögen, die mindestens einmal durchlaufen werden müssen. Gesucht sind geschlossene Ketten von Kanten und gleichgerichteten Bögen, so dass jede dieser Folgen mindestens (oder genau) einen der Knoten aus  $W$  enthält und die Vereinigung dieser Ketten jeden Knoten aus  $U$  und jede Kante (Bogen) aus  $B$  mindestens einmal enthält und minimale Länge hat.  $\triangle$

Anwendungen dieser Probleme in der Routenplanung von Lieferwagen, von Straßenkehrmaschinen, der Müllabfuhr, von Speditionen etc. sind offensichtlich. Aber auch bei der Steuerung von NC-Maschinen (zum automatischen Bohren, Lötten oder Schweißen) oder der Verdrahtung von Leiterplatten (z. B. von Testbussen) tritt das TSP (oder eine seiner Varianten) auf. Abbildung 3.6 zeigt eine Leiterplatte, durch die 441 Löcher gebohrt werden müssen. Links unten ist der Startpunkt, an den der Bohrkopf nach Beendigung des Arbeitsvorganges zurückkehrt, damit eine neue Platte in die Maschine eingelegt werden kann. Abbildung 3.6 zeigt eine optimale Lösung dieses 442-Städte-TSP. Die Bohrmaschine muss eine Weglänge von 50.069 Einheiten zurückzulegen.

Abbildung 3.7 zeigt 666 Städte auf der Weltkugel. Wählt man die „Luftliniendistanz“ (bezüglich eines Großkreises auf der Kugel) als Entfernung zwischen zwei Städten, so zeigt Abbildung 3.7 eine kürzeste Rundreise durch die 666 Orte dieser Welt. Die Länge dieser Reise ist 294.358 km lang. Abbildungen 3.6 und 3.7 sind Grötschel and Holland (1991) entnommen. Im Internet finden Sie unter der URL

<http://www.math.uwaterloo.ca/tsp>

interessante Informationen zum TSP sowie weitere Bilder von TSP-Beispielen. Daten zu vielen TSP-Beispielen wurden von G. Reinelt gesammelt und sind unter der folgenden URL zu finden:

<http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/>

Der beste zur Zeit verfügbare Code zur Lösung von TSPs ist zu finden unter

<http://www.math.uwaterloo.ca/tsp/concorde.html>

In Abbildung 3.8 sind die eingezeichneten Punkte Standorte von Telefonzellen in der holländischen Stadt Haarlem. Der Stern in der Mitte ist das Postamt. Die Aufgabe ist

### 3 Diskrete Optimierungsprobleme

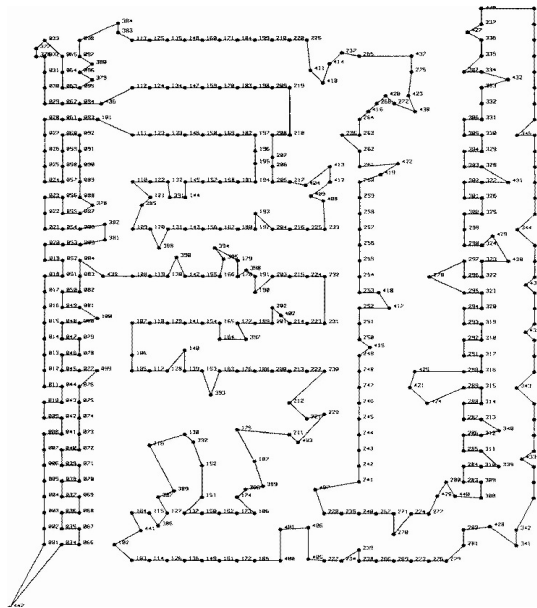


Abbildung 3.6: Optimaler Weg auf einer Leiterplatte um 441 Löcher zu bohren.

hier, eine Routenplanung für den sich wöchentlich wiederholenden Telefonzellenwartungsdienst zu machen. Die einzelnen Touren starten und enden am Postamt (diese Verbindungen sind nicht eingezeichnet) und führen dann so zu einer Anzahl von Telefonzellen, dass die Wartung aller Telefonzellen auf der Tour innerhalb einer Schicht durchgeführt werden kann.

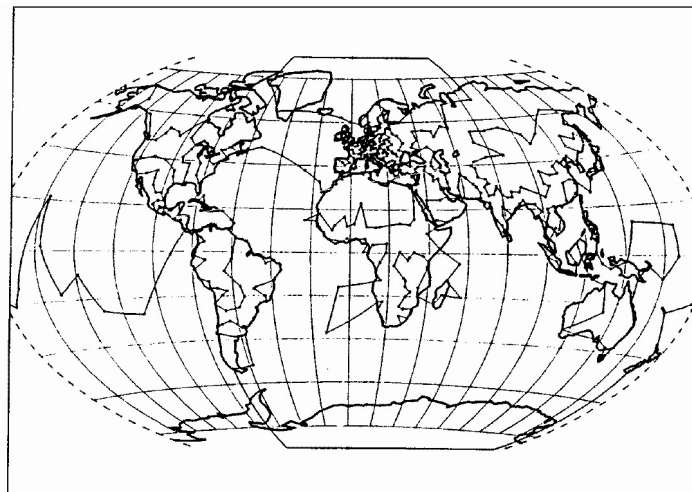


Abbildung 3.7: Optimale Tour für 666 Städte weltweit.

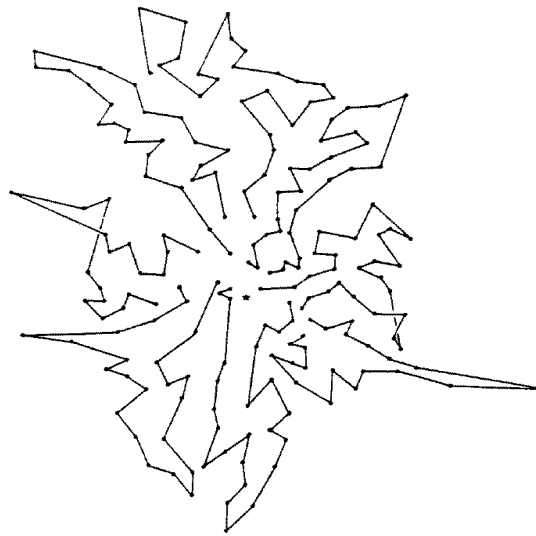


Abbildung 3.8: Optimale Routen für die Wartung der Telefonzellen in der holländischen Stadt Haarlem.

Als Travelling-Salesman-Problem lassen sich auch die folgenden Anwendungsprobleme formulieren:

- Bestimmung einer optimalen Durchlaufreihenfolge der Flüssigkeiten (Chargen) in einer Mehrproduktenpipeline (Minimierung Reinigungszeiten),
- Bestimmung der optimalen Verarbeitungsfolge von Lacken in einer Großlackiererei (Minimierung Reinigungszeiten),
- Bestimmung einer Reihenfolge des Walzens von Profilen in einem Walzwerk, so dass die Umrüstzeiten der Walzstraße minimiert werden,
- Bestimmung der zeitlichen Reihenfolge von archäologischen Fundstätten (Grablegungsreihenfolge von Gräbern in einem Gräberfeld, Besiedlungsreihenfolge von Orten) aufgrund von Ähnlichkeitsmaßen (Distanzen), die durch die aufgefundenen Fundstücke definiert werden, siehe hierzu Gertzen and Grötschel (2012).

Umfangreiche Information über das TSP, seine Varianten und Anwendungen kann man in den Sammelbänden Lawler et al. (1985) und Gutin and Punnen (2002) finden.

**(3.13) Stabile Mengen, Cliques, Knotenüberdeckungen.** Gegeben sei ein Graph  $G = (V, E)$  mit Knotengewichten  $c_v \in \mathbb{R}$  für alle  $v \in V$ . Das *Stabile-Menge-Problem* ist die Aufgabe, eine stabile Menge  $S \subseteq V$  zu suchen, so dass  $c(S)$  maximal ist, das *Cliqueproblem* die Aufgabe, eine Clique  $Q \subseteq V$  zu suchen, so dass  $c(Q)$  maximal ist,

### 3 Diskrete Optimierungsprobleme

und das *Knotenüberdeckungsproblem* die Aufgabe, eine Knotenüberdeckung  $K \subseteq V$  zu suchen, so dass  $c(K)$  minimal ist.  $\triangle$

Die drei oben aufgeführten Probleme sind auf triviale Weise ineinander überföhrbar. Ist nämlich  $S \subseteq V$  eine stabile Menge in  $G$ , so ist  $S$  eine Clique im komplementären Graphen  $\overline{G}$  von  $G$  und umgekehrt. Also ist das Stabile-Menge-Problem für  $G$  mit Gewichtsfunktion  $c$  nicht anders als das Cliquesproblem für  $\overline{G}$  mit derselben Gewichtsfunktion und umgekehrt. Ist ferner  $S \subseteq V$  eine stabile Menge in  $G$ , so ist  $V \setminus S$  eine Knotenüberdeckung von  $G$ . Daraus folgt, dass zu jeder gewichtsmaximalen stabilen Menge  $S$  die zugehörige Knotenüberdeckung  $V \setminus S$  gewichtsm minimal ist und umgekehrt. Das Stabile-Menge-Problem, das Cliquesproblem und das Knotenüberdeckungsproblem sind also drei verschiedene Formulierungen einer Aufgabe. Anwendungen dieser Probleme finden sich z. B. in folgenden Bereichen:

- Einsatzplanung von Flugzeugbesatzungen
- Busfahrereinsatzplanung
- Tourenplanung im Behindertentransport
- Auslegung von Fließbändern
- Investitionsplanung
- Zuordnung von Wirtschaftsprüfern zu Prüffeldern
- Entwurf von optimalen fehlerkorrigierenden Codes
- Schaltkreisentwurf
- Standortplanung
- Wiedergewinnung von Information aus Datenbanken
- Versuchsplanung
- Signalübertragung.

Aber auch das folgende Schachproblem kann als Stabile-Menge-Problem formuliert werden: Bestimme die maximale Anzahl von Damen (oder Türmen, oder Pferden etc.), die auf einem  $n \times n$  Schachbrett so plaziert werden können, dass keine eine andere schlägt.

**(3.14) Färbungsprobleme.** Gegeben sei ein Graph  $G = (V, E)$ . Zusätzlich seien Knotengewichte  $b_v$  für alle  $v \in V$  gegeben. Die Aufgabe, eine Folge von (nicht notwendigerweise verschiedenen) stabilen Mengen  $S_1, \dots, S_t$  von  $G$  zu suchen, so dass jeder Knoten in mindestens  $b_v$  dieser stabilen Mengen enthalten und  $t$  minimal ist, heißt (gewichtetes) *Knotenfärbungsproblem* oder kurz *Färbungsproblem*. Beim (gewichteten) *Kantenfärbungsproblem* sind statt Knotengewichten Kantengewichte  $c_e, e \in E$ , gegeben und gesucht ist eine Folge von (nicht notwendigerweise verschiedenen) Matchings  $M_1, \dots, M_s$ , so dass jede Kante in mindestens  $c_e$  dieser Matchings enthalten und  $s$  so klein wie möglich ist.  $\triangle$



Das geographische Färbungsproblem ist uns schon in (3.6) begegnet. Hat man eine Färbung der Länder, so dass je zwei benachbarte Länder verschieden gefärbt sind, so entspricht jede Gruppe von Ländern gleicher Farbe einer stabilen Menge in  $G$ . Hat man umgekehrt eine Zerlegung der Knotenmenge von  $G$  in stabile Mengen, so kann man jeweils die Länder, die zu den Knoten einer stabilen Menge gehören mit derselben Farbe belegen und erhält dadurch eine zulässige Landkartenfärbung. Das Landkartenfärbungsproblem ist also das Knotenfärbungsproblem des zugehörigen Graphen mit  $b_v = 1$  für alle  $v \in V$ .

Die Aufgabe, in einer geographischen Region die Sendefrequenzen von Rundfunksendern (oder Mobilfunkantennen) so zu verteilen, dass sich die Sender gegenseitig nicht stören und alle Rundfunkteilnehmer die für sie gedachten Programme auch empfangen können, kann man als Färbungsproblem (mit weiteren Nebenbedingungen) formulieren.

**(3.15) Schnittprobleme.** Gegeben sei ein Graph  $G = (V, E)$  mit Kantengewichten  $c_e \in \mathbb{R}$  für alle  $e \in E$ . Das Problem, einen Schnitt  $\delta(W)$  in  $G$  zu finden mit maximalem Gewicht  $c(\delta(W))$ , heißt *Max-Cut-Problem*. Sind alle Kantengewichte  $c_e$  nicht-negativ, so nennt man das Problem, einen Schnitt minimalen Gewichts in  $G$  zu finden, *Min-Cut-Problem*. △

Das Min-Cut-Problem ist in der Theorie der Netzwerkflüsse sehr wichtig (siehe Kapitel 6).

Das Max-Cut-Problem hat z. B. eine interessante Anwendung in der Physik, und zwar kann man beim Studium magnetischer Eigenschaften von Spingläsern im Rahmen des Ising Modells die Aufgabe, einen Grundzustand (energieminimale Konfiguration bei  $0^\circ$  K) zu bestimmen, als Max-Cut-Problem formulieren. Ich will diese Anwendung kurz skizzieren.

Ein *Spinglas* besteht aus nichtmagnetischem Material, das an einigen Stellen durch magnetische Atome „verunreinigt“ ist. Man interessiert sich für die Energie des Systems und die Orientierung der magnetischen Atome (Verunreinigungen) bei  $0^\circ$  K, also für den so genannten (gefrorenen) Grundzustand des Spinglases. Dieser Grundzustand ist experimentell nicht herstellbar, und die Physiker haben unterschiedliche, sich z. T. widersprechende Theorien über einige Eigenschaften dieses Grundzustandes.

Mathematisch wird dieses Problem wie folgt modelliert. Jeder Verunreinigung  $i$  wird ein Vektor  $S_i \in \mathbb{R}^3$  zugeordnet, der (bei einem gegebenen Bezugssystem) die Orientierung des Atomes im Raum, d. h. den magnetischen Spin, beschreibt. Zwischen zwei Verunreinigungen  $i, j$  besteht eine magnetische Interaktion, die durch

$$H_{ij} = J(r_{ij}) S_i \cdot S_j$$

beschrieben wird, wobei  $J(r_{ij})$  eine Funktion ist, die vom Abstand  $r_{ij}$  der Verunreinigungen abhängt, und  $S_i \cdot S_j$  das innere Produkt der Vektoren  $S_i, S_j$  ist. In der Praxis wird  $J$  (bei gewissen physikalischen Modellen) wie folgt bestimmt:

$$J(r_{ij}) := \cos(Kr_{ij})/r_{ij}^3,$$

wobei  $K$  eine materialabhängige Konstante ist (z. B.  $K = 2.4 \cdot 10^8$ ). Die gesamte Energie einer Spinkonfiguration ist gegeben durch

$$H = - \sum J(r_{ij}) S_i \cdot S_j + \sum F \cdot S_i,$$

### 3 Diskrete Optimierungsprobleme

wobei  $F$  ein äußeres magnetisches Feld ist. (Der Einfachheit halber nehmen wir im folgenden an  $F = 0$ .) Ein Zustand minimaler Energie ist also dadurch charakterisiert, dass  $\sum J(r_{ij}) S_i \cdot S_j$  maximal ist.

Das hierdurch gegebene Maximierungsproblem ist mathematisch kaum behandelbar. Von Ising wurde folgende Vereinfachung vorgeschlagen. Statt jeder beliebigen räumlichen Orientierung werden jeder Verunreinigung nur zwei Orientierungen erlaubt: „Nordpol oben“ oder „Nordpol unten“. Die dreidimensionalen Vektoren  $S_i$  werden dann in diesem Modell durch Variablen  $s_i$  mit Werten in der zweielementigen Menge  $\{1, -1\}$  ersetzt. Unter Physikern besteht Übereinstimmung darüber, dass dieses Ising-Modell das wahre Verhalten gewisser Spingläser gut widerspiegelt. Das obige Maximierungsproblem lautet dann bezüglich des Ising Modells:

$$\max \left\{ \sum J(r_{ij}) s_i s_j \mid s_i \in \{-1, 1\} \right\}.$$

Nach dieser durch die Fachwissenschaftler vorgenommenen Vereinfachung ist der Schritt zum Max-Cut-Problem leicht. Wir definieren einen Graphen  $G = (V, E)$ , wobei jeder Knoten aus  $V$  eine Verunreinigung repräsentiert, je zwei Knoten  $i, j$  sind durch eine Kante verbunden, die das Gewicht  $c_{ij} = -J(r_{ij})$  trägt. (Ist  $r_{ij}$  groß, so ist nach Definition  $c_{ij}$  sehr klein, und üblicherweise werden Kanten mit kleinen Gewichten  $c_{ij}$  gar nicht berücksichtigt). Eine Partition von  $V$  in  $V_1$  und  $V_2$  entspricht einer Orientierungsfestlegung der Variablen, z. B.  $V_1 := \{i \in V \mid i \text{ repräsentiert eine Verunreinigung mit Nordpol oben}\}$ ,  $V_2 := \{i \in V \mid \text{der Nordpol von } i \text{ ist unten}\}$ . Bei gegebenen Orientierungen der Atome (Partition  $V_1, V_2$  von  $V$ ) ist die Energie des Spinglaszustandes also wie folgt definiert:

$$\sum_{i \in V_1, j \in V_2} c_{ij} - \sum_{i, j \in V_1} c_{ij} - \sum_{i, j \in V_2} c_{ij}.$$

Der Zustand minimaler Energie kann also durch Maximierung des obigen Ausdrucks bestimmt werden. Addieren wir zu diesem Ausdruck die Konstante  $C := \sum_{i, j \in V} c_{ij}$ , so folgt daraus, dass der Grundzustand eines Spinglases durch die Lösung des Max-Cut Problems

$$\max \left\{ \sum_{i \in V_1} \sum_{j \in V_2} c_{ij} \mid V_1, V_2 \text{ Partition von } V \right\}$$

bestimmt werden kann. Eine genauere Darstellung und die konkrete Berechnung von Grundzuständen von Spingläsern (und weitere Literaturhinweise) kann man in Barahona et al. (1988) finden. Dieses Paper beschreibt auch eine Anwendung des Max-Cut-Problems im VLSI-Design und bei der Leiterplattenherstellung: Die Lagenzuweisung von Leiterbahnen, so dass die Anzahl der Kontaktlöcher minimal ist.

**(3.16) Standortprobleme.** Probleme dieses Typs tauchen in der englischsprachigen Literatur z. B. unter den Namen Location oder Allocation Problems, Layout Planning, Facilities Allocation, Plant Layout Problems oder Facilities Design auf. Ihre Vielfalt ist (ähnlich wie bei (3.12)) kaum in wenigen Zeilen darstellbar. Ein (relativ allgemeiner) Standardtyp ist der folgende. Gegeben sei ein Graph (oder Digraph), dessen Knoten Städte, Wohnbezirke, Bauplätze, mögliche Fabrikationsstätten etc. repräsentieren, und

dessen Kanten Verkehrsverbindungen, Straßen, Kommunikations- oder Transportmöglichkeiten etc. darstellen. Die Kanten besitzen „Gewichte“, die z. B. Entfernungen etc. ausdrücken. Wo sollen ein Krankenhaus, ein Flughafen, mehrere Polizei- oder Feuerwehrstationen, Warenhäuser, Anlieferungslager, Fabrikationshallen, ... errichtet werden, so dass ein „Optimalitätskriterium“ erfüllt ist? Hierbei tauchen häufig Zielfunktionen auf, die nicht linear sind. Z. B. soll ein Feuerwehrdepot so stationiert werden, dass die maximale Entfernung vom Depot zu allen Wohnbezirken minimal ist; drei Auslieferungslager sollen so errichtet werden, dass jedes Lager ein Drittel der Kunden bedienen kann und die Summe der Entfernungen der Lager zu ihren Kunden minimal ist bzw. die maximale Entfernung minimal ist.  $\triangle$

**(3.17) Lineare Anordnungen und azyklische Subdigraphen.** Gegeben sei ein vollständiger Digraph  $D_n = (V, A)$  mit Bogengewichten  $c((i, j))$  für alle  $(i, j) \in A$ . Das Problem, eine lineare Reihenfolge der Knoten, sagen wir  $i_1, \dots, i_n$ , zu bestimmen, so dass die Summe der Gewichte der Bögen, die konsistent mit der linearen Ordnung sind (also  $\sum_{p=1}^{n-1} \sum_{q=p+1}^n c((i_p, i_q))$ ), maximal ist, heißt *Linear-Ordering-Problem*. Das *Azyklische-Subdigraphen-Problem* ist die Aufgabe, in einem Digraphen  $D = (V, A)$  mit Bogengewichten eine Bogenmenge  $B \subseteq A$  zu finden, die keinen gerichteten Kreis enthält und deren Gewicht maximal ist. Beim *Feedback-Arc-Set-Problem* sucht man eine Bogenmenge minimalen Gewichts, deren Entfernung aus dem Digraphen alle gerichteten Kreise zerstört.  $\triangle$

Die drei in (3.17) genannten Probleme sind auf einfache Weise ineinander transformierbar. Diese Probleme haben interessante Anwendungen z. B. bei der Triangulation von Input-Output-Matrizen, der Rangbestimmung in Turniersportarten, im Marketing und der Psychologie. Weitergehende Informationen finden sich in Grötschel et al. (1984) und in Reinelt (1985). Einige konkrete Anwendungsbeispiele werden in den Übungen behandelt.

**(3.18) Entwurf kostengünstiger und ausfallsicherer Telekommunikationsnetzwerke.** Weltweit wurden in den letzten Jahren (und das geschieht weiterhin) die Kupferkabel, die Telefongespräche etc. übertragen, durch Glasfaserkabel ersetzt. Da Glasfaserkabel enorm hohe Übertragungskapazitäten haben, wurden anfangs die stark „vermaschten“ Kupferkabelnetzwerke durch Glasfasernetzwerke mit Baumstruktur ersetzt. Diese Netzwerkstrukturen haben jedoch den Nachteil, dass beim Ausfall eines Verbindungskabels (z. B. bei Baggerarbeiten) oder eines Netzknotens (z. B. durch einen Brand) große Netzteile nicht mehr miteinander kommunizieren können. Man ist daher dazu übergegangen, Telekommunikationsnetzwerke mit höherer Ausfallsicherheit wie folgt auszulegen. Zunächst wird ein Graph  $G = (V, E)$  bestimmt; hierbei repräsentiert  $V$  die Knotenpunkte, die in einem Telekommunikationsnetz verknüpft werden sollen, und  $E$  stellt die Verbindungen zwischen Knoten dar, die durch das Ziehen eines direkten (Glasfaser-) Verbindungskabels realisiert werden können. Gleichzeitig wird geschätzt, was das Legen einer direkten Kabelverbindung kostet. Anschließend wird festgelegt, welche Sicherheitsanforderungen das Netz erfüllen soll. Dies wird so gemacht, dass man für je zwei Knoten bestimmt, ob das Netz noch eine Verbindung zwischen diesen beiden Knoten besitzen soll, wenn ein, zwei,

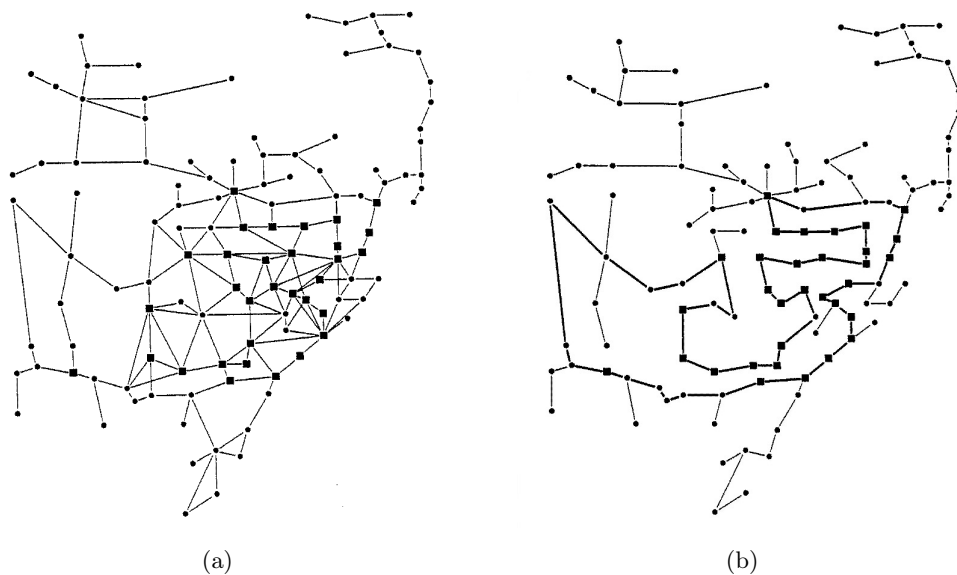


Abbildung 3.9: Abbildung (a) stellt ein Telekommunikationsnetz dar. In Abbildung (b) ist eine Optimallösung des kostengünstigsten Telekommunikationsnetzes in diesem Graphen zu sehen.

drei, ... Kanten oder einige andere Knoten ausfallen. Dann wird ein Netzwerk bestimmt, also eine Teilmenge  $F$  von  $E$ , so dass alle Knoten miteinander kommunizieren können, alle Sicherheitsanforderungen erfüllt werden und die Baukosten minimal sind.  $\triangle$

Mit Hilfe dieses Modells (und zu seiner Lösung entwickelter Algorithmen) werden z. B. in den USA Glasfasernetzwerke für so genannte LATA-Netze entworfen und ausgelegt, siehe Grötschel et al. (1992) und Grötschel et al. (1995). Abbildung 3.9(a) zeigt das Netzwerk der möglichen direkten Kabelverbindungen in einer großen Stadt in den USA, Abbildung 3.9(b) zeigt eine optimale Lösung. Hierbei sind je zwei durch ein Quadrat gekennzeichnete Knoten gegen den Ausfall eines beliebigen Kabels geschützt (d. h. falls ein Kabel durchschnitten wird, gibt es noch eine (nicht notwendig direkte, sondern auch über Zwischenknoten verlaufende) Verbindung zwischen je zwei dieser Knoten, alle übrigen Knotenpaare wurden als relativ unwichtig erachtet und mussten nicht gegen Kabelausfälle geschützt werden.

Die gerade genannten, in den 1990er Jahren behandelten LATA-Netzwerke sind noch relativ einfach unter Weglassung vieler Nebenbedingungen mathematisch modelliert worden. Im Laufe der Jahre sind durch technische Entwicklungen und immer komplizierter werdende Anforderungen wesentlich komplexere mathematische Modelle zum Entwurf von Kommunikationsnetzwerken entwickelt und in der Praxis eingesetzt worden. In dem Aufsatz von Andreas Bley und Thorsten Koch *Optimierung des G-WiN: Optimierung in der Planung und beim Aufbau des Gigabit-Wissenschaftsnetzes*, siehe [https://www.dfn.de/fileadmin/5Presse/DFNMitteilungen/mitteilungen\\_bis\\_66/heft54.pdf](https://www.dfn.de/fileadmin/5Presse/DFNMitteilungen/mitteilungen_bis_66/heft54.pdf) aus dem

Jahr 2000 wird z.B. (für den Laien gut verständlich) dargestellt, welche Überlegungen in die Optimierung des G-WiN-Netzes des DFN-Vereins eingeflossen sind, das seinerzeit zur Versorgung des Kommunikationsbedarfs von rund 750 deutschen Wissenschaftseinrichtungen diente. Standort- und Netzplanung flossen hier zusammen. Nach ungefähr fünf Jahren ist das G-WiN-Netz durch das X-WiN-Netz abgelöst worden, eine Kurzbeschreibung hierzu findet sich in Andreas Bley und Marcus Pattloch, *Modellierung und Optimierung der X-WiN Plattform*, DFN-Mitteilungen, 4-7(2005), siehe <https://www.dfn.de/fileadmin/5Presse/DFNMitteilungen/heft67.pdf>. Die Neukonfigurationen gehen in Abständen von ungefähr fünf Jahren weiter, und das ist nicht nur beim DFN-Verein so, sondern auch bei den meisten anderen Kommunikationsnetzen.

Glasfaserkabel werden inzwischen nicht nur für sogenannte „Backbone-Netze“ (wie das X-WiN) verwendet. Die stark steigende Nutzung von hochvolumigen Diensten (Video Streaming und Conferencing, etc.) erfordert zunehmend den breitbandigen Anschluss von Industriestandorten, öffentlicher Verwaltung und privater Haushalte an das weltweite Kommunikationsnetz. Glasfaserbasierte Zugangsnetze wie Fiber-to-the-Curb (FTTC), Fiber-to-the-Building (FTTB) und Fiber-to-the-Home (FTTH) (gemeinsam kurz mit FTTx bezeichnet) können diese Bandbreiten bereitstellen. Der Aufbau derartiger Zugangsnetze ist allerdings mit sehr hohen Investitionen verbunden und erfordert eine an die jeweiligen Randbedingungen angepasste strategische Planung. Hierbei muss eine Vielzahl von Faktoren, wie die bereits existierende Infrastruktur, die verfügbare Systemtechnik, die zu erwartenden Nutzer und Dienste, die örtlichen Gegebenheiten und nicht zuletzt die Kosten für den Aufbau und den Betrieb des Netzes in Betracht gezogen werden. In zwei Projekten (FTTx-PLAN, 2009-2011 und Multi-FTTx, 2012-2015) wurden am Zuse-Institut verschiedene Planungstools zum optimalen Entwurf derartiger Netzwerke entwickelt. Weitere Informationen finden sich auf den Projektwebseiten <http://www.zib.de/projects/tools-planning-fttx-networks> und <http://www.zib.de/projects/multicriteria-optimization-fttx-networks>.

Ein Überblick über die Anwendung von Optimierung und auch anderer Zweige der Mathematik in der Telekommunikation findet sich in dem Artikel Aurzada et al. (2014) im MATHEON-Buch.

**(3.19) Betrieb von Gasnetzen.** Gasnetze sind ein wichtiger Teil der europäischen Gasversorgung. Durch die physikalischen Eigenschaften von Gasen sowie die zur Gasbeförderung eingesetzte Technik unterscheiden sich adäquate mathematische Modelle deutlich von denen für z.B. Telekommunikationsnetze. Insbesondere sind nicht nur (Gas-)Flüsse relevant, sondern auch die Drücke des Gases an verschiedenen Stellen des Netzes. In Rohren verringert sich der Druck des Gases wegen der Reibung am Rohr. Daher muss in bestimmten Abständen der Druck durch sogenannte Kompressoren wieder erhöht werden. Aus technischen Gründen gibt es außerdem Druckregler, die den Druck reduzieren, sowie Ventile, um den Gasfluss zu steuern.

Ein Gasnetz kann (stark vereinfacht) modelliert werden als ein Digraph  $G = (V, A)$ , wobei jeder Bogen  $a \in A$  ein Element (Rohr, Ventil, Druckregler, Kompressor) des Gasnetzes repräsentiert. Der Gasfluss über einen Bogen  $a \in A$  sei  $x_a \in \mathbb{R}$ , der Druck an einem Knoten  $v \in V$  sei  $p_v \in \mathbb{R}_+$ , und die Ein- bzw. Ausspeisung in  $v$  sei  $d_v \in \mathbb{R}$ . An

## Literaturverzeichnis

jedem Knoten  $v \in V$  gilt die *Flusserhaltungsgleichung*

$$d_v + \sum_{a \in \delta^-(v)} x_a - \sum_{a \in \delta^+(v)} x_a = 0,$$

d. h. Gas, das in einen Knoten hineinströmt, verlässt ihn auch wieder.

Für jedes Element  $a = (u, v) \in A$  gilt, je nach Elementtyp, eine Beziehung zwischen Fluss und Druck. Für ein Rohr  $a = (u, v)$  gilt approximativ der Zusammenhang

$$p_v^2 - p_u^2 = \alpha_a x_a |x_a|,$$

wobei die Konstante  $\alpha_a$  die physikalischen Eigenschaften des Rohres abbildet. Ein Regler  $a = (u, v)$  kann in seiner Betriebsrichtung den Druck in gewissen Grenzen unabhängig vom Fluss verringern, d. h. es gilt

$$p_v \geq p_u - \Delta_a$$

mit maximaler Druckverringern  $\Delta_a$ , falls  $x_a > 0$  und  $p_v = p_u$  sonst. Analog kann ein Verdichter den Druck in Betriebsrichtung erhöhen. Ein Ventil  $a = (u, v)$  erlaubt es, einzelne Komponenten zu sperren und damit Druck und Fluss vor und hinter dem Ventil zu entkoppeln, d. h. es gilt

$$\begin{aligned} \text{Ventil } a \text{ geschlossen: } & x_a = 0, p_u, p_v \text{ beliebig,} \\ \text{Ventil } a \text{ geöffnet: } & x_a \text{ beliebig, } p_u = p_v. \end{aligned}$$

Eine interessante Frage ist zum Beispiel, ob der Ein- und Ausspeisevektor  $d = (d_v)_{v \in V}$  im Gasnetz realisiert werden kann. Durch die (nichtlineare) Abhängigkeit des Gasflusses von den Drücken und der Möglichkeit, das Verhalten von Elementen des Netzes zwischen sehr verschiedenen Zuständen umzuschalten, ist diese Frage viel schwerer zu beantworten als für Netzwerkflussprobleme ohne Druck-Fluss-Koppelung, die wir später betrachten werden.  $\triangle$

## Literaturverzeichnis

- M. Aigner. *Graphentheorie: Eine Entwicklung aus dem 4-Farben-Problem*. Teubner Verlag, Studienbücher: Mathematik, Stuttgart, 1984. ISBN 3-519-02068-8.
- D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook. *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2006.
- F. Aurzada, A. Bley, A. Eisenblätter, H.-F. Geerdes, M. Guillemand, G. Kutyniok, F. Philipp, C. Raack, M. Scheutzwow, and A. Werner. Mathematics for telecommunications. In P. Deuffhard, M. Grötschel, D. Hömberg, U. Horst, J. Kramer, V. Mehrmann, K. Polthier, F. Schmidt, C. Schütte, M. Skutella, and J. Sprekels, editors, *MATHEON – Mathematics for Key Technologies*, volume 1 of *EMS Series in Industrial and Applied Mathematics*, pages 75–89. European Mathematical Society, 2014. doi: 10.4171/137.

- F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt. An application of combinatorial optimization to statistical physics and circuit layout design. *Operations Research*, 36(3):493–513, 1988.
- Berge and Ghouila-Houri. *Programme, Spiele, Transportnetze*. Teubner Verlag, Leipzig, 1969.
- C. Berge. *Hypergraphs, Combinatorics of finite sets*, volume 45. North-Holland Mathematical Library, Amsterdam, 1989.
- B. Bollobás. *Graph Theory: An Introductory Course*. Springer Verlag, New York, 1979.
- J. A. Bondy and U. S. R. Murty. *Graph Theory*. Springer, Berlin, 2008.
- W. J. Cook. *In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation*. Princeton University Press, 2012.
- W. Domschke. *Logistik: Rundreisen und Touren*. Oldenbourg-Verlag, München – Wien, 4., erweiterte Auflage, 1997.
- J. Ebert. Effiziente Graphenalgorithmen. *Studentexte: Informatik*, 1981.
- T. L. Gertzen and M. Grötschel. Flinders Petrie, the Travelling Salesman Problem, and the beginning of mathematical modeling in archaeology. In *Optimization Stories*, Documenta Mathematica, pages 199–210. DMV, 2012. Elektronisch verfügbar unter <http://www.zib.de/groetschel/pubnew/paper/gertzgroe2012.pdf>.
- M. C. Golombic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York. ISBN 1980.
- R. L. Graham, M. Grötschel, and L. Lovász, editors. *Handbook of Combinatorics, Volume I*. Elsevier (North-Holland); The MIT Press, Cambridge, Massachusetts, 1995. ISBN 0-444-82346-8/v.1 (Elsevier); ISBN 0-262-07170-3/v.1 (MIT).
- M. Grötschel and O. Holland. Solution of large-scale symmetric travelling salesman problems. *Mathematical Programming, Series A*, 51(2):141–202, 1991.
- M. Grötschel and Y. Yuan. Euler, Mei-Ko Kwan, Königsberg, and a Chinese Postman. In *Optimization Stories*, Documenta Mathematica, pages 43–50. DMV, 2012. Elektronisch verfügbar unter <http://www.zib.de/groetschel/pubnew/paper/groeyuan2012.pdf>.
- M. Grötschel, M. Jünger, and G. Reinelt. A Cutting Plane Algorithm for the Linear Ordering Problem. *Operations Research*, 32(6):1195–1220, 1984.
- M. Grötschel, C. L. Monma, and M. Stoer. Computational Results with a Cutting Plane Algorithm for Designing Communication Networks with Low-Connectivity Constraints. *Operations Research*, 40(2):309–330, 1992.

## Literaturverzeichnis

- M. Grötschel, C. L. Monma, and M. Stoer. Design of Survivable Networks. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, pages 617–672. North-Holland, 1995.
- G. Gutin and A. P. Punnen, editors. *The Traveling Salesman Problem and Its Variations*. Kluwer Academic Publishers, 2002.
- R. Halin. *Graphentheorie*. Akademie-Verlag Berlin, 2. Auflage, 1989.
- K. Hässig. *Graphentheoretische Methoden des Operations Research*. Teubner-Verlag, Stuttgart, 1979.
- D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig, 1936. Mehrfach auf deutsch und in englischer Übersetzung nachgedruckt.
- E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys. *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley, Chichester, 1985.
- T. Lengauer. *Combinatorial Algorithms for Integrated Circuit Layout*. Teubner, Stuttgart und Wiley, Chichester, 1990.
- J. K. Lenstra. *Sequencing by Enumerative Methods*. PhD thesis, Mathematisch Centrum, Amsterdam, 1976.
- J. G. Oxley. *Matroid Theory*. Oxford University Press, Oxford, 1992.
- G. Reinelt. *The Linear Ordering Problem: Algorithms and Applications*. Heldermann Verlag, Berlin, 1985.
- H. Sachs. *Einführung in die Theorie der endlichen Graphen*. Teubner, Leipzig, 1970, und Hanser, München, 1971, 1970.
- M. Stoer. Design of survivable networks. *Lecture Notes for Mathematics*, 1992.
- K. Wagner. *Graphentheorie*. BI Wissenschaftsverlag, Mannheim, 1970.
- H. Walther and G. Nägler. *Graphen, Algorithmen, Programme*. VEB Fachbuchverlag, Leipzig, 1987.



## 4 Komplexitätstheorie und Speicherung von Graphen

In diesem Kapitel führen wir einige der Grundbegriffe der Komplexitätstheorie ein, die für die algorithmische Graphentheorie und die kombinatorische Optimierung von Bedeutung sind. Wir behandeln insbesondere die Klassen  $\mathcal{P}$  und  $\mathcal{NP}$  und das Konzept der  $\mathcal{NP}$ -Vollständigkeit. Die Darstellung erfolgt auf informellem Niveau. Gute Bücher zur Einführung in die Komplexitätstheorie sind Garey and Johnson (1979), Papadimitriou (1994) und Wegener (2003), weiterführende Aspekte werden unter anderem in Wagner and Wechsung (1986) und van Leeuwen (1990) behandelt. Shmoys and Tardos (1995) ist ein guter Übersichtsartikel. Einen detaillierten Überblick über Komplexitätsklassen gibt Johnson (1990), noch mehr Komplexitätsklassen findet man unter der URL:

[http://complexityzoo.uwaterloo.ca/Complexity\\_Zoo](http://complexityzoo.uwaterloo.ca/Complexity_Zoo).

Komplexitätstheorie kann man nicht ohne Grundkenntnisse in Kodierungstechniken betreiben. Noch wichtiger sind Datenstrukturen beim Entwurf effizienter Algorithmen. Diesen Themenkreis streifen wir kurz in Abschnitt 4.3. Zur Vertiefung dieses Gebietes empfehlen wir Aho et al. (1974), Cormen et al. (2001), Mehlhorn (1984), Mehlhorn and Sanders (2008), Meinel (1991), Ottmann and Widmayer (2012) und Tarjan (1983). Ein Handbuch zu Datenstrukturen mit breiten Übersichtsartikeln ist Mehta and Sahni (2004).

### 4.1 Probleme, Komplexitätsmaße, Laufzeiten

In der Mathematik (und nicht nur hier) kann das Wort „Problem“ sehr verschiedene Bedeutungen haben. Für unsere Zwecke benötigen wir eine (einigermaßen) präzise Definition. Ein *Problem* ist eine allgemeine Fragestellung, bei der mehrere Parameter offen gelassen sind und für die eine Lösung oder Antwort gesucht wird.

Ein Problem ist dadurch definiert, dass alle seine Parameter beschrieben werden und dass genau angegeben wird, welche Eigenschaften eine Antwort (Lösung) haben soll. Sind alle Parameter eines Problems mit expliziten Daten belegt, dann spricht man im Englischen von „problem instance“. Im Deutschen hat sich hierfür bisher kein Standardbegriff ausgeprägt. Es sind u. a. die Wörter Einzelfall, Fallbeispiel, Problembeispiel, Probleminstanz, Problemausprägung oder kurz Instanz gebräuchlich. Ebenso wird auch das (allgemeine) Problem manchmal als Problemtyp oder Problemklasse bezeichnet. In diesem Skript werden wir keiner starren Regel folgen. Aus dem Kontext heraus dürfte i. A. klar sein, was gemeint ist.

Das Travelling-Salesman-Problem ist in diesem Sinne ein Problem. Seine offenen Parameter sind die Anzahl der Städte und die Entfernungen zwischen diesen Städten. Eine

#### 4 Komplexitätstheorie und Speicherung von Graphen

Entfernungstabelle in einem Autoatlas definiert ein konkretes Beispiel für das Travelling-Salesman-Problem.

Aus mathematischer Sicht kann man es sich einfach machen: Ein Problem (wir verwenden dafür häufig das Symbol  $\Pi$ ) ist die Menge aller Problembeispiele. Das Travelling-Salesman-Problem ist also die Menge aller TSP-Instanzen. Das ist natürlich nicht sonderlich tief Sinnig, vereinfacht aber die mathematische Notation.

Wir sagen, dass ein Algorithmus ein Problem  $\Pi$  *löst*, wenn er für jedes Problembeispiel  $\mathcal{I} \in \Pi$  eine Lösung findet. Das Ziel des Entwurfs von Algorithmen ist natürlich, möglichst „effiziente“ Verfahren zur Lösung von Problemen zu finden.

Um dieses Ziel mit Inhalt zu füllen, müssen wir den Begriff „Effizienz“ messbar machen. Mathematiker und Informatiker haben hierzu verschiedene *Komplexitätsmaße* definiert. Wir werden uns hier nur mit den beiden für uns wichtigsten Begriffen *Zeit-* und *Speicherkomplexität* beschäftigen. Hauptsächlich werden wir uns mit Zeitkomplexität befassen.

Es ist trivial, dass die Laufzeit eines Algorithmus abhängt von der „Größe“ eines Problembeispiels, d. h. vom Umfang der Eingabedaten. Bevor wir also Laufzeitanalysen anstellen können, müssen wir beschreiben, wie wir unsere Problembeispiele darstellen, bzw. kodieren wollen. Allgemein kann man das durch die Angabe von *Kodierungsschemata* bewerkstelligen. Da wir uns jedoch ausschließlich mit Problemen beschäftigen, die mathematisch darstellbare Strukturen haben, reicht es für unsere Zwecke aus, Kodierungsvorschriften für die uns interessierenden Strukturen anzugeben. Natürlich gibt es für jede Struktur beliebig viele Kodierungsmöglichkeiten. Wir werden die geläufigsten benutzen und merken an, dass auf diesen oder dazu (in einem spezifizierbaren Sinn) äquivalenten Kodierungsvorschriften die gesamte derzeitige Komplexitätstheorie aufbaut.

Ganze Zahlen kodieren wir *binär*. Die binäre Darstellung einer nicht-negativen ganzen Zahl  $n$  benötigt  $\lceil \log_2(|n|+1) \rceil$  Bits (oder Zellen). Hinzu kommt ein Bit für das Vorzeichen. Die *Kodierungslänge*  $\langle n \rangle$  einer ganzen Zahl  $n$  ist die Anzahl der zu ihrer Darstellung notwendigen Bits, d. h.

$$\langle n \rangle := \lceil \log_2(|n| + 1) \rceil + 1. \quad (4.1)$$

Jede rationale Zahl  $r$  hat eine Darstellung  $r = \frac{p}{q}$  mit  $p, q \in \mathbb{Z}$ ,  $p$  und  $q$  teilerfremd und  $q > 0$ . Wir nehmen an, dass jede rationale Zahl so dargestellt ist, und können daher sagen, dass die Kodierungslänge von  $r = \frac{p}{q}$  gegeben ist durch

$$\langle r \rangle := \langle p \rangle + \langle q \rangle.$$

Wir werden im Weiteren auch sagen, dass wir eine ganze oder rationale Zahl  $r$  in einem *Speicherplatz* (oder *Register*) speichern, und wir gehen davon aus, dass der Speicherplatz für  $r$  die benötigten  $\langle r \rangle$  Zellen besitzt.

Die Kodierungslänge eines Vektors  $x = (x_1, \dots, x_n)^T \in \mathbb{Q}^n$  ist

$$\langle x \rangle := \sum_{i=1}^n \langle x_i \rangle,$$

und die Kodierungslänge einer Matrix  $A \in \mathbb{Q}^{(m,n)}$  ist

$$\langle A \rangle := \sum_{i=1}^m \sum_{j=1}^n \langle a_{ij} \rangle.$$

Für diese Vorlesung besonders wichtig sind die Datenstrukturen zur Kodierung von Graphen und Digraphen. Auf diese werden wir in Abschnitt 4.3 genauer eingehen.

Sind alle Kodierungsvorschriften festgelegt, so müssen wir ein *Rechnermodell* entwerfen, auf dem unsere Speicher- und Laufzeitberechnungen durchgeführt werden sollen. In der Komplexitätstheorie benutzt man hierzu i. A. *Turing-Maschinen* oder *RAM-Maschinen*. Wir wollen auf diese Rechnermodelle nicht genauer eingehen. Wir nehmen an, dass der Leser weiß, was Computer sind und wie sie funktionieren, und unterstellen einfach, dass jeder eine naive Vorstellung von einer „vernünftigen“ Rechenmaschine hat. Dies reicht für unsere Zwecke aus.

Wir stellen uns den Ablauf eines Algorithmus  $A$  (der in Form eines Rechnerprogramms vorliegt) auf einer Rechenanlage wie folgt vor: Der Algorithmus soll Problembeispiele  $\mathcal{I}$  des Problems  $\Pi$  lösen. Alle Problembeispiele liegen in kodierter Form vor. Die Anzahl der Zellen, die notwendig sind, um  $\mathcal{I}$  vollständig anzugeben, nennen wir die *Kodierungslänge* oder *Inputlänge*  $\langle \mathcal{I} \rangle$  von  $\mathcal{I}$ . Der Algorithmus liest diese Daten und beginnt dann Operationen auszuführen, d. h. Zahlen zu berechnen, zu speichern, zu löschen, usw. Die Anzahl der Zellen, die während der Ausführung des Algorithmus  $A$  mindestens einmal benutzt wurden, nennen wir den *Speicherbedarf von  $A$  zur Lösung von  $\mathcal{I}$* . Üblicherweise schätzt man den Speicherbedarf eines Algorithmus  $A$  dadurch nach oben ab, dass man die Anzahl der von  $A$  benutzten Speicherplätze bestimmt und diesen Wert mit der größten Anzahl von Zellen (Bits) multipliziert, die einer der Speicherplätze beim Ablauf des Algorithmus benötigte.

Die *Laufzeit von  $A$  zur Lösung von  $\mathcal{I}$*  ist (etwas salopp gesagt) die Anzahl der elementaren Operationen, die  $A$  bis zur Beendigung des Verfahrens ausgeführt hat. Dabei wollen wir als *elementare Operationen* zählen:

Lesen, Schreiben und Löschen,

Addieren, Subtrahieren, Multiplizieren, Dividieren und Vergleichen

von rationalen (oder ganzen) Zahlen. Da ja zur Darstellung derartiger Zahlen mehrere Zellen benötigt werden, muss zur genauen Berechnung der Laufzeit jede elementare Operation mit den Kodierungslängen der involvierten Zahlen multipliziert werden. Die *Laufzeit von  $A$  zur Lösung von  $\mathcal{I}$*  ist die Anzahl der elementaren Rechenoperationen, die  $A$  ausgeführt hat, um eine Lösung von  $\mathcal{I}$  zu finden, multipliziert mit der Kodierungslänge der bezüglich der Kodierungslänge größten ganzen oder rationalen Zahl, die während der Ausführung des Algorithmus aufgetreten ist. Wir tun also so, als hätten wir alle elementaren Operationen mit der in diesem Sinne größten Zahl ausgeführt und erhalten somit eine Abschätzung der „echten“ Laufzeit nach oben.

**(4.2) Definition.** Sei  $A$  ein Algorithmus zur Lösung eines Problems  $\Pi$ .

(a) Die Funktion  $f_A : \mathbb{N} \rightarrow \mathbb{N}$ , definiert durch

$$f_A(n) := \max\{\text{Laufzeit von } A \text{ zur Lösung von } \mathcal{I} \mid \mathcal{I} \in \Pi \text{ und } \langle \mathcal{I} \rangle \leq n\},$$

heißt Laufzeitfunktion von  $A$ .

## 4 Komplexitätstheorie und Speicherung von Graphen

(b) Die Funktion  $s_A : \mathbb{N} \rightarrow \mathbb{N}$ , definiert durch

$$s_A(n) := \max\{\text{Speicherbedarf von } A \text{ zur Lösung von } \mathcal{I} \mid \mathcal{I} \in \Pi \text{ und } \langle \mathcal{I} \rangle \leq n\},$$

heißt Speicherplatzfunktion von  $A$ .

(c) Der Algorithmus  $A$  hat eine polynomiale Laufzeit (kurz:  $A$  ist ein polynomialer Algorithmus), wenn es ein Polynom  $p : \mathbb{N} \rightarrow \mathbb{N}$  gibt mit

$$f_A(n) \leq p(n) \quad \text{für alle } n \in \mathbb{N}.$$

Wir sagen  $f_A$  ist von der Ordnung höchstens  $n^k$  (geschrieben  $f_A = O(n^k)$ ), falls das Polynom  $p$  den Grad  $k$  hat.

(d) Der Algorithmus  $A$  hat polynomialen Speicherplatzbedarf, wenn es ein Polynom  $q : \mathbb{N} \rightarrow \mathbb{N}$  gibt mit  $s_A(n) \leq q(n)$  für alle  $n \in \mathbb{N}$ .  $\triangle$

Wir werden uns in dieser Vorlesung hauptsächlich mit Problemen beschäftigen, für die polynomiale Algorithmen existieren. Wir werden aber auch Probleme behandeln, die in einem noch zu präzisierenden Sinne „schwieriger“ sind und für die (bisher) noch keine polynomialen Verfahren gefunden worden sind.

Eine triviale Bemerkung sei hier gemacht. Ein Algorithmus, dessen Speicherplatzfunktion nicht durch ein Polynom beschränkt werden kann, kann keine polynomiale Laufzeit haben, da nach Definition jede Benutzung eines Speicherplatzes in die Berechnung der Laufzeitfunktion eingeht.

**Hausaufgabe.** Bestimmen Sie die Laufzeitfunktion und die Speicherplatzfunktion des folgenden Algorithmus:

**Eingabe:** ganze Zahl  $n$

- 1:  $k := \langle n \rangle$
- 2: **for**  $i = 1, \dots, k$  **do**
- 3:      $n := n \cdot n \cdot n$
- 4: **end for**
- 5: Gib  $n$  aus.

## 4.2 Die Klassen $\mathcal{P}$ und $\mathcal{NP}$ , $\mathcal{NP}$ -Vollständigkeit

Wir wollen nun einige weitere Begriffe einführen, um zwischen „einfachen“ und „schwierigen“ Problemen unterscheiden zu können. Wir werden dabei zunächst – aus technischen Gründen – nur Entscheidungsprobleme behandeln und später die Konzepte auf (die uns eigentlich interessierenden) Optimierungsprobleme erweitern. Ein *Entscheidungsproblem* ist ein Problem, das nur zwei mögliche Antworten besitzt, nämlich „ja“ oder „nein“. Die Fragen „Enthält ein Graph einen Kreis?“, „Enthält ein Graph einen hamiltonschen Kreis?“, „Ist die Zahl  $n$  eine Primzahl?“ sind z. B. Entscheidungsprobleme. Da wir uns nicht mit (für uns unwichtigen) Feinheiten der Komplexitätstheorie beschäftigen wollen,

werden wir im weiteren nur solche Entscheidungsprobleme betrachten, für die Lösungsalgorithmen mit endlicher Laufzeitfunktion existieren.

Die Klasse aller derjenigen Entscheidungsprobleme, für die ein polynomialer Lösungsalgorithmus existiert, wird mit  $\mathcal{P}$  bezeichnet. Diese Definition ist recht informell. Wenn wir genauer wären, müssten wir  $\mathcal{P}$  relativ zu einem Kodierungsschema und zu einem Rechnermodell definieren. Die Definition würde dann etwa wie folgt lauten. Gegeben sei ein Kodierungsschema  $E$  und ein Rechnermodell  $M$ ,  $\Pi$  sei ein Entscheidungsproblem, wobei jedes Problembeispiel aus  $\Pi$  durch das Kodierungsschema  $E$  kodiert werden kann.  $\Pi$  gehört zur Klasse  $\mathcal{P}$  (bezüglich  $E$  und  $M$ ), wenn es einen auf  $M$  implementierbaren Algorithmus zur Lösung der Problembeispiele aus  $\Pi$  gibt, dessen Laufzeitfunktion auf  $M$  polynomial ist. Wir wollen im weiteren derartig komplizierte und unübersichtliche Definitionen vermeiden und werden auf dem (bisherigen) informellen Niveau bleiben in der Annahme, die wesentlichen Anliegen ausreichend klar machen zu können.

Wir werden im Abschnitt 4.3 sehen, dass das Problem „Enthält ein Graph einen Kreis?“ zur Klasse  $\mathcal{P}$  gehört. Aber trotz enormer Anstrengungen sehr vieler Forscher ist es noch nicht gelungen, das Problem „Enthält ein Graph einen hamiltonschen Kreis“ in polynomialer Zeit zu lösen.

Diese Frage ist „offenbar“ schwieriger. Um zwischen den Problemen in  $\mathcal{P}$  und den „schwierigeren“ formal unterscheiden zu können, sind die folgenden Begriffe geprägt worden.

Wir sagen – zunächst einmal informell –, dass ein Entscheidungsproblem  $\Pi$  zur Klasse  $\mathcal{NP}$  gehört, wenn es die folgende Eigenschaft hat: Ist die Antwort für ein Problembeispiel  $\mathcal{I} \in \Pi$  „ja“, dann kann die Korrektheit der Antwort in polynomialer Zeit überprüft werden.

Bevor wir etwas genauer werden, betrachten wir ein Beispiel. Wir wollen herausfinden, ob ein Graph einen hamiltonschen Kreis enthält. Jeden Graphen können wir (im Prinzip) auf ein Blatt Papier zeichnen. Hat der gegebene Graph einen hamiltonschen Kreis, und hat jemand (irgendwie) einen solchen gefunden und alle Kanten des Kreises rot angestrichen, dann können wir auf einfache Weise überprüfen, ob die rot angemalten Kanten tatsächlich einen hamiltonschen Kreis darstellen. Bilden sie einen solchen Kreis, so haben wir die Korrektheit der „ja“-Antwort in polynomialer Zeit verifiziert.

Nun die ausführliche Definition:

**(4.3) Definition.** *Ein Entscheidungsproblem  $\Pi$  gehört zur Klasse  $\mathcal{NP}$ , wenn es die folgenden Eigenschaften hat:*

1. *Für jedes Problembeispiel  $\mathcal{I} \in \Pi$ , für das die Antwort „ja“ lautet, gibt es mindestens ein Objekt  $Q$ , mit dessen Hilfe die Korrektheit der „ja“-Antwort überprüft werden kann.*
2. *Es gibt einen Algorithmus, der Problembeispiele  $\mathcal{I} \in \Pi$  und Zusatzobjekte  $Q$  als Input akzeptiert und der in einer Laufzeit, die polynomial in  $\langle \mathcal{I} \rangle$  ist, überprüft, ob  $Q$  ein Objekt ist, aufgrund dessen Existenz eine „ja“-Antwort für  $\mathcal{I}$  gegeben werden muss.  $\triangle$*

#### 4 Komplexitätstheorie und Speicherung von Graphen

Die Probleme „Hat ein Graph  $G$  einen Kreis?“, „Hat ein Graph  $G$  einen hamiltonschen Kreis?“ sind somit in  $\mathcal{NP}$ . Hat nämlich  $G$  einen Kreis oder hamiltonschen Kreis, so wählen wir diesen als Objekt  $Q$ . Dann entwerfen wir einen polynomialen Algorithmus, der für einen Graphen  $G$  und eine zusätzliche Kantenmenge  $Q$  entscheidet, ob  $Q$  ein Kreis oder hamiltonscher Kreis von  $G$  ist. Auch die Frage „Ist  $n \in \mathbb{N}$  eine zusammengesetzte Zahl?“ ist in  $\mathcal{NP}$ , denn liefern wir als „Objekt“ zwei Zahlen  $\neq 1$ , deren Produkt  $n$  ist, so ist  $n$  keine Primzahl. Die Überprüfung der Korrektheit besteht somit in diesem Fall aus einer einzigen Multiplikation.

Die obige Definition der Klasse  $\mathcal{NP}$  enthält einige Feinheiten, auf die ich ausdrücklich hinweisen möchte.

- Es wird nichts darüber gesagt, wie das Zusatzobjekt  $Q$  zu finden ist. Es wird lediglich postuliert, dass es existiert, aber nicht, dass man es z. B. mit einem polynomialen Algorithmus finden kann.
- Die Laufzeit des Algorithmus aus 2. in (4.3) ist nach Definition polynomial in  $\langle \mathcal{I} \rangle$ . Da der Algorithmus  $Q$  lesen muss, folgt daraus, dass die Kodierungslänge von  $Q$  durch ein Polynom in der Kodierungslänge von  $\mathcal{I}$  beschränkt sein muss. Auf die Frage „Hat die Gleichung  $x^2 + y^2 = n^2$  eine Lösung  $x, y$ ?“ ist „ $x = n$  und  $y = 0$ “ ein geeignetes Zusatzobjekt  $Q$ , aber weder „ $x = y = n\sqrt{0.5}$ “ ( $\sqrt{0.5}$  kann nicht endlich binär kodiert werden) noch „ $x = \frac{n2^{2^n}}{2^{2^n}}, y = 0$ “ (die Kodierungslänge von  $x$  ist exponentiell in der Inputlänge des Problems) wären als Zusatzobjekt  $Q$  geeignet, um die Korrektheit der „ja“-Antwort in polynomialer Zeit verifizieren zu können.
- Ferner ist die Definition von  $\mathcal{NP}$  unsymmetrisch in „ja“ und „nein“. Die Definition impliziert nicht, dass wir auch für die Problembeispiele mit „nein“-Antworten Objekte  $Q$  und polynomiale Algorithmen mit den in (4.3) spezifizierten Eigenschaften finden können.

Wir sagen, dass die Entscheidungsprobleme, die Negationen von Problemen aus der Klasse  $\mathcal{NP}$  sind, zur Klasse  $\text{co-}\mathcal{NP}$  gehören. Zu  $\text{co-}\mathcal{NP}$  gehören folglich die Probleme „Hat  $G$  keinen Kreis?“, „Hat  $G$  keinen hamiltonschen Kreis?“, „Ist  $n \in \mathbb{N}$  eine Primzahl?“. Es ist bekannt, dass das erste und das letzte dieser drei Probleme ebenfalls zu  $\mathcal{NP}$  gehören. Diese beiden Probleme gehören also zu  $\mathcal{NP} \cap \text{co-}\mathcal{NP}$ . Vom Problem „Hat  $G$  keinen hamiltonschen Kreis?“ weiß man nicht, ob es zu  $\mathcal{NP}$  gehört. Niemand hat bisher Objekte  $Q$  und einen Algorithmus finden können, die den Forderungen 1 und 2 aus 4.3 genügen.

Das Symbol  $\mathcal{NP}$  ist abgeleitet vom Begriff „nichtdeterministischer polynomialer Algorithmus“. Dies sind – grob gesagt – Algorithmen, die am Anfang „raten“, also einen nichtdeterministischen Schritt ausführen können und dann wie übliche Algorithmen ablaufen. Ein nichtdeterministischer Algorithmus „löst“ z. B. das hamiltonsche Graphenproblem wie folgt: Am Anfang rät er einen hamiltonschen Kreis. Gibt es keinen, so hört das Verfahren auf. Gibt es einen, so überprüft er, ob das geratene Objekt tatsächlich ein hamiltonscher Kreis ist. Ist das so, so antwortet er mit „ja“.

Trivialerweise gilt  $\mathcal{P} \subseteq \mathcal{NP}$ , da für Probleme in  $\mathcal{P}$  Algorithmen existieren, die ohne Zusatzobjekte  $Q$  in polynomialer Zeit eine „ja“- oder „nein“-Antwort liefern. Also

gilt auch  $\mathcal{P} \subseteq \text{co-}\mathcal{NP}$ . Eigentlich sollte man meinen, dass Algorithmen, die raten können, mächtiger sind als übliche Algorithmen. Trotz gewaltiger Forschungsanstrengungen seit den 1970er Jahren ist die Frage, ob  $\mathcal{P} = \mathcal{NP}$  gilt oder nicht, immer noch ungeklärt. Meiner Meinung nach ist dieses Problem eines der wichtigsten offenen Probleme der heutigen Mathematik und Informatik. Das Clay Mathematics Institute hat im Jahr 2000 einen Preis von 1 Mio US\$ für die Lösung des  $\mathcal{P} = \mathcal{NP}$ -Problems ausgesetzt, siehe: <http://www.claymath.org/millennium-problems>. Jeder, der sich mit diesem Problem beschäftigt hat, glaubt, dass  $\mathcal{P} \neq \mathcal{NP}$  gilt. (Eine für die allgemeine Leserschaft geschriebene Diskussion dieser Frage ist in Grötschel (2002) zu finden.) Könnte diese Vermutung bestätigt werden, so würde das – wie wir gleich sehen werden – bedeuten, dass für eine sehr große Zahl praxisrelevanter Probleme niemals wirklich effiziente Lösungsalgorithmen gefunden werden können. Wir werden uns also mit der effizienten Auffindung suboptimaler Lösungen zufrieden geben und daher auf den Entwurf von Heuristiken konzentrieren müssen. Deswegen wird auch im weiteren Verlauf des Vorlesungszyklus viel Wert auf die Untersuchung und Analyse von Heuristiken gelegt.

Wir haben gesehen, dass  $\mathcal{P} \subseteq \mathcal{NP} \cap \text{co-}\mathcal{NP}$  gilt. Auch bezüglich der Verhältnisse dieser drei Klassen zueinander gibt es einige offene Fragen.

Gilt  $\mathcal{P} = \mathcal{NP} \cap \text{co-}\mathcal{NP}$ ?

Gilt  $\mathcal{NP} = \text{co-}\mathcal{NP}$ ?

Aus  $\mathcal{NP} \neq \text{co-}\mathcal{NP}$  würde  $\mathcal{P} \neq \mathcal{NP}$  folgen, da offenbar  $\mathcal{P} = \text{co-}\mathcal{P}$  gilt.

Die Klassenzugehörigkeit des oben erwähnten und bereits von den Griechen untersuchten Primzahlproblems war lange Zeit offen. Dass das Primzahlproblem in  $\text{co-}\mathcal{P}$  ist, haben wir oben gezeigt. Rivest gelang es 1977 zu zeigen, dass das Primzahlproblem auch in  $\mathcal{NP}$  ist. Beginnend mit dem Sieb des Erathostenes sind sehr viele Testprogramme entwickelt worden. Erst 2002 gelang es drei Indern, einen polynomialen Algorithmus zu entwickeln, der in polynomialer Zeit herausfindet, ob eine ganze Zahl eine Primzahl ist oder nicht, siehe Agrawal et al. (2004) oder URL:

[http://www.cse.iitk.ac.in/users/manindra/algebra/primality\\_v6.pdf](http://www.cse.iitk.ac.in/users/manindra/algebra/primality_v6.pdf)

Wir wollen nun innerhalb der Klasse  $\mathcal{NP}$  eine Klasse von besonders schwierigen Problemen auszeichnen.

**(4.4) Definition.** Gegeben seien zwei Entscheidungsprobleme  $\Pi$  und  $\Pi'$ . Eine polynomiale Transformation von  $\Pi$  in  $\Pi'$  ist ein polynomialer Algorithmus, der, gegeben ein (kodierte) Problembeispiel  $\mathcal{I} \in \Pi$ , ein (kodierte) Problembeispiel  $\mathcal{I}' \in \Pi'$  produziert, so dass folgendes gilt:

Die Antwort auf  $\mathcal{I}$  ist genau dann „ja“, wenn die Antwort auf  $\mathcal{I}'$  „ja“ ist.  $\triangle$

Offenbar gilt Folgendes: Ist  $\Pi$  in  $\Pi'$  polynomial transformierbar und gibt es einen polynomialen Algorithmus zur Lösung von  $\Pi'$ , dann kann man auch  $\Pi$  in polynomialer Zeit lösen. Man transformiert einfach jedes Problembeispiel aus  $\Pi$  in ein Problembeispiel

aus  $\Pi'$  und wendet den Algorithmus für  $\Pi'$  an. Da sowohl der Transformationsalgorithmus als auch der Lösungsalgorithmus polynomial sind, hat auch die Kombination beider Algorithmen eine polynomiale Laufzeit.

Nun kommen wir zu einem der wichtigsten Begriffe dieser Theorie, der spezifiziert, welches die schwierigsten Probleme in der Klasse  $\mathcal{NP}$  sind.

**(4.5) Definition.** Ein Entscheidungsproblem  $\Pi$  heißt  $\mathcal{NP}$ -vollständig, falls  $\Pi \in \mathcal{NP}$  und falls jedes andere Problem aus  $\mathcal{NP}$  polynomial in  $\Pi$  transformiert werden kann.  $\triangle$

Jedes  $\mathcal{NP}$ -vollständige Entscheidungsproblem  $\Pi$  hat also die folgende Eigenschaft. Falls  $\Pi$  in polynomialer Zeit gelöst werden kann, dann kann auch jedes andere Problem aus  $\mathcal{NP}$  in polynomialer Zeit gelöst werden; in Formeln:

$$\Pi \text{ } \mathcal{NP}\text{-vollständig und } \Pi \in \mathcal{P} \implies \mathcal{P} = \mathcal{NP}.$$

Diese Eigenschaft zeigt, dass – bezüglich polynomialer Lösbarkeit – kein Problem in  $\mathcal{NP}$  schwieriger ist als ein  $\mathcal{NP}$ -vollständiges. Natürlich stellt sich sofort die Frage, ob es überhaupt  $\mathcal{NP}$ -vollständige Probleme gibt. Dies hat Cook (1971) in einer für die Komplexitätstheorie fundamentalen Arbeit bewiesen. In der Tat sind (leider) fast alle praxisrelevanten Probleme  $\mathcal{NP}$ -vollständig. Ein Beispiel: Das hamiltonsche Graphenproblem „Enthält  $G$  einen hamiltonschen Kreis?“ ist  $\mathcal{NP}$ -vollständig.

Wir wollen nun Optimierungsprobleme in unsere Betrachtungen einbeziehen. Aus jedem Optimierungsproblem kann man wie folgt ein Entscheidungsproblem machen. Ist  $\Pi$  ein Maximierungsproblem (Minimierungsproblem), so legt man zusätzlich zu jedem Problembeispiel  $\mathcal{I}$  noch eine Schranke, sagen wir  $B$ , fest und fragt:

Gibt es für  $\mathcal{I}$  eine Lösung, deren Wert nicht kleiner (nicht größer) als  $B$  ist?

Aus dem Travelling-Salesman-Problem wird auf diese Weise ein Entscheidungsproblem, man fragt einfach: „Enthält das Problembeispiel eine Rundreise, deren Länge nicht größer als  $B$  ist?“.

Wir nennen ein Optimierungsproblem  $\mathcal{NP}$ -schwer, wenn das (wie oben angegebene) zugeordnete Entscheidungsproblem  $\mathcal{NP}$ -vollständig ist. Diese Bezeichnung beinhaltet die Aussage, dass alle  $\mathcal{NP}$ -schweren Optimierungsprobleme mindestens so schwierig sind wie die  $\mathcal{NP}$ -vollständigen Probleme. Könnten wir nämlich ein  $\mathcal{NP}$ -schweres Problem (wie das Travelling-Salesman-Problem) in polynomialer Zeit lösen, dann könnten wir auch das zugehörige Entscheidungsproblem in polynomialer Zeit lösen. Wir berechnen den Wert  $w$  einer Optimallösung und vergleichen ihn mit  $B$ . Ist bei einem Maximierungsproblem (Minimierungsproblem)  $w \geq B$  ( $w \leq B$ ), so antworten wir „ja“, andernfalls „nein“.

Häufig kann man Entscheidungsprobleme dazu benutzen, um Optimierungsprobleme zu lösen. Betrachten wir als Beispiel das TSP-Entscheidungsproblem und nehmen wir an, dass alle Problembeispiele durch ganzzahlige Entfernungen zwischen den Städten gegeben sind. Ist  $s$  die kleinste vorkommende Zahl, so ziehen wir von allen Entfernungen den Wert  $s$  ab. Damit hat dann die kürzeste Entfernung den Wert 0. Ist nun  $t$  die längste aller (so modifizierten) Entfernungen, so ist offensichtlich, dass jede Tour eine nichtnegative Länge hat und, da jede Tour  $n$  Kanten enthält, ist keine Tourlänge größer



als  $n \cdot t$ . Wir fragen nun den Algorithmus zur Lösung des TSP-Entscheidungsproblems, ob es eine Rundreise gibt, deren Länge nicht größer als  $\frac{nt}{2}$  ist. Ist das so, fragen wir, ob es eine Rundreise gibt, deren Länge höchstens  $\frac{nt}{4}$  ist, andernfalls fragen wir, ob es eine Rundreise gibt mit Länge höchstens  $\frac{3nt}{4}$ . Wir fahren auf diese Weise fort, bis wir das Intervall ganzer Zahlen, die als mögliche Länge einer kürzesten Rundreise in Frage kommen, auf eine einzige Zahl reduziert haben. Diese Zahl muss dann die Länge der kürzesten Rundreise sein. Insgesamt haben wir zur Lösung des Optimierungsproblems das zugehörige TSP-Entscheidungsproblem ( $\lceil \log_2(nt) \rceil + 1$ )-mal aufgerufen, also eine polynomiale Anzahl von Aufrufen eines Algorithmus vorgenommen. Dieser Algorithmus findet also in polynomialer Zeit die Länge einer kürzesten Rundreise – bei einem gegebenen polynomialen Algorithmus für das zugehörige Entscheidungsproblem. (Überlegen Sie sich, ob – und gegebenenfalls wie – man eine kürzeste Rundreise finden kann, wenn man ihre Länge kennt!)

Dem aufmerksamen Leser wird nicht entgangen sein, dass die oben beschriebene Methode zur Reduktion des Travelling-Salesman-Problems auf das zugehörige Entscheidungsproblem nichts anderes ist als das bekannte Verfahren der *binären Suche*.

Mit diesem oder ähnlichen „Tricks“ lassen sich viele Optimierungsprobleme durch mehrfachen Aufruf von Algorithmen für Entscheidungsprobleme lösen. Wir nennen ein Optimierungsproblem  $\Pi$   *$\mathcal{NP}$ -leicht*, falls es ein Entscheidungsproblem  $\Pi'$  in  $\mathcal{NP}$  gibt, so dass  $\Pi$  durch polynomial viele Aufrufe eines Algorithmus zur Lösung von  $\Pi'$  gelöst werden kann.  $\mathcal{NP}$ -leichte Probleme sind also nicht schwerer als die Probleme in  $\mathcal{NP}$ . Unser Beispiel oben zeigt, dass das TSP auch  $\mathcal{NP}$ -leicht ist.

Wir nennen ein Optimierungsproblem  *$\mathcal{NP}$ -äquivalent*, wenn es sowohl  $\mathcal{NP}$ -leicht als auch  $\mathcal{NP}$ -schwer ist. Diese Bezeichnung ist im folgenden Sinne gerechtfertigt. Ein  $\mathcal{NP}$ -äquivalentes Problem ist genau dann in polynomialer Zeit lösbar, wenn  $\mathcal{P} = \mathcal{NP}$  gilt. Wenn jemand einen polynomialen Algorithmus für das TSP findet, hat er damit gleichzeitig  $\mathcal{P} = \mathcal{NP}$  bewiesen.

Wir wollen im Weiteren dem allgemein üblichen Brauch folgen und die feinen Unterschiede zwischen den oben eingeführten Bezeichnungen für Entscheidungs- und Optimierungsprobleme nicht so genau nehmen. Wir werden häufig einfach von  *$\mathcal{NP}$ -vollständigen Optimierungsproblemen* sprechen, wenn diese  $\mathcal{NP}$ -schwer sind. Die Begriffe  $\mathcal{NP}$ -leicht und  $\mathcal{NP}$ -äquivalent werden wir kaum gebrauchen, da sie für unsere Belange nicht so wichtig sind.

Die folgenden Beispiele von kombinatorischen Optimierungsproblemen, die wir in früheren Abschnitten eingeführt haben, sind  $\mathcal{NP}$ -schwer:

- das symmetrische Travelling Salesman Problem
- das asymmetrische Travelling Salesman Problem
- das Chinesische Postbotenproblem für gemischte Graphen
- fast alle Routenplanungsprobleme
- das Stabile-Mengen-Problem

#### 4 Komplexitätstheorie und Speicherung von Graphen

- das Cliquesproblem
- das Knotenüberdeckungsproblem
- das Knotenfärbungsproblem
- das Kantenfärbungsproblem
- das Max-Cut-Problem
- die meisten Standortprobleme
- das Linear-Ordering-Problem
- das azyklische Subdigraphenproblem
- das Feedback-Arc-Set-Problem.

Einige hundert weitere  $\mathcal{NP}$ -vollständige bzw.  $\mathcal{NP}$ -schwere Probleme und einige tausend Varianten von diesen sind in dem bereits zitierten Buch von Garey and Johnson (1979) aufgeführt. Probleme, die mit diesem Themenkreis zusammenhängen, wurden auch in einer von 1981–1992 laufenden Serie von Aufsätzen von D. S. Johnson mit dem Titel „The  $\mathcal{NP}$ -completeness column: an ongoing guide“ im *Journal of Algorithms* behandelt. Seit 2005 sind drei weitere Artikel der Serie in *ACM Trans. Algorithms* publiziert worden. Der derzeit letzte Artikel ist 2007 erschienen. Alle (bisher) 26 Artikel sind unter der folgenden URL zu finden:

<http://davidjohnson.net/columns/>

Im Internet finden Sie unter der URL

<http://www.nada.kth.se/~viggo/problemlist/compendium.html>

ein „Compendium of  $\mathcal{NP}$  Optimization Problems“. Eine aktuelle Zusammenfassung der Geschichte des Konzepts der  $\mathcal{NP}$ -Vollständigkeit findet sich in Johnson (2012).

Der wichtigste Aspekt der hier skizzierten Theorie ist, dass man zwischen „einfachen“ und „schwierigen“ Problemen zu unterscheiden lernt, und dass man – sobald man weiß, dass ein Problem schwierig ist – andere Wege (Heuristiken, etc.) als bei Problemen in  $\mathcal{P}$  suchen muss, um das Problem optimal oder approximativ zu lösen. In dieser Vorlesung soll versucht werden, einige der Methoden zu beschreiben, mit denen man derartige Probleme angreifen kann.

Zum Schluss dieses Abschnitts sei angemerkt, dass es noch viel schwierigere Probleme als die  $\mathcal{NP}$ -schweren Probleme gibt. Ein „Klassiker“ unter den wirklich schwierigen Problemen ist das Halteproblem von Turing-Maschinen, die wir in Abschnitt 4.1 erwähnt haben. Wir skizzieren die Fragestellung kurz.

Wir nennen ein Entscheidungsproblem *entscheidbar*, wenn es eine Turing-Maschine (genauer einen Algorithmus, der auf einer Turing-Maschine läuft) gibt, die für jede Eingabe

### 4.3 Datenstrukturen zur Speicherung von Graphen

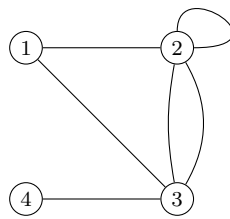


Abbildung 4.1: Ungerichteter Graph zur Kantenliste 4, 6, 1, 2, 2, 3, 4, 3, 3, 2, 2, 2, 1, 3.

terminiert und „ja“ oder „nein“ antwortet. Das *Halteproblem* ist nun das folgende: Gegeben seien eine Turing-Maschine  $M$  und eine Eingabe  $w$ , hält  $M$  auf  $w$ ? Dieses Halteproblem ist beweisbar unentscheidbar. Der *Satz von Rice* zeigt noch etwas viel Stärkeres. Alle „interessanten“ (das kann man präzise definieren) Eigenschaften von Turing-Maschinen sind unentscheidbar.

David Hilbert hat im Jahr 1900 eine Liste von 23 wichtigen, seinerzeit ungelösten, mathematischen Problemen vorgelegt. Sein zehntes Problem lautete: Gibt es einen Algorithmus, der für eine beliebige diophantische Gleichung entscheidet, ob sie lösbar ist? Diophantische Gleichungen sind von der Form  $p(x_1, \dots, x_n) = 0$ , wobei  $p$  ein Polynom mit ganzzahligen Koeffizienten ist. J. Matijassewitsch hat 1970 in seiner Dissertation bewiesen, dass es keinen solchen Algorithmus gibt. Um Lösungen von allgemeinen diophantischen Gleichungen zu finden, muss man also spezielle Fälle betrachten, gute Einfälle und Glück haben.

## 4.3 Datenstrukturen zur Speicherung von Graphen

Wir wollen hier einige Methoden skizzieren, mit deren Hilfe man Graphen und Digraphen speichern kann, und ihre Vor- und Nachteile besprechen. Kenntnisse dieser Datenstrukturen sind insbesondere dann wichtig, wenn man lauffzeit- und speicherplatzeffiziente (oder gar -optimale) Codes von Algorithmen entwickeln will.

### 4.3.1 Kanten- und Bogenlisten

Die einfachste Art, einen Graphen oder Digraphen zu speichern, ist die *Kantenliste* für Graphen bzw. die *Bogenliste* für Digraphen. Ist  $G = (V, E)$  ein Graph mit  $n = |V|$  Knoten und  $m = |E|$  Kanten, so sieht eine Kantenliste wie folgt aus:

$$n, m, a_1, e_1, a_2, e_2, a_3, e_3, \dots, a_m, e_m,$$

wobei  $a_i, e_i$  die beiden Endknoten der Kante  $i$  sind. Die Reihenfolge des Aufführens der Endknoten von  $i$  bzw. der Kanten selbst ist beliebig. Bei Schleifen wird der Endknoten zweimal aufgelistet. Eine mögliche Kantenliste für den Graphen aus Abbildung 4.1 ist die folgende:

$$4, 6, 1, 2, 2, 3, 4, 3, 3, 2, 2, 2, 1, 3.$$

#### 4 Komplexitätstheorie und Speicherung von Graphen

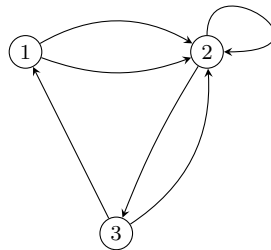


Abbildung 4.2: Gerichteter Graph zur Kantenliste 3, 6, 1, 2, 2, 3, 3, 2, 2, 2, 1, 2, 3, 1.

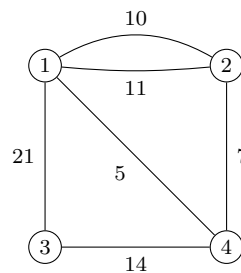


Abbildung 4.3: Gewichteter Graph mit den Kantenlisten 4, 6, 1, 2, 10, 2, 1, 11, 2, 4, 7, 4, 3, 14, 3, 1, 21, 1, 4, 5 und 4, 6, 1, 2, 1, 2, 2, 4, 3, 4, 1, 4, 1, 3, 11, 10, 7, 14, 5, 21.

Bei der Bogenliste eines Digraphen verfahren wir genauso; wir müssen jedoch darauf achten, dass ein Bogen immer zunächst durch seinen Anfangs- und dann durch seinen Endknoten repräsentiert wird. Eine Bogenliste des Digraphen aus Abbildung 4.2 ist

3, 6, 1, 2, 2, 3, 3, 2, 2, 2, 1, 2, 3, 1.

Haben die Kanten oder Bögen Gewichte, so repräsentiert man eine Kante (einen Bogen) entweder durch Anfangsknoten, Endknoten, Gewicht oder macht eine Kanten- bzw. Bogenliste wie oben und hängt an diese noch eine Liste mit den  $m$  Gewichten der Kanten  $1, 2, \dots, m$  an. Der gewichtete Graph aus Abbildung 4.3 ist in den beiden folgenden Kantenlisten mit Gewichten gespeichert:

4, 6, 1, 2, 10, 2, 1, 11, 2, 4, 7, 4, 3, 14, 3, 1, 21, 1, 4, 5  
 4, 6, 1, 2, 1, 2, 2, 4, 3, 4, 1, 4, 1, 3, 11, 10, 7, 14, 5, 21.

Der Speicheraufwand einer Kanten- bzw. Bogenliste beträgt  $2(m + 1)$  Speicherplätze, eine Liste mit Gewichten erfordert  $3m + 2$  Speicherplätze.

### 4.3.2 Adjazenzmatrizen

Ist  $G = (V, E)$  ein ungerichteter Graph mit  $V = \{1, 2, \dots, n\}$ , so ist die symmetrische  $(n, n)$ -Matrix  $A = (a_{ij})$  mit

$$a_{ji} = a_{ij} = \text{Anzahl der Kanten, die } i \text{ und } j \text{ verbinden, falls } i \neq j,$$

$$a_{ii} = 2 \cdot \text{Anzahl der Schleifen, die } i \text{ enthalten, } i = 1, \dots, n$$

die *Adjazenzmatrix* von  $G$ . Aufgrund der Symmetrie kann man etwa die Hälfte der Speicherplätze sparen. Hat ein Graph keine Schleifen (unser Normalfall), dann genügt es, die obere (oder untere) Dreiecksmatrix von  $A$  zu speichern. Man macht das z. B. in der Form

$$a_{12}, a_{13}, \dots, a_{1n}, a_{23}, a_{24}, \dots, a_{2n}, a_{34}, \dots, a_{n-1,n}.$$

Die Adjazenzmatrix des Graphen in Abbildung 4.1 sieht wie folgt aus:

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 2 & 2 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Hat ein Graph Kantengewichte, und ist er einfach, so setzt man

$$a_{ij} = \text{Gewicht der Kante } ij, \quad i, j = 1, \dots, n.$$

Ist eine Kante  $ij$  nicht im Graphen  $G$  enthalten, so setzt man

$$a_{ij} = 0, \quad a_{ij} = -\infty \quad \text{oder} \quad a_{ij} = +\infty,$$

je nachdem, welche Gewichtszuordnung für das Problem sinnvoll ist und die Benutzung dieser Kante ausschließt. Wiederum braucht man von der so definierten Matrix  $A$  nur die obere Dreiecksmatrix zu speichern.

Die Adjazenzmatrix  $A = (a_{ij})$  eines Digraphen  $D = (V, E)$  mit  $V = \{1, 2, \dots, n\}$  ohne Schleifen ist definiert durch

$$a_{ii} = 0, \quad i = 1, 2, \dots, n$$

$$a_{ij} = \text{Anzahl der Bögen in } E \text{ mit Anfangsknoten } i \text{ und Endknoten } j, \quad i \neq j$$

Bogengewichte werden wie bei Adjazenzmatrizen von Graphen behandelt. Der Speicheraufwand von Adjazenzmatrizen beträgt  $n^2$  bzw.  $\binom{n}{2}$  Speicherplätze, falls bei einem ungerichteten Graphen nur die obere Dreiecksmatrix gespeichert wird.

### 4.3.3 Adjazenzlisten

Speichert man für einen Graphen  $G = (V, E)$  die Anzahl der Knoten und für jeden Knoten  $v \in V$  seinen Grad und die Namen der Nachbarknoten, so nennt man eine

#### 4 Komplexitätstheorie und Speicherung von Graphen

solche Datenstruktur *Adjazenzliste* von  $G$ . Für den Graphen aus Abbildung 4.1 sieht die Adjazenzliste wie folgt aus

	Knotennummer		
	↓		
4	1	2	2, 3
	2	5	1, 3, 3, 2, 2
	3	4	1, 2, 2, 4
	4	1	3
	↑		
	Grad		

Die Liste der Nachbarknoten eines Knoten  $v$  heißt *Nachbarliste* von  $v$ . Jede Kante  $ij$ ,  $i \neq j$ , ist zweimal repräsentiert, einmal auf der Nachbarliste von  $i$ , einmal auf der von  $j$ . Bei Gewichten hat man wieder zwei Möglichkeiten. Entweder man schreibt direkt hinter jeden Knoten  $j$  auf der Nachbarliste von  $i$  das Gewicht der Kante  $ij$ , oder man legt eine kompatible Gewichtsliste an.

Bei Digraphen geht man analog vor, nur speichert man auf der Nachbarliste eines Knoten  $i$  nur die Nachfolger von  $i$ . Man nennt diese Liste daher auch *Nachfolgerliste*. Ein Bogen wird also nur einmal in der Adjazenzliste eines Digraphen repräsentiert. (Wenn es für das Problem günstiger ist, kann man natürlich auch Vorgängerlisten statt Nachfolgerlisten oder beides anlegen.)

Zur Speicherung einer Adjazenzliste eines Graphen braucht man  $2n + 1 + 2m$  Speicherplätze, für die Adjazenzliste eines Digraphen werden  $2n + 1 + m$  Speicherplätze benötigt.

Kanten- bzw. Bogenlisten sind die kompaktesten aber unstrukturiertesten Speicherformen. Will man wissen, ob  $G$  die Kante  $ij$  enthält, muss man im schlechtesten Fall die gesamte Liste durchlaufen und somit  $2m$  Abfragen durchführen. Eine derartige Frage benötigt bei der Speicherung von  $G$  in einer Adjazenzmatrix nur eine Abfrage, während man bei einer Adjazenzliste die Nachbarliste von  $i$  (oder  $j$ ) durchlaufen und somit Grad von  $i$ , im schlechtesten Fall also  $n - 1$  Abfragen durchführen muss.

Für dünn besetzte Graphen ist die Speicherung in einer Adjazenzmatrix i. A. zu aufwendig. Es wird zu viel Platz vergeudet. Außerdem braucht fast jeder Algorithmus, der für Adjazenzmatrizen konzipiert ist, mindestens  $O(n^2)$  Schritte, da er ja jedes Element von  $A$  mindestens einmal anschauen muss, um zu wissen, ob die zugehörige Kante im Graphen ist oder nicht. Mit Hilfe von Adjazenzlisten kann man dagegen dünn besetzte Graphen in der Regel sehr viel effizienter bearbeiten. Das Buch Aho et al. (1974) informiert sehr ausführlich über dieses Thema.

Wir wollen hier als Beispiel nur einen einzigen einfachen, aber vielseitig und häufig anwendbaren Algorithmus zur Untersuchung von Graphen erwähnen: das Depth-First-Search-Verfahren (kurz DFS-Verfahren bzw. auf deutsch: Tiefensuche).

Wir nehmen an, dass ein Graph  $G = (V, E)$  gegeben ist und alle Knoten *unmarkiert* sind. Alle Kanten seien *unbenutzt*. Wir wählen einen Startknoten, sagen wir  $v$ , und *markieren* ihn. Dann wählen wir eine Kante, die mit  $v$  inzidiert, sagen wir  $vw$ , gehen zu  $w$  und *markieren*  $w$ . Die Kante  $vw$  ist nun *benutzt* worden. Allgemein verfahren wir wie folgt. Ist  $x$  der letzte von uns markierte Knoten, dann versuchen wir eine mit  $x$  inzidente

Kante  $xy$  zu finden, die noch nicht benutzt wurde. Ist  $y$  markiert, so suchen wir eine weitere mit  $x$  inzidente Kante, die noch nicht benutzt wurde. Ist  $y$  nicht markiert, dann gehen wir zu  $y$ , markieren  $y$  und beginnen von neuem ( $y$  ist nun der letzte markierte Knoten). Wenn die Suche nach Kanten, die mit  $y$  inzidieren und die noch nicht benutzt wurden, beendet ist (d. h. alle Kanten, auf denen  $y$  liegt, wurden einmal berührt), kehren wir zu  $x$  zurück und fahren mit der Suche nach unbenutzten Kanten, die mit  $x$  inzidieren fort, bis alle Kanten, die  $x$  enthalten, abgearbeitet sind. Diese Methode nennt man Tiefensuche, da man versucht, einen Knoten so schnell wie möglich zu verlassen und „tiefer“ in den Graphen einzudringen.

Eine derartige Tiefensuche teilt die Kanten des Graphen in zwei Teilmengen auf. Eine Kante  $xy$  heißt *Vorwärtskante*, falls wir bei der Ausführung des Algorithmus von einem markierten Knoten  $x$  entlang  $xy$  zum Knoten  $y$  gegangen sind und dabei  $y$  markiert haben. Andernfalls heißt  $xy$  *Rückwärtskante*. Man überlegt sich leicht, dass die Menge der Vorwärtskanten ein Wald von  $G$  ist, der in jeder Komponente von  $G$  einen aufspannenden Baum bildet. Ist der Graph zusammenhängend, so nennt man die Menge der Vorwärtskanten *DFS-Baum* von  $G$ . Mit Hilfe von Adjazenzlisten kann die Tiefensuche sehr effizient rekursiv implementiert werden.

#### (4.6) Algorithmus Depth-First-Search.

**Eingabe:** Graph  $G = (V, E)$  in Form einer Adjazenzliste, d. h. für jeden Knoten  $v \in V$  ist eine Nachbarliste  $N(v)$  gegeben.

**Ausgabe:** Kantenmenge  $T$  (=DFS-Baum, falls  $G$  zusammenhängend ist).

- 1: Alle Knoten  $v \in V$  seien unmarkiert.
- 2: Setze  $T := \emptyset$ .
- 3: **for all**  $v \in V$  **do**
- 4:     Ist  $v$  unmarkiert, dann rufe SEARCH( $v$ ) auf.
- 5: **end for**
- 6: Gib  $T$  aus.
- 7: **procedure** SEARCH( $v$ ) ▷ Rekursives Unterprogramm
- 8:     Markiere  $v$ .
- 9:     **for all**  $w \in N(v)$  **do**
- 10:         Ist  $w$  unmarkiert, setze  $T := T \cup \{vw\}$  und rufe SEARCH( $w$ ) auf.
- 11:     **end for**
- 12: **end procedure**

In Algorithmus (4.6) wird im Hauptprogramm jeder Knoten einmal berührt, und im Unterprogramm jede Kante genau zweimal. Hinzu kommt die Ausgabe von  $T$ . Die Laufzeit des Verfahrens ist also  $O(|V| + |E|)$ . Diese Laufzeit könnte man niemals bei der Speicherung von  $G$  in einer Adjazenzmatrix erreichen.

Mit Hilfe des obigen Verfahrens können wir unser mehrmals zitiertes Problem „Enthält  $G$  einen Kreis?“ lösen. Offensichtlich gilt:  $G$  enthält genau dann einen Kreis, wenn  $E \setminus T$  nicht leer ist. Wir haben somit einen polynomialen Algorithmus zur Lösung des Kreisproblems gefunden. Der DFS-Baum, von Algorithmus (4.6) produziert, hat einige interessante Eigenschaften, die man dazu benutzen kann, eine ganze Reihe von weiteren

Graphenproblemen sehr effizient zu lösen. Der hieran interessierte Leser sei z. B. auf das Buch Aho et al. (1974) verwiesen.

## Literaturverzeichnis

- M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in  $\mathcal{P}$ . *Annals of Mathematics*, 160: 781–793, 2004.
- A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *Design and Analysis of Computer Algorithms*. Addison-Wesley, Reading, 1974.
- W. J. Cook. The complexity of theorem proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158, Ohio, 1971. Shaker Heights.
- T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, Cambridge, Mass., 2001.
- M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of  $\mathcal{NP}$ -completeness*. W. H. Freeman and Company, New York, 1979. ISBN 0716710455.
- M. Grötschel.  $\mathcal{P} = \mathcal{NP}$ ?. *Elemente der Mathematik, Eine Zeitschrift der Schweizerischen Mathematischen Gesellschaft*, 57(3):96–102, 2002. (siehe <http://www.zib.de/groetschel/pubnew/biblio.html>).
- D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen (1990), editor, *Algorithms and Complexity, Handbook of Theoretical Computer Science*, volume Vol. A, pages 67–161. Elsevier, Amsterdam, 1990.
- D. S. Johnson. A brief history of  $\mathcal{NP}$ -completeness, 1954–2012. In *Optimization Stories*, Documenta Mathematica, pages 359–376. DMV, 2012.
- K. Mehlhorn. *Data Structures and Algorithms*, volume 1–3. Springer-Verlag, EATCS Monographie edition, 1984. (dreibändige Monographie, Band I liegt auch auf deutsch im Teubner-Verlag (1986) vor).
- K. Mehlhorn and P. Sanders. *Algorithms and Data Structures: The Basic Toolbox*. Springer, 2008.
- D. P. Mehta and S. Sahni, editors. *Handbook of Data Structures and Applications*. Chapman & Hall, 2004.
- C. Meinel. *Effiziente Algorithmen. Entwurf und Analyse*. Fachbuchverlag, Leipzig, 1991.
- T. Ottmann and P. Widmayer. *Algorithmen und Datenstrukturen*. Spektrum Akademischer Verlag, 2012.
- C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, Amsterdam, 1994.



- D. B. Shmoys and E. Tardos. Computational complexity. In *Handbook of Combinatorics*, chapter 29, pages 1599–1645. North-Holland, Amsterdam, 1995.
- R. E. Tarjan. Data structures and network algorithms. *SIAM*, 1983.
- J. van Leeuwen. Algorithms and complexity. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A, pages 525–631. Elsevier, Amsterdam, 1990.
- K. Wagner and G. Wechsung. *Computational Complexity*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1986.
- I. Wegener. *Komplexitätstheorie: Grenzen der Effizienz von Algorithmen*. Springer, 2003.



# 5 Bäume und Wege

Dieses Kapitel ist einem Thema gewidmet, das algorithmisch sehr einfach zu lösen ist: Bestimme einen kostenminimalen aufspannenden Baum in einem Graphen. Wir werden Varianten dieser Aufgabe und auch „gerichtete Versionen“ betrachten.

Bevor wir jedoch algorithmischen Fragen nachgehen, sollen Wälder aus graphentheoretischer Sicht analysiert werden. Das Ziel des ersten Abschnitts dieses Kapitels ist nicht eine umfassende Behandlung des Themenkreises, sondern das Einüben typischer graphentheoretischer Beweisargumente. Bäume sind sehr einfache Objekte. Die meisten Eigenschaften von Bäumen können mit minimalem Aufwand nachgewiesen werden. Die dabei benutzten Argumente tauchen jedoch in der Graphentheorie – meistens in etwas komplizierterer Form – immer wieder auf. Wir hoffen, diese Beweistechniken hier sichtbar machen zu können.

## 5.1 Graphentheoretische Charakterisierungen von Bäumen und Arboreszenzen

Wir erinnern daran, dass ein Graph, der keinen Kreis enthält, *Wald* genannt wird, dass ein Graph  $G$  *zusammenhängend* heißt, wenn es in  $G$  zwischen je zwei Knoten eine sie verbindende Kette (oder äquivalent dazu, einen sie verbindenden Weg) gibt, und dass ein *Baum* ein zusammenhängender Wald ist. Ein Untergraph eines Graphen  $G = (V, E)$ , der ein Baum ist und alle Knoten  $V$  enthält, heißt *aufspannender Baum* (von  $G$ ). Eine *Zusammenhangskomponente* (kurz: *Komponente*) eines Graphen  $G$  ist ein maximaler zusammenhängender Untergraph von  $G$ . Wir werden nun einige Eigenschaften von Bäumen und Wäldern beweisen. Wir beginnen mit trivialen Beobachtungen.

**(5.1) Lemma.** *Ein Baum  $G = (V, E)$  mit mindestens zwei Knoten hat mindestens zwei Knoten mit Grad 1.* △

**Beweis.** Da ein Baum zusammenhängend ist, liegt jeder Knoten auf mindestens einer Kante. Wir wählen einen beliebigen Knoten, sagen wir  $v$ . Wir starten in  $v$  einen (vereinfachten) DFS-Algorithmus. Wir markieren  $v$  und gehen zu einem Nachbarn, sagen wir  $w$ , von  $v$ . Wir markieren  $w$ . Hat  $w$  den Grad 1, stoppen wir die Suche. Andernfalls gehen wir zu einem von  $v$  verschiedenen Nachbarn von  $w$  und fahren so fort. Da ein Baum keinen Kreis enthält, kehrt dieses Verfahren niemals zu einem bereits markierten Knoten zurück. Da der Graph endlich ist, muss das Verfahren irgendwann mit einem Knoten mit Grad 1 aufhören. Hat der Anfangsknoten  $v$  auch Grad 1, können wir aufhören. Falls nicht, gehen wir zu einem von  $w$  verschiedenen Nachbarn von  $v$  und wiederholen das obige Verfahren. Auf diese Weise finden wir einen zweiten Knoten mit Grad 1. □

## 5 Bäume und Wege

Der Beweis ist etwas länglich geraten. Der Grund dafür ist, einmal zu zeigen, wie durch einfache Analyse eines sehr einfachen Algorithmus Eigenschaften von Graphen nachgewiesen werden können. Was können Sie aus diesem Beweisverfahren „herausholen“, wenn Sie den Algorithmus statt mit einem beliebigen Knoten  $v$  mit einem Knoten  $v$  mit maximalem Grad beginnen?

**(5.2) Lemma.** Sei  $G = (V, E)$  ein Graph.

(a) Es gilt:  $2|E| = \sum_{v \in V} \deg(v)$ .

(b) Ist  $G$  ein Baum, so gilt:  $|E| = |V| - 1$ .

(c) Gilt  $\deg(v) \geq 2$  für alle  $v \in V$ , so enthält  $G$  einen Kreis. △

**Beweis.** (a) Da jede Kante genau zwei (nicht notwendig verschiedene) Knoten enthält, wird bei der Summe der Knotengerade jede Kante genau zweimal gezählt.

(b) Beweis durch Induktion! Die Behauptung ist offensichtlich richtig für  $|V| = 1$  und  $|V| = 2$ . Wir nehmen an, dass die Behauptung korrekt ist für alle Bäume mit höchstens  $n \geq 2$  Knoten. Sei  $G = (V, E)$  ein Baum mit  $n + 1$  Knoten. Nach Lemma (5.1) enthält  $G$  einen Knoten  $v$  mit Grad 1.  $G - v$  ist dann ein Baum mit  $n$  Knoten. Nach Induktionsvoraussetzung hat  $G - v$  genau  $n - 1$  Kanten, also enthält  $G$  genau  $n$  Kanten.

(c) Enthält  $G$  Schlingen, so ist nichts zu zeigen, da Schlingen Kreise sind. Andernfalls führen wir folgenden Markierungsalgorithmus aus. Wir wählen einen beliebigen Knoten  $v_1$  und markieren ihn. Wir wählen einen beliebigen Nachbarn  $v_2$  von  $v_1$  und markieren sowohl  $v_2$  als auch die Kante  $v_1v_2$ . Haben wir die Knoten  $v_1, \dots, v_k$ ,  $k \geq 2$ , markiert, so enthält  $G$ , da  $\deg(v) \geq 2$ , eine von der markierten Kante  $v_{k-1}v_k$  verschiedene Kante  $v_kv_{k+1}$  zu einem  $v_k$ -Nachbarn  $v_{k+1}$ . Ist  $v_{k+1}$  bereits markiert, so ist  $v_{k+1}$  einer der bereits markierten Knoten, sagen wir  $v_i$ ,  $1 \leq i \leq k$ . Daher bildet die Knotenfolge  $(v_i, v_{i+1}, \dots, v_k)$  einen Kreis. Ist  $v_{k+1}$  nicht markiert, so markieren wir  $v_{k+1}$  und die Kante  $v_kv_{k+1}$  und fahren mit unserem Markierungsalgorithmus fort. Da  $|V|$  endlich ist, wird nach endlich vielen Schritten die Situation eintreten, dass jeder Nachbar des gerade markierten Knotens  $v_k$  bereits markiert ist. Und damit ist ein Kreis gefunden. □

**(5.3) Lemma.** Ein Graph  $G = (V, E)$  mit mindestens 2 Knoten und mit weniger als  $|V| - 1$  Kanten ist unzusammenhängend. △

**Beweis.** Sei  $m := |E|$ . Wäre  $G$  zusammenhängend, müsste es in  $G$  von jedem Knoten zu jedem anderen einen Weg geben. Wir führen einen Markierungsalgorithmus aus. Wir wählen einen beliebigen Knoten  $v \in V$  und markieren  $v$ . Wir markieren alle Nachbarn von  $v$  und entfernen die Kanten, die von  $v$  zu seinen Nachbarn führen. Wir gehen nun zu einem markierten Knoten, markieren dessen Nachbarn und entfernen die Kanten, die

noch zu diesen Nachbarn führen. Wir setzen dieses Verfahren fort, bis wir keinen Knoten mehr markieren können. Am Ende haben wir höchstens  $m$  Kanten entfernt sowie  $v$  und maximal  $m$  weitere Knoten markiert. Da  $m < |V| - 1$  gilt, ist mindestens ein Knoten unmarkiert, also nicht von  $v$  aus auf einem Weg erreichbar. Daher ist  $G$  unzusammenhängend.  $\square$

Der nächste Satz zeigt, dass die Eigenschaft, ein Baum zu sein, auf viele äquivalente Weisen charakterisiert werden kann.

**(5.4) Satz.** Sei  $G = (V, E)$ ,  $|V| = n \geq 2$  ein Graph. Dann sind äquivalent:

- (1)  $G$  ist ein Baum.
- (2)  $G$  enthält keinen Kreis und genau  $n - 1$  Kanten.
- (3)  $G$  ist zusammenhängend und enthält genau  $n - 1$  Kanten.
- (4)  $G$  ist zusammenhängend und enthält keinen Kreis.
- (5) Jedes Knotenpaar aus  $V$  ist durch genau einen Weg miteinander verbunden.
- (6)  $G$  enthält keinen Kreis; wird irgendeine Kante  $uv$  mit  $u, v \in V$  und  $uv \notin E$  zu  $G$  hinzugefügt, so entsteht genau ein Kreis.
- (7)  $G$  ist zusammenhängend, und für alle  $e \in E$  ist  $G - e$  unzusammenhängend.  $\triangle$

**Beweis.** (1)  $\iff$  (4) Definition.

(4)  $\implies$  (5) Da  $G$  zusammenhängend ist, ist jedes Knotenpaar durch einen Weg miteinander verbunden. Gibt es zwischen einem Knotenpaar zwei verschiedene Wege, so ist die Verknüpfung dieser beiden Wege eine geschlossene Kette, die offensichtlich einen Kreis enthält. Widerspruch!

(5)  $\implies$  (6) Enthielte  $G$  einen Kreis, so gäbe es Knotenpaare, die durch zwei verschiedene Wege miteinander verbunden sind. Also enthält  $G$  keinen Kreis. Sei  $uv \notin E$ . Da  $G$  einen  $[u, v]$ -Weg  $P$  enthält, ist  $P \cup uv$  ein Kreis. Gäbe es in  $G + uv$  einen weiteren Kreis, so gäbe es in  $G$  zwei verschiedene  $[u, v]$ -Wege, ein Widerspruch!

(6)  $\implies$  (7) Gibt es für  $uv \notin E$  in  $G + uv$  einen Kreis, so gibt es in  $G$  einen  $[u, v]$ -Weg. Daraus folgt, dass  $G$  zusammenhängend ist. Gibt es eine Kante  $uv \in E$  mit  $G - uv$  zusammenhängend, so gibt es in  $G - uv$  einen  $[u, v]$ -Weg  $P$ . Dann aber ist  $P \cup uv$  ein Kreis in  $G$ , Widerspruch!

(7)  $\implies$  (4) Gäbe es in  $G$  einen Kreis, so wäre  $G - e$  für jede Kante  $e$  dieses Kreises zusammenhängend. Also enthält  $G$  keinen Kreis.

(4)  $\implies$  (2) Folgt aus Lemma (5.2)(b).

## 5 Bäume und Wege

- (2)  $\implies$  (3) Sei  $k \geq 1$  die Anzahl der Zusammenhangskomponenten von  $G$ ,  $n_1, \dots, n_k$  ihre jeweiligen Knotenanzahlen und  $m_1, \dots, m_k$  ihre Kantenanzahlen. Jede Komponente  $i$  ist kreisfrei, da  $G$  kreisfrei ist, und damit ein Baum. Nach Lemma (5.2) gilt dann  $m_i = n_i - 1$ . Die Komponenten sind eine Partition von  $G$ , also gilt  $n = n_1 + \dots + n_k$  sowie  $|E| = m_1 + \dots + m_k = n - k$ . Wegen  $|E| = n - 1$  folgt  $k = 1$ , also ist  $G$  zusammenhängend.
- (3)  $\implies$  (4) Angenommen  $G$  enthält einen Kreis. Sei  $e$  eine Kante des Kreises, dann ist  $G - e$  zusammenhängend und hat  $n - 2$  Kanten. Dies widerspricht Lemma (5.3).  $\square$

Eine Kante  $e$  eines Graphen  $G$ , die die Eigenschaft besitzt, dass  $G - e$  unzusammenhängend ist, heißt *Brücke*. Satz (5.4)(7) zeigt insbesondere, dass  $G$  ein Baum genau dann ist, wenn jede Kante von  $G$  eine Brücke ist.

### (5.5) Korollar.

- (a) *Ein Graph ist zusammenhängend genau dann, wenn er einen aufspannenden Baum enthält.*
- (b) *Sei  $G = (V, E)$  ein Wald und sei  $p$  die Anzahl der Zusammenhangskomponenten von  $G$ , dann gilt  $|E| = |V| - p$ .  $\triangle$*

Die hier bewiesenen Eigenschaften kann man auf analoge Weise auch auf gerichtete Bäume und Wälder übertragen. Wir geben hier nur die Resultate an und laden den Leser ein, die Beweise selbst auszuführen.

Ein Digraph, der keinen Kreis enthält und bei dem jeder Knoten den Innengrad höchstens 1 hat (also  $\deg^-(v) \leq 1$ ), heißt *Branching*. Ein zusammenhängendes Branching heißt *Arboreszenz*. Jedes Branching ist also ein Wald, jede Arboreszenz ein Baum. Ein Knoten  $v$  in einem Digraphen heißt *Wurzel*, wenn jeder Knoten des Digraphen von  $v$  aus auf einem gerichteten Weg erreicht werden kann.

Ein Digraph  $D$  heißt *quasi-stark zusammenhängend*, falls es zu jedem Paar  $u, v$  von Knoten einen Knoten  $w$  in  $D$  (abhängig von  $u$  und  $v$ ) gibt, so dass es von  $w$  aus einen gerichteten Weg zu  $u$  und einen gerichteten Weg zu  $v$  gibt.

Es ist einfach zu sehen, dass jede Arboreszenz genau eine Wurzel hat, und dass ein Digraph genau dann quasi-stark zusammenhängend ist, wenn er eine Wurzel besitzt.

**(5.6) Satz.** *Sei  $D = (V, A)$  ein Digraph mit  $n \geq 2$  Knoten. Dann sind die folgenden Aussagen äquivalent:*

- (1)  *$D$  ist eine Arboreszenz.*
- (2)  *$D$  ist ein Baum mit Wurzel.*
- (3)  *$D$  enthält genau  $n - 1$  Bögen und ist quasi-stark zusammenhängend.*
- (4)  *$D$  enthält keinen Kreis und ist quasi-stark zusammenhängend.*

- (5)  $D$  enthält einen Knoten  $r$ , so dass es in  $D$  für jeden anderen Knoten  $v$  genau einen gerichteten  $(r, v)$ -Weg gibt.
- (6)  $D$  ist quasi-stark zusammenhängend, und für alle  $a \in A$  ist  $D - a$  nicht quasi-stark zusammenhängend.
- (7)  $D$  ist quasi-stark zusammenhängend, besitzt einen Knoten  $r$  mit  $\deg^-(r) = 0$  und erfüllt  $\deg^-(v) = 1$  für alle  $v \in V \setminus \{r\}$ .
- (8)  $D$  ist ein Baum, besitzt einen Knoten  $r$  mit  $\deg^-(r) = 0$  und erfüllt  $\deg^-(v) = 1$  für alle  $v \in V \setminus \{r\}$ .
- (9)  $D$  enthält keinen Kreis, einen Knoten  $r$  mit  $\deg^-(r) = 0$  und erfüllt  $\deg^-(v) = 1$  für alle  $v \in V \setminus \{r\}$ .  $\triangle$

## 5.2 Optimale Bäume und Wälder

Das Problem, in einem Graphen mit Kantengewichten einen aufspannenden Baum minimalen Gewichts oder einen Wald maximalen Gewichts zu finden, haben wir bereits in (3.11) eingeführt. Beide Probleme sind sehr effizient lösbar und haben vielfältige Anwendungen. Umfassende Überblicke über die Geschichte dieser Probleme, ihre Anwendungen und die bekannten Lösungsverfahren geben die Aufsätze Graham and Hell (1982) und Nešetřil and Nešetřilová (2012) sowie Kapitel 50 des Buches von Schrijver (2003), volume B.

Wir wollen hier jedoch nur einige dieser Lösungsmethoden besprechen. Zunächst wollen wir uns überlegen, dass die beiden Probleme auf sehr direkte Weise äquivalent sind. Angenommen wir haben einen Algorithmus zur Lösung eines Maximalwald-Problems, und wir wollen in einem Graphen  $G = (V, E)$  mit Kantengewichten  $c_e$ ,  $e \in E$ , einen minimalen aufspannenden Baum finden, dann gehen wir wie folgt vor. Wir setzen

$$M := \max\{|c_e| \mid e \in E\} + 1,$$

$$c'_e := M - c_e$$

und bestimmen einen maximalen Wald  $W$  in  $G$  bezüglich der Gewichtsfunktion  $c'$ . Falls  $G$  zusammenhängend ist, ist  $W$  ein aufspannender Baum, denn andernfalls gäbe es eine Kante  $e \in E$ , so dass  $W' := W \cup \{e\}$  ein Wald ist, und wegen  $c'_e > 0$ , wäre  $c'(W') > c'(W)$ . Aus der Definition von  $c'$  folgt direkt, dass  $W$  ein minimaler aufspannender Baum von  $G$  bezüglich  $c$  ist. Ist  $W$  nicht zusammenhängend, so ist auch  $G$  nicht zusammenhängend, also existiert kein aufspannender Baum.

Haben wir einen Algorithmus, der einen minimalen aufspannenden Baum in einem Graphen findet, und wollen wir einen maximalen Wald in einem Graphen  $G = (V, E)$  mit Kantengewichten  $c_e$ ,  $e \in E$ , bestimmen, so sei  $K_n = (V, E_n)$  der vollständige Graph mit  $n = |V|$  Knoten und folgenden Kantengewichten

$$c'_e := -c_e \text{ für alle } e \in E \text{ mit } c_e > 0,$$

$$c'_e := M \text{ für alle } e \in E_n \setminus \{e \in E \mid c_e > 0\},$$

## 5 Bäume und Wege

wobei wir z. B. setzen

$$M := n \cdot (\max\{c_e \mid e \in E\} + 1).$$

Ist  $B$  ein minimaler aufspannender Baum von  $K_n$  bezüglich der Kantengewichte  $c'$ , dann ist offenbar aufgrund unserer Konstruktion  $W := B \setminus \{e \in B \mid c'_e = M\}$  ein Wald in  $G$  maximalen Gewichts.

Der folgende sehr einfache Algorithmus findet einen maximalen Wald.

### (5.7) Algorithmus GREEDY-MAX.

**Eingabe:** Graph  $G = (V, E)$  mit Kantengewichten  $c(e)$  für alle  $e \in E$ .

**Ausgabe:** Wald  $W \subseteq E$  mit maximalem Gewicht  $c(W)$ .

1. (Sortieren): Ist  $k$  die Anzahl der Kanten von  $G$  mit positivem Gewicht, so nummeriere diese  $k$  Kanten, so dass gilt  $c(e_1) \geq c(e_2) \geq \dots \geq c(e_k) > 0$ .
2. Setze  $W := \emptyset$ .
3. FOR  $i = 1$  TO  $k$  DO:  
Falls  $W \cup \{e_i\}$  keinen Kreis enthält, setze  $W := W \cup \{e_i\}$ .
4. Gib  $W$  aus. △

**(5.8) Satz.** *Der Algorithmus GREEDY-MAX arbeitet korrekt.* △

**Beweis.** Hausaufgabe! □

Versuchen Sie, einen direkten Beweis für die Korrektheit von Algorithmus (5.7) zu finden. Im nachfolgenden Teil dieses Abschnitts und in späteren Kapiteln werden wir Sätze angeben, aus denen Satz (5.8) folgt.

Der obige Algorithmus heißt „Greedy-Max“ („greedy“ bedeutet „gierig“ oder „gefräßig“), weil er versucht, das bezüglich der Zielfunktionskoeffizienten jeweils „Beste“ zu nehmen, das im Augenblick noch verfügbar ist. Wir werden später noch andere Algorithmen vom Greedy-Typ kennenlernen, die bezüglich anderer Kriterien das „Beste“ wählen.

Der Greedy-Algorithmus funktioniert auf analoge Weise für das Minimum Spanning Tree Problem.

### (5.9) Algorithmus GREEDY-MIN.

**Eingabe:** Graph  $G = (V, E)$  mit Kantengewichten  $c(e)$  für alle  $e \in E$ .

**Ausgabe:** Maximaler Wald  $T \subseteq E$  mit minimalem Gewicht  $c(T)$ .

1. (Sortieren): Nummeriere die  $m$  Kanten des Graphen  $G$ , so dass gilt:  
 $c(e_1) \leq c(e_2) \leq \dots \leq c(e_m)$ .
2. Setze  $T := \emptyset$ .



3. FOR  $i = 1$  TO  $m$  DO:  
     Falls  $T \cup \{e_i\}$  keinen Kreis enthält, setze  $T := T \cup \{e_i\}$ .
4. Gib  $T$  aus. △

Aus Satz (5.8) und unseren Überlegungen zur Reduktion des Waldproblems auf das Baumproblem und umgekehrt folgt:

**(5.10) Satz.** *Algorithmus (5.9) liefert einen maximalen Wald  $T$  (d. h. für jede Zusammenhangskomponente  $G' = (V', E')$  von  $G$  ist  $T \cap E'$  ein aufspannender Baum), dessen Gewicht  $c(T)$  minimal ist. Ist  $G$  zusammenhängend, so ist  $T$  ein aufspannender Baum von  $G$  minimalen Gewichts  $c(T)$ .* △

Die Laufzeit von Algorithmus (5.7) bzw. (5.9) kann man wie folgt abschätzen. Mit den gängigen Sortierverfahren der Informatik (z. B. HEAP-SORT) kann man die Kanten von  $E$  in  $O(k \log_2 k)$  bzw.  $O(m \log_2 m)$  Schritten in der geforderten Weise ordnen. In Schritt 3 ruft man  $k$ - bzw.  $m$ -mal ein Unterprogramm auf, das überprüft, ob eine Kantenmenge einen Kreis besitzt oder nicht. Durch Benutzung geeigneter Datenstrukturen kann man einen derartigen Aufruf in höchstens  $O(n)$  Schritten abarbeiten. Daraus folgt, dass Schritt 3 in höchstens  $O(mn)$  Schritten ausgeführt werden kann. Dies ist auch die Gesamtlaufzeit des Verfahrens. Mit speziellen „Implementierungstricks“ kann die Laufzeit von Schritt 3 auf  $O(m + n \log n)$  gesenkt und damit die Gesamtlaufzeit sogar auf  $O(m \log m)$  Schritte reduziert werden. In der Literatur wird Algorithmus (5.9) häufig *Kruskal-Algorithmus* genannt.

Einen gewichtsminimalen aufspannenden Baum kann man übrigens auch mit folgendem Verfahren finden.

**(5.11) „Dualer“ Greedy-Algorithmus.**

**Eingabe:** Zusammenhängender Graph  $G = (V, E)$  mit Kantengewichten  $c(e)$  für alle  $e \in E$ .

**Ausgabe:** Aufspannender Baum  $T \subseteq E$  minimalen Gewichts  $c(T)$ .

1. (Sortieren): Numeriere die  $m$  Kanten des Graphen  $G$ , so dass gilt:  
 $c(e_1) \geq c(e_2) \geq \dots \geq c(e_m)$ .
2. Setze  $T := E$ .
3. FOR  $i = 1$  TO  $m$  DO:  
     Falls  $T \setminus \{e_i\}$  zusammenhängend ist, setze  $T := T \setminus \{e_i\}$ .
4. Gib  $T$  aus. △

Der Korrektheitsbeweis für Algorithmus (5.11) bleibt einer Übungsaufgabe überlassen. Wie bereits erwähnt, gibt es eine Vielzahl weiterer Verfahren zur Bestimmung minimaler aufspannender Bäume. Ein gemeinsames Skelett für mehrere dieser Algorithmen kann wie folgt skizziert werden.

**(5.12) Algorithmus META-MST.**

**Eingabe:** Zusammenhängender Graph  $G = (V, E)$  mit Kantengewichten  $c(e)$  für alle  $e \in E$ .

**Ausgabe:** Aufspannender Baum  $T$  von  $G$  minimalen Gewichts.

1. (Initialisierung)  
FOR ALL  $i \in V$  DO: Setze  $V_i := \{i\}$  und  $T_i := \emptyset$ .
2. DO  $|V| - 1$  TIMES:
  - (a) Wähle eine nicht-leere Menge  $V_i$ .
  - (b) Wähle eine Kante  $uv \in E$  mit  $u \in V_i$ ,  $v \in V \setminus V_i$  und  $c(uv) \leq c(pq)$  für alle  $pq \in E$  mit  $p \in V_i$ ,  $q \in V \setminus V_i$ .
  - (c) Bestimme  $j$ , so dass  $v \in V_j$ .
  - (d) Setze  $V_i := V_i \cup V_j$  und  $V_j := \emptyset$ .
  - (e) Setze  $T_i := T_i \cup T_j \cup \{uv\}$  und  $T_j := \emptyset$ .
3. Gib diejenige Kantenmenge  $T_i$  mit  $T_i \neq \emptyset$  aus. △

Algorithmus (5.9) ist ein Spezialfall von Algorithmus (5.12). Überlegen Sie sich wieso!

**(5.13) Satz.** *Algorithmus (5.12) bestimmt einen minimalen aufspannenden Baum.* △

**Beweis.** Wir zeigen durch Induktion über  $p = |T_1| + \dots + |T_n|$ , dass  $G$  einen minimalen aufspannenden Baum  $T$  enthält mit  $T_i \subseteq T$  für alle  $i$ . Ist  $p = 0$ , so ist nichts zu zeigen. Sei  $uv$  eine Kante, die bei einem Durchlauf von Schritt 2 in (b) gewählt wurde. Nach Induktionsvoraussetzung sind alle bisher bestimmten Mengen  $T_i$  in einem minimalen aufspannenden Baum  $T$  enthalten. Gilt  $uv \in T$ , so sind wir fertig. Ist  $uv \notin T$ , so enthält  $T \cup \{uv\}$  einen Kreis. Folglich muss es eine Kante  $rs \in T$  geben mit  $r \in V_i$ ,  $s \in V \setminus V_i$ . Aufgrund unserer Wahl in (b) gilt  $c(uv) \leq c(rs)$ . Also ist  $T := (T \setminus \{rs\}) \cup \{uv\}$  ebenfalls ein minimaler aufspannender Baum und der neue Baum  $T$  erfüllt unsere Bedingungen. Die Korrektheit des Algorithmus folgt aus dem Fall  $p = n - 1$ . □

Die Laufzeit des Algorithmus (5.12) hängt natürlich sehr stark von den Datenstrukturen ab, die man zur Ausführung des Schrittes 2 implementiert. Wir können an dieser Stelle nicht ausführlich auf Implementierungstechniken eingehen und verweisen hierzu auf Mehlhorn (1984), Vol 2, Kapitel IV, Abschnitt 8. Hier wird gezeigt, dass bei geeigneten Datenstrukturen eine Laufzeit von  $O(n \log \log m)$  Schritten erreicht werden kann. Für planare Graphen ergibt sich sogar eine  $O(n)$ -Laufzeit.

Spanning-Tree-Algorithmen werden häufig als Unterprogramme zur Lösung von Travelling-Salesman- und anderen Problemen benötigt. Speziell ist hier eine Implementation dieser Algorithmen für vollständige Graphen erforderlich. Der nachfolgende Algorithmus lässt sich gerade für diesen Fall vollständiger Graphen einfach implementieren und hat sowohl empirisch wie theoretisch günstige Rechenzeiten aufzuweisen. Dieses Verfahren, das offensichtlich ebenfalls eine Spezialisierung von (5.12) ist, wird häufig PRIM-Algorithmus genannt.



## 5 Bäume und Wege

```

* Mode=2 : The matrix of edge weights is given as lower triangular      *
*          matrix. The entries are stored row by row.                    *
*
* Mode=3 : The edge weights are given in an edge list of the           *
*          form: 1st endnode, 2nd endnode, edge weight. Edges which     *
*          are not present are assumed to have 'infinite' weight.       *
*          The input is ended if the first endnode is less than 1.      *
*
*****/
CONST  max_n = 100;           { maximum number of nodes }
      max_n2 = 4950;         { max_n choose 2 = number of edges of K_n }
                                   { to process larger graphs only max_n and
                                   max-n2 have to be changed }

      inf    = maxint;       { infinity }

TYPE   arrn2 = ARRAY[1..max_n2] OF integer;
      arrn  = ARRAY[1..max_n]  OF integer;

VAR    i, j,
      mode,
                                   { input mode of weights:
                                   0 : full matrix
                                   1 : upper triangular matrix
                                   2 : lower triangular matrix
                                   3 : edge list
                                   }
      min,
                                   { minimum distance }
      ind,
                                   { index of entering edge }
      newnode,
                                   { entering tree node }
      t1, t2,
                                   { entering tree edge }
      outnodes,
                                   { number of nodes not in tree }
      c,
      weight,
                                   { weight of tree }
      nchoose2,
      n      : integer;         { number of nodes }
      w      : arrn2;          { vector of weights }
      dope,
                                   { dope vector for index calculations }
      dist,
                                   { shortest distances to non-tree nodes }
      in_t,
                                   { in-tree node of shortest edge }
      out_t  : arrn;           { out-tree node of shortest edge }
                                   { minimum tree is also stored in in_t & out_t}
      connected : boolean;     { true <=> input graph is connected? }
      inp,
                                   { input file }
      outp    : text;          { output file }

BEGIN {MAIN PROGRAM}

{===== Input of complete graph =====}

reset(inp);
rewrite(outp);

{- number of nodes -}
writeln(outp,'Enter number of nodes:');
read(inp,n);

IF (n<1) OR (n>max_n) THEN
  BEGIN
    writeln(outp,'Number of nodes too large or not positive!');
    HALT;
  END;

{- initialize dope vector -}
nchoose2 := (n * (n-1)) DIV 2;
FOR i:=1 TO nchoose2 DO
  w[i] := inf;
dope[1] := -1;
FOR i:=2 TO n DO
  dope[i] := dope[i-1] + n - i;

{- input mode -}
writeln(outp,'Enter input mode:');
read(inp,mode);

```

```

{- edge weights -}
CASE mode OF
  0 : { * full matrix * }
    BEGIN
      FOR i:=1 TO n DO
        FOR j:=1 TO n DO
          BEGIN
            read(inp, c);
            IF i<j THEN w[dope[i]+j] := c;
          END
        END;
      END;
  1 : { * upper triangular matrix * }
    BEGIN
      FOR i:=1 TO nchoose2 DO
        read(inp, w[i]);
      END;
  2 : { * lower triangular matrix * }
    BEGIN
      FOR i:=2 TO n DO
        FOR j:=1 TO i-1 DO
          BEGIN
            read(inp, c);
            w[dope[j]+i] := c;
          END;
        END;
      END;
  3 : { * edge list * }
    BEGIN
      read(inp, i, j, c);
      WHILE (i>0) DO
        BEGIN
          IF (i<1) OR (i>n) OR
             (j<1) OR (j>n)
          THEN BEGIN
            writeln(outp, 'Input error, node out of range!');
            HALT;
          END;
          IF i<j
          THEN w[dope[i]+j] := c
          ELSE w[dope[j]+i] := c;
          read(inp, i, j, c);
        END;
      END;
ELSE: { * invalid mode * }
    BEGIN
      writeln(outp, 'Invalid input mode!');
      HALT;
    END;
END; {OF CASE}

{===== Initialization =====}

connected := true;
outnodes := n-1;
weight := 0;
FOR i:=1 TO outnodes DO
  BEGIN
    in_t[i] := 1;
    out_t[i] := i+1;
    dist[i] := w[i];
  END;

{===== Prim's Algorithm =====}

WHILE (outnodes > 1) AND connected DO
  BEGIN
    {- determine entering node -}
    min := inf;
    ind := 0;
    FOR i:=1 TO outnodes DO
      IF dist[i] < min
      THEN BEGIN
        min := dist[i];
        ind := i;
      END;
    END;
  END;

```

## 5 Bäume und Wege

```

IF ind = 0
THEN connected := false
ELSE BEGIN
  {- augment tree -}
  weight      := weight + min;
  newnode     := out_t[ind];
  t1          := in_t[ind];
  t2          := out_t[ind];
  c           := dist[ind];
  in_t[ind]   := in_t[outnodes];
  out_t[ind]  := out_t[outnodes];
  dist[ind]   := dist[outnodes];
  in_t[outnodes] := t1;
  out_t[outnodes] := t2;
  dist[outnodes] := c;
  outnodes   := outnodes - 1;
  {- update dist[] and in_t[] -}
  FOR i:=1 TO outnodes DO
  BEGIN
    IF newnode < out_t[i]
    THEN c := w[dope[newnode]+out_t[i]]
    ELSE c := w[dope[out_t[i]]+newnode];
    IF c < dist[i]
    THEN BEGIN
      in_t[i] := newnode;
      dist[i] := c;
    END;
  END;
END;
END;
{- insert the last edge -}
IF connected
THEN IF dist[1]>=inf
THEN connected := false
ELSE weight := weight + dist[1];

{===== Output of minimum spanning tree =====}

writeln(outp);
IF NOT connected
THEN writeln(outp,'The graph is disconnected.')
ELSE BEGIN
  writeln(outp,'Minimum spanning tree:');
  writeln(outp,'=====');
  writeln(outp);
  FOR i:=n-1 DOWNTO 1 DO
    writeln(outp, in_t[i]:5, ' - ', out_t[i]:3,
      ' (', dist[i]:1,')');
  writeln(outp);
  writeln(outp,'Weight: ', weight:6);
  writeln(outp);
END;
END.

```

Wir wollen nun noch ein Beispiel angeben, das die Vorgehensweise der Algorithmen (5.9), (5.11) und (5.14) verdeutlicht.

**(5.16) Beispiel.** Wir betrachten den in Abbildung 5.1(a) dargestellten Graphen. Wir wenden Algorithmus (5.9) an. Zunächst sortieren wir die Kanten in nicht absteigender Reihenfolge  $hi, bc, ab, ac, de, ef, eg, be, bf, cf, dg, ad, ae, hf, he, hg$ . In Schritt 5.9 von (5.9) werden die in der Abbildung 5.1(b) gezeichneten Kanten ausgewählt. Den Prim-Algorithmus (5.14) starten wir mit dem Knoten  $w = a$ . Es ergibt sich der in Abbildung 5.1(c) gezeichnete minimale aufspannende Baum.  $\triangle$

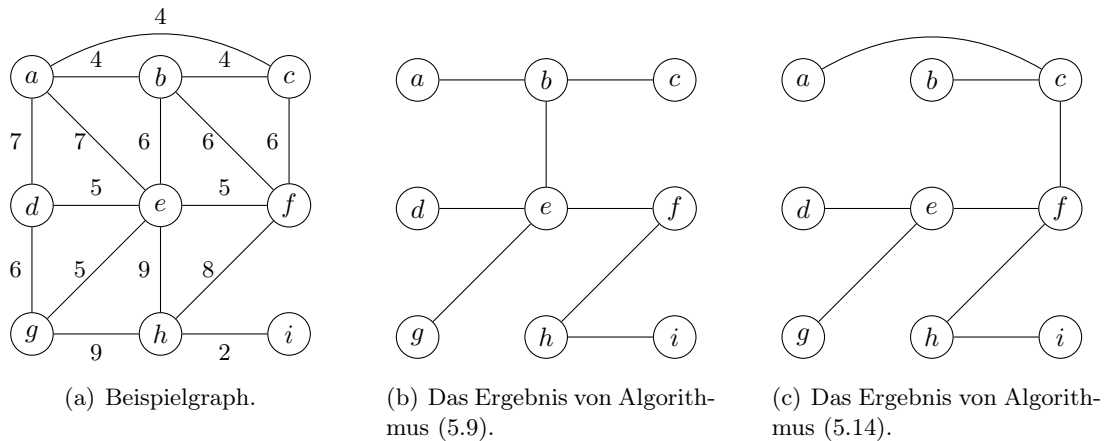


Abbildung 5.1: Ein Beispielgraph und die MSTs, die sich als Ergebnis der Algorithmen (5.9) und (5.14) ergeben.

Wie Beispiel (5.16) zeigt, muss ein minimaler Baum nicht eindeutig bestimmt sein. Überlegen Sie sich bitte, wie man feststellen kann, ob ein minimaler aufspannender Baum eindeutig ist.

Weitergehende Informationen über Branchings und Arboreszenzen (sowie Wälder und Bäume) finden sich im Buch Schrijver (2003) in Part V.

Im Internet finden sich viele „Graph Libraries“ oder „Algorithm Repositories“, in denen fertig implementierte Algorithmen angeboten werden, die das „Minimum Spanning Tree“- oder „Maximum Weighted Branching“-Problem lösen. Einige der Algorithmensammlungen sind kommerziell (und kosten Geld), einige sind frei verfügbar, einige interaktiv abrufbar und viele haben Visualisierungskomponenten. Die Halbwertzeit der Webseiten ist häufig nicht besonders hoch. Es folgen einige Webseiten, die Baum-, Branching- und viele andere Graphenalgorithmen anbieten:

- COIN-OR::LEMON: <http://lemon.cs.elte.hu>
- QuickGraph: <http://quickgraph.codeplex.com>
- The Stony Brook Algorithm Repository: <http://www.cs.sunysb.edu/~algorithm/>
- LEDA: <http://www.algorithmic-solutions.com/leda/index.htm>

## 5.3 Kürzeste Wege

Wir wollen uns nun mit der Aufgabe beschäftigen, in einem Digraphen mit Bogengewichten kürzeste gerichtete Wege zu finden. Wir werden Algorithmen vorstellen, die kürzeste Wege von einem Knoten zu einem anderen oder zu allen anderen oder kürzeste Wege zwischen zwei Knoten finden. Wir beschränken uns auf Digraphen, da derartige Probleme in

ungerichteten Graphen auf einfache Weise auf gerichtete Probleme reduziert werden können. Denn ist ein Graph  $G = (V, E)$  mit Kantenlängen  $c(e) \geq 0$  für alle  $e \in E$  gegeben, so ordnen wir diesem Graphen den Digraphen  $D = (V, A)$  mit  $A = \{(i, j), (j, i) \mid ij \in E\}$  und  $c((i, j)) := c((j, i)) := c(ij)$  zu. Den (ungerichteten)  $[u, v]$ -Wegen in  $G$  entsprechen dann die gerichteten  $(u, v)$ -Wege bzw.  $(v, u)$ -Wege in  $D$  und umgekehrt. Einander entsprechende Wege in  $G$  und  $D$  haben nach Definition gleiche Längen. Also liefert uns ein kürzester  $(u, v)$ -Weg (oder ein kürzester  $(v, u)$ -Weg) in  $D$  einen kürzesten  $[u, v]$ -Weg in  $G$ .

Kürzeste-Wege-Probleme spielen in der kombinatorischen Optimierung eine große Rolle. Es ist daher nicht überraschend, dass es zu diesem Problemkreis eine außerordentlich umfangreiche Literatur und sehr viele Lösungsvorschläge gibt. Wenn man dann noch Variationen hinzunimmt wie: Berechnung längster Wege oder zuverlässiger Wege, von Wegen maximaler Kapazität, der  $k$  kürzesten Wege, von Wegen mit gerader oder ungerader Bogenzahl etc., so liefert das den Stoff einer gesamten Vorlesung. Wir wollen in dieser Vorlesung lediglich drei Algorithmen (für unterschiedliche Spezialfälle) behandeln. Der Leser, der sich für umfassendere Darstellungen interessiert, sei auf die Bücher Ahuja et al. (1993), Krumke and Noltemeier (2005), Lawler (1976), Mehlhorn (1984), Domschke (1972), Schrijver (2003), Syslo et al. (1983) verwiesen. Es werden derzeit immer noch neue Algorithmen oder Modifikationen bekannter Algorithmen entdeckt, die aus theoretischer oder praktischer Sicht schneller als die bekannten Verfahren sind oder sonstige Vorzüge haben.

Ein kürzlich als Preprint erschienener Artikel, siehe Bast et al. (2014), gibt einen Überblick über Kürzeste-Wege-Algorithmen (und deren Varianten), die bei der Wegesuche in Transportnetzen eingesetzt werden. Die vollständige Beschreibung (mit Testergebnissen) eines solchen Verfahrens für länderübergreifende Straßennetzwerke ist in Geisberger et al. (2012) zu finden.

Es gibt keinen Algorithmus zur Bestimmung eines kürzesten  $(s, t)$ -Weges, der nicht (zumindest implizit) auch alle übrigen kürzesten Wege von  $s$  nach  $v$ ,  $s \neq v \neq t$ , berechnet. Die Algorithmen für Kürzeste-Wege-Probleme kann man in zwei Kategorien einteilen, und zwar solche, die negative Bogenlängen zulassen, und solche, die nur nichtnegative Bogenlängen behandeln können. Von jedem der beiden Typen stellen wir einen Vertreter vor. Ferner wollen wir noch einen Algorithmus behandeln, der kürzeste Wege zwischen allen Knoten berechnet.

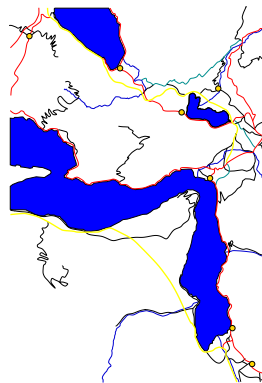
Vermutlich haben sich die Menschen schon in grauer Vorzeit mit der Bestimmung kürzester Wege beschäftigt, um z. B. Transporte zu vereinfachen, den Handel zu erleichtern etc. Mathematik – im heutigen Sinne – wurde dabei sicherlich nicht verwendet. Eines der ältesten (uns bekannten) Wegeprobleme der (belletristischen) Literatur kommt aus einer klassischen Quelle: Friedrich Schillers (1759–1805) Schauspiel „Wilhelm Tell“. Dieser konnte bereits 1291 nicht nur gut schießen, sondern auch optimieren. Und nur mit dieser Kombination konnte er die Schweiz befreien! Tell befindet sich nach dem Apfelschuss am Ufer des Vierwaldstätter Sees unweit des Ortes Altdorf. Er muss unbedingt vor dem Reichsvogt Hermann Gessler die Hohle Gasse in Küssnacht erreichen, siehe Abbildung 5.2(b).

Schiller berichtet:





(a) F. Schiller.



(b) Vierwaldstätter See.



(c) W. Tell.

Abbildung 5.2: Mathematik in der Belletristik: Schillers Schauspiel „Wilhelm Tell“.

Tell. Nennst mir den nächsten Weg nach Arth und Küssnacht  
 Fischer. Die offene Straße zieht sich über Steinen  
 Den kürzern Weg und heimlichern  
 Kann Euch mein Knabe über Lowenz führen.  
 Tell (gibt ihm die Hand). Gott lohn Euch Eure Guttat. Lebet wohl.

Der Fischer löst für Tell in dieser Szene offensichtlich ein graphentheoretisches Optimierungsproblem. In einem Graphen (Wegenetz am Vierwaldstätter See) mit Kantenlängen (Reisezeit) soll der kürzeste Weg zwischen zwei vorgegebenen Punkten (Altdorf und Küssnacht) bestimmt werden. Tell behandelt sogar eine kompliziertere Variante mit einer zusätzlichen Nebenbedingung: Die Summe von „Verhaftungskoeffizienten“ muss unterhalb eines sicheren Grenzwertes bleiben. Man kann dies auch als multikriterielles Optimierungsproblem auffassen (Weglänge und Sicherheit gleichzeitig optimieren). Dies ist ein Aufgabentyp, den wir auch heute noch nicht gut beherrschen. (In der Vorlesung wird mehr dazu berichtet).

### 5.3.1 Ein Startknoten, nichtnegative Gewichte

Das Verfahren, das wir nun darstellen wollen, ist mehrfach entdeckt worden. Es wird allgemein nach Dijkstra (1959) benannt. Wir gehen davon aus, dass ein Digraph  $D = (V, A)$  mit „Gewichten“ bzw. „Längen“ oder „Entfernungen“  $c(a) \geq 0$  für alle  $a \in A$  gegeben ist. Ferner seien ein Startknoten  $s$  und möglicherweise ein Endknoten  $t$  gegeben. Das Verfahren findet einen kürzesten gerichteten Weg von  $s$  zu allen anderen Knoten bzw. einen kürzesten  $(s, t)$ -Weg.

Der Algorithmus wird häufig als *Markierungsmethode* bezeichnet. (Warum, wird aus dem Weiteren klar.) Seine Idee kann man wie folgt beschreiben.

Wir beginnen im Startknoten  $s$ , markieren  $s$  und ordnen  $s$  die *permanente Distanz* Null (= Länge des kürzesten Weges von  $s$  zu sich selbst) zu. Alle übrigen Knoten  $v$

seien unmarkiert, und wir ordnen ihnen als *temporäre Distanz* (= Länge des kürzesten bisher gefundenen  $(s, v)$ -Weges) entweder  $+\infty$  oder die Länge des Bogens  $(s, v)$ , falls dieser in  $D$  existiert, zu. Der unmarkierte Knoten mit der kleinsten temporären Distanz ist dann der Knoten, der am nächsten zu  $s$  liegt. Nennen wir den Knoten  $u$ . Da alle Bogenlängen nicht-negativ sind, ist der Bogen  $(s, u)$  der kürzeste Weg von  $s$  nach  $u$ . Wir markieren daher  $u$  und erklären die temporäre Distanz von  $u$  als *permanent*, weil wir den (global) kürzesten  $(s, u)$ -Weg gefunden haben. Nun bestimmen wir alle Nachfolger  $v$  von  $u$  und vergleichen die temporäre Distanz von  $v$  mit der permanenten Distanz von  $u$  plus der Länge des Bogens  $(u, v)$ . Ist diese Summe kleiner als die bisherige temporäre Distanz, wird sie die neue temporäre Distanz, weil der bisher bekannte Weg von  $s$  nach  $v$  länger ist als der Weg von  $s$  über  $u$  nach  $v$ . Wir wählen nun wieder eine kleinste der temporären Distanzen, erklären sie als permanent, da der bisher gefundene Weg durch Umwege über andere Knoten nicht verkürzt werden kann, markieren den zugehörigen Knoten und fahren so fort bis entweder alle Knoten oder der gesuchte Endknoten  $t$  markiert sind. Etwas formaler kann man diesen Algorithmus wie folgt aufschreiben.

### (5.17) DIJKSTRA-Algorithmus.

**Eingabe:** Digraph  $D = (V, A)$ , Gewichte  $c(a) \geq 0$  für alle  $a \in A$ , ein Knoten  $s \in V$  (und möglicherweise ein Knoten  $t \in V \setminus \{s\}$ ).

**Ausgabe:** Kürzeste gerichtete Wege von  $s$  nach  $v$  für alle  $v \in V$  und ihre Länge (bzw. ein kürzester  $(s, t)$ -Weg).

**Datenstrukturen:** DIST( $v$ ) Länge des kürzesten  $(s, v)$ -Weges  
 VOR( $v$ ) Vorgänger von  $v$  im kürzesten  $(s, v)$ -Weg

1. Setze:

$$\begin{aligned} \text{DIST}(s) &:= 0 \\ \text{DIST}(v) &:= c((s, v)) \quad \forall v \in V \text{ mit } (s, v) \in A \\ \text{DIST}(v) &:= +\infty \quad \forall v \in V \text{ mit } (s, v) \notin A \\ \text{VOR}(v) &:= s \quad \forall v \in V \setminus \{s\} \end{aligned}$$

Markiere  $s$ , alle übrigen Knoten seien unmarkiert.

2. Bestimme einen unmarkierten Knoten  $u$ , so dass

$$\text{DIST}(u) = \min\{\text{DIST}(v) \mid v \text{ unmarkiert}\}.$$

Markiere  $u$ .

(Falls  $u = t$ , gehe zu 5.)

3. Für alle unmarkierten Knoten  $v$  mit  $(u, v) \in A$  führe aus:

Falls  $\text{DIST}(v) > \text{DIST}(u) + c((u, v))$   
 setze:  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ .

4. Sind noch nicht alle Knoten markiert, gehe zu 2.

5. Für alle markierten Knoten  $v$  ist  $\text{DIST}(v)$  die Länge eines kürzesten  $(s, v)$ -Weges. Falls  $v$  markiert ist und  $\text{DIST}(v) < +\infty$ , so ist  $\text{VOR}(v)$  der Vorgänger von  $v$  in

einem kürzesten  $(s, v)$ -Weg, d. h. durch Rückwärtsgehen bis  $s$  kann ein kürzester  $(s, v)$ -Weg bestimmt werden. (Brechen wir das Verfahren nicht in Schritt 2 ab und gilt am Ende  $\text{DIST}(v) = +\infty$ , so heißt das, dass es in  $D$  keinen  $(s, v)$ -Weg gibt.)  $\triangle$

Zur Notationsvereinfachung für den nachfolgenden Beweis bezeichnen wir mit  $\text{DIST}_k(v)$  den Wert der in (5.17) berechneten Distanzfunktion nach dem  $k$ -ten Durchlaufen der Schritte 2, 3 und 4. Für den DIJKSTRA-Algorithmus gilt aufgrund der Auswahlvorschrift nach der  $k$ -ten Markierungsphase Folgendes: Sind die Knoten in der Reihenfolge  $v_1, v_2, \dots, v_k$  markiert worden, so gilt  $\text{DIST}_k(v_1) \leq \dots \leq \text{DIST}_k(v_k) \leq \text{DIST}_k(v)$  für alle bisher unmarkierten Knoten  $v$ .

**(5.18) Satz.** *Der Dijkstra-Algorithmus arbeitet korrekt.*  $\triangle$

**Beweis.** Wir zeigen durch Induktion über die Anzahl  $k$  markierter Knoten Folgendes: Ist  $v$  markiert, so enthält  $\text{DIST}_k(v)$  die Länge eines kürzesten  $(s, v)$ -Weges; ist  $v$  unmarkiert, so enthält  $\text{DIST}_k(v)$  die Länge eines kürzesten  $(s, v)$ -Weges, wobei als innere Knoten nur markierte Knoten zugelassen sind. (Falls  $\text{DIST}_k(v) = +\infty$ , so wird dies als Nichtexistenz eines  $(s, v)$ -Weges bzw. eines  $(s, v)$ -Weges über markierte innere Knoten interpretiert). Hieraus folgt offenbar die Behauptung.

Ist nur ein Knoten (also  $s$ ) markiert, so ist unsere Behauptung aufgrund der Definition in Schritt 1 korrekt. Wir nehmen nun an, dass die Behauptung richtig ist für  $k$  markierte Knoten und dass das Verfahren in Schritt 2 einen  $(k+1)$ -ten Knoten, sagen wir  $u$ , markiert und Schritt 3 durchlaufen hat. Nach Induktionsvoraussetzung ist  $\text{DIST}_k(u)$  die Länge eines kürzesten  $(s, u)$ -Weges, der als innere Knoten nur die ersten  $k$  markierten Knoten benutzen darf. Gäbe es einen kürzeren gerichteten Weg, sagen wir  $P$ , von  $s$  nach  $u$ , so müsste dieser einen Bogen von einem markierten Knoten zu einem bisher nicht markierten Knoten enthalten. Sei  $(v, w)$  der erste derartige Bogen auf dem Weg  $P$ . Der Teilweg  $\bar{P}$  des Weges  $P$  von  $s$  nach  $w$  ist also ein  $(s, w)$ -Weg, dessen innere Knoten markiert sind. Folglich gilt nach Induktionsvoraussetzung  $\text{DIST}_{k+1}(w) \leq c(\bar{P})$ . Aus  $\text{DIST}_{k+1}(u) \leq \text{DIST}_{k+1}(w)$  und der Nichtnegativität der Bogenlängen folgt  $\text{DIST}_{k+1}(u) \leq c(\bar{P}) \leq c(P)$ , ein Widerspruch.

Es bleibt noch zu zeigen, dass für derzeit unmarkierte Knoten  $v$  der Wert  $\text{DIST}_{k+1}(v)$  die Länge eines kürzesten  $(s, v)$ -Weges ist, der nur markierte innere Knoten enthalten darf. Im Update-Schritt 3 wird offenbar die Länge eines  $(s, v)$ -Weges über markierte Knoten verschieden von  $u$  verglichen mit der Länge eines  $(s, v)$ -Weges über markierte Knoten, der als vorletzten Knoten den Knoten  $u$  enthält. Angenommen es gibt einen  $(s, v)$ -Weg  $P$  über markierte Knoten (inclusive  $u$ ), dessen vorletzter Knoten  $w$  verschieden von  $u$  ist und dessen Länge geringer ist als die kürzeste Länge der oben betrachteten Wege. Da  $\text{DIST}_{k+1}(w)$  die Länge eines kürzesten  $(s, w)$ -Weges ist und es einen solchen, sagen wir  $P'$ , gibt, der nur markierte Knoten enthält, die verschieden von  $u$  sind ( $w$  wurde vor  $u$  markiert), kann der  $(s, w)$ -Weg auf  $P$  nicht kürzer als  $P'$  sein, also ist  $P$  nicht kürzer als die Länge von  $P' \cup \{(w, v)\}$ . Widerspruch.  $\square$

In der Datenstruktur VOR merken wir uns zu jedem Knoten  $v$  seinen Vorgänger in einem kürzesten  $(s, v)$ -Weg. Einen kürzesten  $(s, v)$ -Weg erhält man also in umgekehrter

Reihenfolge durch die Knotenfolge

$$v, \text{VOR}(v), \text{VOR}(\text{VOR}(v)), \dots, \text{VOR}(\text{VOR}(\dots \text{VOR}(v) \dots)).$$

Durch VOR ist offenbar eine Arboreszenz mit Wurzel  $s$  in  $D$  definiert. Daraus folgt sofort:

**(5.19) Satz.** *Sei  $D = (V, A)$  ein Digraph mit nichtnegativen Bogengewichten und  $s \in V$ , dann gibt es eine Arboreszenz  $B$  mit Wurzel  $s$ , so dass für jeden Knoten  $v \in V$ , für den es einen  $(s, v)$ -Weg in  $D$  gibt, der (eindeutig bestimmte) gerichtete Weg in  $B$  von  $s$  nach  $v$  ein kürzester  $(s, v)$ -Weg ist.  $\triangle$*

An dieser Stelle sei darauf hingewiesen, dass der PRIM-Algorithmus (5.14) und der DIJKSTRA-Algorithmus (5.17) (im Wesentlichen) identische Algorithmen sind. Sie unterscheiden sich lediglich bezüglich einer Gewichtstransformation. In Schritt 3 von (5.14) wird  $\min\{c(e) \mid e \in \delta(W)\}$  gesucht, in Schritt 2 von (5.17) wird auch ein derartiges Minimum gesucht, jedoch sind vorher in Schritt 3 die Gewichte der Bögen des Schnittes modifiziert worden.

Den DIJKSTRA-Algorithmus kann man ohne Schwierigkeiten so implementieren, dass seine Laufzeit  $O(|V|^2)$  beträgt. Bei Digraphen mit geringer Bogenzahl kann die Laufzeit durch Benutzung spezieller Datenstrukturen beschleunigt werden, siehe hierzu z. B. Ahuja et al. (1993) oder Schrijver (2003).

### 5.3.2 Ein Startknoten, beliebige Gewichte

Das Problem, einen kürzesten Weg in einem Digraphen mit beliebigen Bogengewichten zu bestimmen, ist trivialerweise äquivalent zum Problem, einen längsten Weg in einem Digraphen mit beliebigen Bogengewichten zu finden. Gäbe es für das letztere Problem einen polynomialen Algorithmus, so könnte man in polynomialer Zeit entscheiden, ob ein Digraph einen gerichteten hamiltonschen Weg enthält. Dieses Problem ist aber  $\mathcal{NP}$ -vollständig, also ist das Kürzester-Weg-Problem für beliebige Gewichte  $\mathcal{NP}$ -schwer.

Andererseits kann man dennoch in beliebig gewichteten Digraphen kürzeste Wege finden, wenn die negativen Gewichte „gut verteilt“ sind oder der Digraph bestimmte Eigenschaften hat. Der DIJKSTRA-Algorithmus funktioniert bei negativen Gewichten nicht (im Induktionsschritt des Beweises von (5.18) wurde von der Nichtnegativität explizit Gebrauch gemacht). Wir wollen nun auf ein Verfahren eingehen, das unabhängig voneinander von Moore (1959) und Bellman (1958) vorgeschlagen wurde. Zu diesem Verfahren gibt es eine Vielzahl von Verbesserungsvorschlägen (siehe hierzu z. B. Lawler (1976), Syslo et al. (1983), Glover et al. (1985)). Die Idee hinter diesem Verfahren lässt sich wie folgt beschreiben. Wir wollen vom Startknoten  $s$  aus zu allen anderen Knoten  $v$  einen kürzesten  $(s, v)$ -Weg bestimmen. Wir initialisieren  $\text{DIST}(v)$  wieder mit  $+\infty$  oder mit  $c((s, v))$  ( $\text{DIST}(v)$  enthält also die Länge des kürzesten zur Zeit bekannten  $(s, v)$ -Weges mit einer bestimmten Eigenschaft) und setzen wie in (5.17)  $\text{VOR}(v) = s$ . Nun versuchen wir, die Distanzen  $\text{DIST}(v)$  sukzessive zu reduzieren. Finden wir einen Bogen  $(u, v)$  mit  $\text{DIST}(u) + c((u, v)) < \text{DIST}(v)$ , so setzen wir  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ . Wir führen diese Iteration so lange fort, bis kein Wert  $\text{DIST}(u)$  mehr

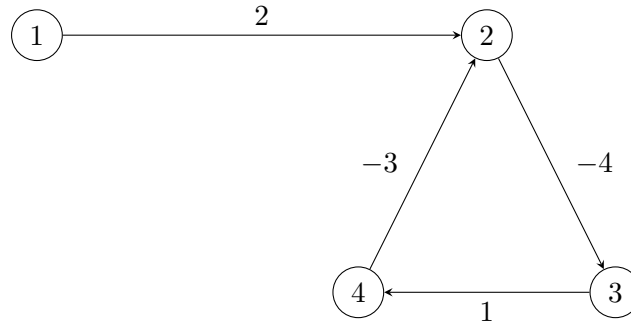


Abbildung 5.3: Digraph mit negativen Kantengewichten

reduziert werden kann. Die verschiedenen Versionen des Moore-Bellman-Algorithmus unterscheiden sich durch die Art, wie diese Basisiteration ausgeführt wird (d. h. in welcher Reihenfolge die Knoten und Bögen (u. U. mehrfach) abgearbeitet werden).

Wir wollen uns zunächst überlegen, unter welchen Umständen der MOORE-BELLMAN-Algorithmus bei allgemeinen Gewichten nicht funktioniert. Wir betrachten den Digraphen aus Abbildung 5.3 mit den dort eingetragenen Gewichten. Wir initialisieren mit  $\text{DIST}(1) = 0$ ,  $\text{DIST}(2) = 2$ ,  $\text{DIST}(3) = \text{DIST}(4) = +\infty$ ,  $\text{VOR}(i) = 1$ ,  $i = 1, 2, 3, 4$ . Wir stellen fest, dass  $\text{DIST}(3) > \text{DIST}(2) + c((2, 3)) = -2$ , und setzen  $\text{DIST}(3) = -2$ ,  $\text{VOR}(3) = 2$ . Wir setzen analog  $\text{DIST}(4) = \text{DIST}(3) + c((3, 4)) = -1$ ,  $\text{VOR}(4) = 3$ . Nun gilt  $\text{DIST}(2) = 2 > \text{DIST}(4) + c((4, 2)) = -4$ , also setzen wir  $\text{DIST}(2) = -4$ . Was ist passiert? Der kürzeste Weg von 1 nach 2 besteht offensichtlich aus dem Bogen (1, 2) und hat die Länge 2. Wenden wir unser Verfahren an, so stellen wir fest, dass wir von 1 nach 4 mit der Weglänge  $-1$  gelangen können. Dieser Weg enthält den Knoten 2. Aber nun können wir von 4 nach 2 zurückgehen, und unsere gerichtete Kette von 1 nach 2, nach 3, nach 4 und wieder zu 2 hat eine geringere Länge als der direkte Weg von 1 nach 2. Der Grund für diese Wegverkürzung liegt darin, dass wir einen Kreis, hier den Kreis (2, 3, 4), entdeckt haben, dessen Gesamtlänge negativ ist. Laufen wir nun noch einmal durch diesen Kreis, so können wir die „Weglänge“ noch weiter verkürzen, d. h. unser Verfahren wird eine immer kleinere „Weglänge“ produzieren und nicht enden. Nennen wir einen gerichteten Kreis  $C$  *negativ*, wenn sein Gewicht  $c(C)$  negativ ist, so zeigt die obige Überlegung, dass negative Kreise in einem Digraphen zum Scheitern des Verfahrens führen. Hat ein Digraph überhaupt keinen gerichteten Kreis, ist er also azyklisch, so gibt es insbesondere keine negativen Kreise, und das MOORE-BELLMAN-Verfahren funktioniert.

#### (5.20) MOORE-BELLMAN-Algorithmus für azyklische Digraphen.

**Eingabe:** Azyklischer Digraph  $D = (V, A)$ , Gewichte  $c(a)$  für alle  $a \in A$  (beliebige negative Gewichte sind zugelassen), ein Knoten  $s \in V$ .

**Ausgabe:** Für jeden Knoten  $v \in V$  ein kürzester  $(s, v)$ -Weg und seine Länge.

**Datenstrukturen:**  $\text{DIST}(v)$ ,  $\text{VOR}(v)$  für alle  $v \in V$ . O. B. d. A. nehmen wir an, dass  $V = \{1, 2, \dots, n\}$  gilt und alle Bögen die Form  $(u, v)$  mit  $u < v$  haben.

## 5 Bäume und Wege

1. Setze:

$$\text{DIST}(v) := \begin{cases} 0 & \text{falls } s = v \\ +\infty & \text{falls } s \neq v \text{ und } (s, v) \notin A \\ c((s, v)) & \text{andernfalls} \end{cases}$$

$$\text{VOR}(v) := s.$$

2. FOR  $v = s + 2$  TO  $n$  DO:

3. FOR  $u = s + 1$  TO  $v - 1$  DO:

Falls  $(u, v) \in A$  und  $\text{DIST}(u) + c((u, v)) < \text{DIST}(v)$

setze:  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ .

END 3.

END 2.

4. Falls  $\text{DIST}(v) < +\infty$ , so enthält  $\text{DIST}(v)$  die Länge des kürzesten gerichteten Weges von  $s$  nach  $v$ , und aus  $\text{VOR}$  kann ein kürzester  $(s, v)$ -Weg entnommen werden. Falls  $\text{DIST}(v) = +\infty$ , so existiert in  $D$  kein  $(s, v)$ -Weg.  $\triangle$

**(5.21) Satz.** *Algorithmus (5.20) funktioniert für beliebige azyklische Digraphen  $D$  und beliebige Bogengewichte.*  $\triangle$

**Beweis.** Nach Voraussetzung haben alle Bögen in  $D$  die Form  $(u, v)$  mit  $u < v$ . Folglich gibt es in  $D$  keinen  $(s, v)$ -Weg für  $v < s$ . Nach Definition ist die Länge eines  $(s, s)$ -Weges gleich Null. Ferner enthält jeder  $(s, v)$ -Weg mit  $v > s$  nur innere Knoten  $u$  mit  $s < u < v$ . Es gibt höchstens einen  $(s, s+1)$ -Weg, nämlich den Bogen  $(s, s+1)$ , falls er in  $D$  existiert, also enthält  $\text{DIST}(v)$  für  $1 \leq v \leq s+1$  die Länge eines kürzesten  $(s, v)$ -Weges in  $D$ .

Ist  $v > s+1$ , so folgt durch Induktion über die Schleifenindizes der Schleife 2, dass  $\text{DIST}(u)$  die Länge eines kürzesten  $(s, u)$ -Weges für  $1 \leq u \leq v$  enthält. Aus formalen Gründen lassen wir Schleife 2 mit  $v = s+1$  beginnen. Dadurch wird kein Wert  $\text{DIST}(u)$  in Schritt 3 geändert. Für  $v = s+1$  ist somit nach obiger Bemerkung die Behauptung korrekt. Sei also die Behauptung für  $v$  richtig und betrachten wir den Knoten  $v+1$ . Nach Induktionsvoraussetzung enthält  $\text{DIST}(u)$ ,  $1 \leq u \leq v$ , die Länge eines kürzesten  $(s, u)$ -Weges. Da ein  $(s, v+1)$ -Weg entweder von  $s$  direkt nach  $v+1$  führt (das Gewicht dieses Bogens ist gegenwärtig in  $\text{DIST}(v+1)$  gespeichert) oder zunächst zu Zwischenknoten  $u$  im Intervall  $s < u \leq v$  und dann auf einen Bogen nach  $v+1$  führt, ist also die Länge des kürzesten  $(s, v+1)$ -Weges gegeben durch das Minimum der folgenden beiden Werte:

$$c((s, v+1)) = \text{DIST}(v+1),$$

$$\text{Länge des kürzesten } (s, u)\text{-Weges} + c((u, v+1)) = \text{DIST}(u) + c((u, v+1)).$$

Dieses Minimum wird offenbar bei Ausführung der Schleife 3 für  $v+1$  berechnet. Daraus folgt die Behauptung.  $\square$

Da das Verfahren (5.20) im wesentlichen aus zwei Schleifen besteht, die beide über maximal  $n - 2$  Indizes laufen, ist die Laufzeit des Verfahrens  $O(n^2)$ .

Wir geben nun den MOORE-BELLMAN-Algorithmus für beliebige Digraphen in zwei verschiedenen Varianten an:

**(5.22) MOORE-BELLMAN-Algorithmus.**

**Eingabe:** Digraph  $D = (V, A)$ , Gewichte  $c(a)$  für alle  $a \in A$  (können auch negativ sein), ein Knoten  $s \in V$ .

**Ausgabe:** Für jeden Knoten  $v \in V$  ein kürzester  $(s, v)$ -Weg und seine Länge. Korrektheit des Output ist nur dann garantiert, wenn  $D$  keinen negativen Kreis enthält.

**Datenstrukturen:**  $\text{DIST}(v)$ ,  $\text{VOR}(v)$  für alle  $v \in V$  (wie in Algorithmus (5.17))

**YEN-VARIANTE** Wir nehmen hier zur Vereinfachung der Darstellung o. B. d. A. an, dass  $V = \{1, \dots, n\}$  und  $s = 1$  gilt.

1. Setze:

$$\begin{aligned} \text{DIST}(s) &:= 0 \\ \text{DIST}(v) &:= c((s, v)) \quad \text{falls } (s, v) \in A \\ \text{DIST}(v) &:= \infty \quad \text{sonst} \\ \text{VOR}(v) &:= s \quad \forall v \in V. \end{aligned}$$

2. FOR  $m = 0$  TO  $n - 2$  DO:

3. Falls  $m$  gerade: FOR  $v = 2$  TO  $n$  DO:

4. FOR  $u = 1$  TO  $v - 1$  DO:

Falls  $(u, v) \in A$  und  $\text{DIST}(u) + c((u, v)) < \text{DIST}(v)$ ,  
setze:  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ .

END 4.

END 3.

5. Falls  $m$  ungerade: FOR  $v = n - 1$  TO 1 BY  $-1$  DO:

6. FOR  $u = n$  TO  $v + 1$  BY  $-1$  DO:

Falls  $(u, v) \in A$  und  $\text{DIST}(u) + c((u, v)) < \text{DIST}(v)$ ,  
setze:  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ .

END 6.

END 5.

END 2.

Gehe zu 7.

**D'ESOPO-PAPE-VARIANTE**

1'. Setze:

$$\begin{aligned}\text{DIST}(s) &:= 0 \\ \text{DIST}(v) &:= \infty \quad \forall v \in V \setminus \{s\} \\ \text{VOR}(v) &:= s \quad \forall v \in V.\end{aligned}$$

2'. Initialisiere eine Schlange  $Q$  und setze  $s$  in  $Q$ .3'. Hole das erste Element aus der Schlange, sagen wir  $u$ .4'. Für alle Bögen  $(u, v)$ , die in  $u$  beginnen, führe aus:5'. Falls  $\text{DIST}(u) + c((u, v)) < \text{DIST}(v)$ a) Setze  $\text{DIST}(v) := \text{DIST}(u) + c((u, v))$  und  $\text{VOR}(v) := u$ ,b) Falls  $v$  noch nicht in  $Q$  war, setze  $v$  an das Ende von  $Q$ ,c) Falls  $v$  schon in  $Q$  war, aber gegenwärtig nicht in  $Q$  ist, setze  $v$  an den Anfang von  $Q$ .

END 4'

6'. Ist die Schlange nicht leer, gehe zu 3', andernfalls zu 7.

7. Falls  $\text{DIST}(v) < +\infty$ , so enthält  $\text{DIST}(v)$  die Länge eines kürzesten  $(s, v)$ -Weges, und aus  $\text{VOR}(v)$  kann wie üblich ein kürzester  $(s, v)$ -Weg rekonstruiert werden. Ist  $\text{DIST}(v) = +\infty$ , so gibt es in  $D$  keinen  $(s, v)$ -Weg.  $\triangle$ 

Es ist intuitiv einsichtig, dass das MOORE-BELLMAN-Verfahren ein korrektes Ergebnis liefert, falls keine negativen Kreise vorliegen. Ebenso leuchtet ein, dass die D'ESOPO-PAPE-Variante eine Spezialisierung dieses Verfahrens ist mit einer konkreten Angabe der Bearbeitungsreihenfolge. Wir wollen nun noch die Korrektheit der YEN-Variante vorführen.

**(5.23) Satz.** *Die YEN-Variante des MOORE-BELLMAN-Verfahrens arbeitet korrekt, falls  $D$  keinen negativen gerichteten Kreis enthält.*  $\triangle$

**Beweis.** Wir geben dem Vektor  $\text{DIST}$  eine Interpretation, aus der die Korrektheit einfach folgt. Wir haben in (5.22) angenommen, dass  $s = 1$  gilt und die Knoten mit  $1, 2, \dots, n$  bezeichnet sind. Wir nennen einen Bogen  $(u, v)$  einen *Aufwärtsbogen*, falls  $u < v$  gilt, andernfalls heißt  $(u, v)$  *Abwärtsbogen*. Wir sprechen von einem *Richtungswechsel*, wenn in einem  $(s, v)$ -Weg ein Abwärtsbogen auf einen Aufwärtsbogen folgt oder umgekehrt. Da  $s = 1$ , ist der erste Bogen immer ein Aufwärtsbogen, also ist der erste Richtungswechsel immer aufwärts nach abwärts. Um einfacher argumentieren zu können, bezeichnen wir mit  $\text{DIST}(v, m)$  den Inhalt des Vektors  $\text{DIST}(v)$  nach Beendigung der  $m$ -ten Iteration der äußeren Schleife.



Wir behaupten nun:

$$\text{DIST}(v, m) = \min\{c(W) \mid W \text{ ist ein gerichteter } (1, v)\text{-Weg mit} \\ \text{höchstens } m \text{ Richtungswechseln}\}, \quad 0 \leq m \leq n - 2.$$

Da ein  $(1, v)$ -Weg höchstens  $n - 1$  Bögen und somit höchstens  $n - 2$  Richtungswechsel besitzt, folgt der Satz aus dem Beweis unserer Behauptung.

Wir beweisen unsere Behauptung durch Induktion über  $m$ . Für  $m = 0$  ist der Durchlauf der Schritte 3 und 4 nichts anderes als Algorithmus 5.20 (angewendet auf  $s = 1$  und den azyklischen Digraphen der Aufwärtsbögen, der keinen gerichteten und somit auch keinen gerichteten negativen Kreis enthält), dessen Korrektheit wir in Satz (5.21) bewiesen haben.  $\text{DIST}(v, 0)$  enthält somit die Länge des kürzesten  $(1, v)$ -Weges ohne Richtungswechsel, die Behauptung für  $m = 0$  ist also richtig.

Nehmen wir nun an, dass unsere Behauptung für  $m \geq 0$  richtig ist und dass wir Schleife 2 zum  $(m+1)$ -ten Male durchlaufen haben. Wir müssen zwei Fälle unterscheiden:  $m + 1$  gerade oder ungerade. Wir führen den Fall  $m + 1$  ungerade vor, der andere Fall folgt analog. Die Menge der  $(1, v)$ -Wege mit höchstens  $m + 1$  Richtungswechseln besteht aus folgenden Wegen:

- (a)  $(1, v)$ -Wege mit höchstens  $m$  Richtungswechseln,
- (b)  $(1, v)$ -Wege mit genau  $m + 1$  Richtungswechseln.

Die Minimallänge der Wege in (a) kennen wir nach Induktionsvoraussetzung bereits, sie ist  $\text{DIST}(v, m)$ .

Wir haben angenommen, dass  $s = 1$  gilt, also ist der erste Bogen eines jeden Weges ein Aufwärtsbogen. Für einen  $(1, v)$ -Weg mit  $m + 1$  Richtungswechseln und  $m + 1$  ungerade ist daher der letzte Bogen ein Abwärtsbogen.

Zur Bestimmung des Minimums in (b) führen wir eine weitere Induktion über  $u = n, n - 1, \dots, v + 1$  durch. Da jeder Weg, der in  $n$  endet mit einem Aufwärtsbogen aufhört, gibt es keinen  $(1, n)$ -Weg mit genau  $m + 1$  Richtungswechseln, also gilt  $\text{DIST}(n, m) = \text{DIST}(n, m + 1)$ .

Nehmen wir nun an, dass wir wissen, dass  $\text{DIST}(w, m + 1)$  für  $n \geq w \geq u > v + 1$  die Länge eines kürzesten  $(1, w)$ -Weges mit höchstens  $m + 1$  Richtungswechseln ist. Zur Bestimmung der Länge eines kürzesten  $(1, u - 1)$ -Weges mit höchstens  $m + 1$  Richtungswechseln müssen wir die Länge eines kürzesten  $(1, u - 1)$ -Weges mit höchstens  $m$  Richtungswechseln (diese ist in  $\text{DIST}(u - 1, m)$  gespeichert) vergleichen mit der Länge eines kürzesten  $(1, u - 1)$ -Weges mit genau  $m + 1$  Richtungswechseln.

Sei nun  $P$  ein kürzester  $(1, u - 1)$ -Weg mit genau  $m + 1$  Richtungswechseln. Sei  $r$  der Knoten auf  $P$ , bei dem der letzte Richtungswechsel erfolgt. Da der letzte Bogen auf  $P$ , weil  $m + 1$  ungerade ist, ein Abwärtsbogen ist, gilt  $u \leq r \leq n$ . Der Weg  $P_r$  auf  $P$  von 1 bis  $r$  ist ein gerichteter Weg mit  $m$  Richtungswechseln. Also gilt nach Induktionsvoraussetzung  $c(P_r) \geq \text{DIST}(r, m)$ . Für alle Knoten  $s$ , die auf  $P$  zwischen  $r$  und  $u - 1$  liegen (also  $u - 1 < s < r$ ), ist der Weg  $P_s$  auf  $P$  von 1 bis  $s$  ein gerichteter  $(1, s)$ -Weg mit genau  $m + 1$  Richtungswechseln. Somit ist nach Induktionsvoraussetzung

$c(P_s) \geq \text{DIST}(s, m + 1)$ . Ist  $t$  der vorletzte Knoten auf  $P$ , also  $(t, u - 1) \in P$ , so ist  $c(P) = c(P_t) + c(t, u - 1) \geq \text{DIST}(t, m + 1) + c(t, u - 1)$ . Der letzte Wert geht in die Minimumsbildung in Schritt 6 ein. Also wird in Schritt 6 der kürzeste aller  $(1, u - 1)$ -Wege mit höchstens  $m + 1$  Richtungswechseln berechnet.  $\square$

Wir haben festgestellt, dass die beiden Varianten des MOORE-BELLMAN-Verfahrens korrekt arbeiten, wenn der gegebene Digraph keine negativen Kreise enthält, aber haben bisher verschwiegen, wie man das effektiv entdeckt. Wie man das bei der D'ESOPOPAPE-Variante auf einfache Weise machen kann – ohne andere Algorithmen einzuschalten – ist mir nicht bekannt. Bei der YEN-Variante gibt es eine simple Modifikation, die das Gewünschte leistet.

**(5.24) Bemerkung.** Nehmen wir an, dass jeder Knoten des Digraphen  $D$  von  $s = 1$  auf einem gerichteten Weg erreicht werden kann.  $D$  enthält einen negativen Kreis genau dann, wenn bei einer zusätzlichen Ausführung der Schleife 2 der YEN-Variante (also für  $m = n - 1$ ) der Wert  $\text{DIST}(v)$  für mindestens einen Knoten  $v \in V$  geändert wird.  $\triangle$

Der Beweis dieser Bemerkung sei dem Leser überlassen. Auf das Thema „negative Kreise“ werden wir später noch einmal zurückkommen.

Die YEN-Variante des MOORE-BELLMAN-Algorithmus hat, da drei Schleifen über maximal  $n$  Indizes ineinander geschaltet sind, eine Laufzeit von  $O(n^3)$ . Für die D'ESOPOPAPE-Variante gibt es (konstruierte) Beispiele mit exponentieller Laufzeit. Dennoch hat sie sich in der Praxis als sehr schnell erwiesen und ist fast immer der YEN-Variante überlegen. Sind alle Gewichte positiv und sollen kürzeste  $(s, v)$ -Wege für alle  $v \in V$  bestimmt werden, so ist die DIJKSTRA-Methode für Digraphen mit vielen Bögen (d. h.  $O(n^2)$  Bögen) die bessere Methode; bei Digraphen mit wenigen Bögen haben extensive Testläufe gezeigt, dass die D'ESOPOPAPE-Variante in der Praxis günstigere Laufzeiten erbringt.

### 5.3.3 Kürzeste Wege zwischen allen Knotenpaaren

Natürlich kann man kürzeste Wege zwischen je zwei Knotenpaaren eines Digraphen  $D$  dadurch bestimmen, dass man das DIJKSTRA- oder das MOORE-BELLMAN-Verfahren  $n$ -mal anwendet, d. h. jeder Knoten wird einmal als Startknoten gewählt. Bei Benutzung der DIJKSTRA-Methode (nicht-negative Gewichte vorausgesetzt) hätte dieses Verfahren eine Laufzeit von  $O(n^3)$ . Falls negative Gewichte vorkommen, müsste die YEN-Variante verwendet werden, was zu einer Laufzeit von  $O(n^4)$  führt. Es gibt jedoch einen extrem einfachen  $O(n^3)$ -Algorithmus, der das Gleiche leistet. Dieses Verfahren geht auf Floyd (1962) zurück.

#### (5.25) FLOYD-Algorithmus.

**Eingabe:** Digraph  $D = (V, A)$ ,  $V = \{1, \dots, n\}$  mit Gewichten  $c(a)$  (können auch negativ sein), für alle  $a \in A$ .

**Ausgabe:** Eine  $(n, n)$ -Matrix  $W = (w_{ij})$ , so dass für  $i \neq j$   $w_{ij}$  die Länge des kürzesten  $(i, j)$ -Weges und  $w_{ii}$  die Länge eines kürzesten gerichteten Kreises, der  $i$  enthält, ist (eine Matrix mit diesen Eigenschaften nennt man *Kürzeste-Weglängen-Matrix*)

und eine  $(n, n)$ -Matrix  $P = (p_{ij})$ , so dass  $p_{ij}$  der vorletzte Knoten eines kürzesten  $(i, j)$ -Weges (bzw.  $(i, i)$ -Kreises) ist.

1. FOR  $i = 1$  TO  $n$  DO:  
     FOR  $j = 1$  TO  $n$  DO:

$$w_{ij} := \begin{cases} c((i, j)) & \text{falls } (i, j) \in A \\ +\infty & \text{andernfalls} \end{cases}$$

$$p_{ij} := \begin{cases} i & \text{falls } (i, j) \in A \\ 0 & \text{andernfalls (bedeutet: zur Zeit kein Weg bekannt)} \end{cases}$$

    END

END

2. FOR  $l = 1$  TO  $n$  DO:  
     FOR  $i = 1$  TO  $n$  DO:  
         FOR  $j = 1$  TO  $n$  DO:  
             Falls  $w_{ij} > w_{il} + w_{lj}$ ,  
             setze:  $w_{ij} := w_{il} + w_{lj}$  und  $p_{ij} := p_{lj}$ .  
             (Falls  $i = j$  und  $w_{ii} < 0$ , kann abgebrochen werden.)

        END

    END

END

3. Gib  $W$  und  $P$  aus. △

Für zwei Knoten  $i, j$  kann der in  $P$  gespeicherte kürzeste  $(i, j)$ -Weg wie folgt bestimmt werden. Setze  $k := 1$  und  $v_k := p_{ij}$ . Ist  $v_k = i$  dann STOP, andernfalls setze  $v_{k+1} := p_{iv_k}$ ,  $k := k + 1$  und wiederhole, d. h. wir iterieren so lange bis ein Knoten, sagen wir  $v_s$ , der Knoten  $i$  ist, dann ist

$$(i = v_s, v_{s-1}, v_{s-2}, \dots, v_1, j)$$

ein kürzester  $(i, j)$ -Weg. Überzeugen Sie sich, dass dies stimmt!

**(5.26) Satz.** Sei  $D = (V, A)$  ein Digraph mit beliebigen Bogengewichten  $c(a)$  für alle  $a \in A$ . Sei  $W$  die  $(n, n)$ -Matrix, die vom FLOYD-Algorithmus produziert wird, dann gilt:

- (a) Der FLOYD-Algorithmus liefert genau dann eine Kürzeste-Weglängen-Matrix  $W$ , wenn  $D$  keinen negativen gerichteten Kreis enthält.
- (b)  $D$  enthält genau dann einen negativen gerichteten Kreis, wenn ein Hauptdiagonalelement von  $W$  negativ ist. △

**Beweis.** Zur Notationsvereinfachung bezeichnen wir die Anfangsmatrix  $W$  aus Schritt 1 mit  $W^0$ , die Matrix  $W$  nach Beendigung des  $l$ -ten Durchlaufs der äußeren Schleife von Schritt 2 mit  $W^l$ . Durch Induktion über  $l = 0, 1, \dots, n$  zeigen wir, dass  $W^l$  genau dann die Matrix der kürzesten Längen von  $(i, j)$ -Wegen (bzw.  $(i, i)$ -Kreisen) ist, bei denen die Knoten  $1, \dots, l$  als innere Knoten auftreten können, wenn  $D$  keinen negativen Kreis in der Knotenmenge  $1, \dots, l$  besitzt. Ist letzteres der Fall, so gilt  $w_{ii}^l < 0$  für ein  $i \in \{1, \dots, l\}$ .

Für  $l = 0$  ist die Behauptung offenbar richtig. Angenommen, sie ist für  $l \geq 0$  richtig, und wir haben die äußere Schleife von Schritt 2 zum  $(l + 1)$ -ten Male durchlaufen. Bei diesem Durchlauf haben wir folgenden Schritt ausgeführt.

$$\text{Falls } w_{ij}^l > w_{i,l+1}^l + w_{l+1,j}^l, \text{ dann setze } w_{ij}^{l+1} := w_{i,l+1}^l + w_{l+1,j}^l,$$

d. h. wir haben die (nach Induktionsvoraussetzung) kürzeste Länge eines  $(i, j)$ -Weges über die Knoten  $1, \dots, l$  verglichen mit der Summe der kürzesten Längen eines  $(i, l + 1)$ -Weges und eines  $(l + 1, j)$ -Weges jeweils über die Knoten  $1, \dots, l$ . Die letztere Summe repräsentiert also die Länge eines kürzesten  $(i, j)$ -Weges über  $1, \dots, l + 1$ , der den Knoten  $l + 1$  enthält. Falls diese Summe kleiner als  $w_{ij}^l$  ist, setzen wir  $w_{ij}^{l+1} := w_{i,l+1}^l + w_{l+1,j}^l$ , andernfalls  $w_{ij}^{l+1} = w_{ij}^l$ . Daraus folgt die Behauptung, es sei denn,  $w_{ij}^l > w_{i,l+1}^l + w_{l+1,j}^l$  und die Verkettung, sagen wir  $K$  des  $(i, l + 1)$ -Weges mit dem  $(l + 1, j)$ -Weg ist gar kein Weg, d. h.  $K$  ist eine gerichtete  $(i, j)$ -Kette, die einen Knoten mindestens zweimal enthält. Die Kette  $K$  enthält natürlich einen  $(i, j)$ -Weg, sagen wir  $\bar{K}$ , und  $\bar{K}$  geht aus  $K$  dadurch hervor, dass wir die in  $K$  vorkommenden gerichteten Kreise entfernen. Der Knoten  $l + 1$  ist nach Konstruktion in einem der Kreise enthalten, also ist  $\bar{K}$  ein  $(i, j)$ -Weg, der nur Knoten aus  $\{1, \dots, l\}$  enthält, d. h.  $w_{ij}^l \leq c(\bar{K})$ . Aus  $c(K) = w_{i,l+1}^l + w_{l+1,j}^l < w_{ij}^l$  folgt, dass mindestens einer der aus  $K$  entfernten gerichteten Kreise eine negative Länge hat. Für jeden Knoten  $i$  dieses negativen Kreises muss folglich  $w_{ii}^{l+1} < 0$  gelten. Daraus folgt die Behauptung.  $\square$

Der FLOYD-Algorithmus liefert also explizit einen Kreis negativer Länge, falls ein solcher existiert.

**(5.27) Korollar.** *Für einen Digraphen  $D$  mit Bogengewichten, der keine negativen gerichteten Kreise enthält, kann ein kürzester gerichteter Kreis in  $O(n^3)$  Schritten bestimmt werden.*  $\triangle$

**Beweis.** Wir führen den FLOYD-Algorithmus aus. Nach Beendigung des Verfahrens ist in  $w_{ii}$ ,  $i = 1, \dots, n$  die Länge eines kürzesten gerichteten Kreises, der den Knoten  $i$  enthält, verzeichnet. Wir wählen einen Wert  $w_{ii}$ , der so klein wie möglich ist, und rekonstruieren aus der Matrix  $P$ , wie oben angegeben, den gerichteten Kreis, der  $i$  enthält. Dieser ist ein kürzester gerichteter Kreis in  $D$ . Diese „Nachbearbeitung“ erfordert lediglich  $O(n)$  Operationen, also ist die worst-case-Komplexität des FLOYD-Algorithmus auch die Laufzeitschranke für das Gesamtverfahren.  $\square$

Wendet man Algorithmus (5.25) auf Entfernungstabellen in Straßenatlanten an, so wird man feststellen, dass es häufig Städte  $i, j, k$  gibt mit  $c_{ij} + c_{jk} < c_{ik}$ . Die Entfernungen genügen also nicht der Dreiecksungleichung. Warum ist das so?

### 5.3.4 Min-Max-Sätze und weitere Bemerkungen

Es folgen in einem kurzen Überblick ein paar Zusatzbemerkungen zum Problemkreis „Kürzeste Wege“.

**Zwei Min-Max-Sätze.** In der Optimierungstheorie sind sogenannte Dualitäts- oder Min-Max-Sätze von besonderer Bedeutung. Diese Sätze sind von folgendem Typ: Man hat eine Menge  $P$  und eine Zielfunktion  $c$ , die jedem Element  $x$  von  $P$  einen Wert  $c(x)$  zuordnet. Gesucht wird

$$\min\{c(x) \mid x \in P\}.$$

Dann gelingt es manchmal auf natürliche Weise und unter gewissen technischen Voraussetzungen eine Menge  $D$  und eine Zielfunktion  $b$  zu finden, die jedem  $y \in D$  einen Wert  $b(y)$  zuweist, mit der Eigenschaft

$$\min\{c(x) \mid x \in P\} = \max\{b(y) \mid y \in D\}.$$

Wie wir später sehen werden, ist die Existenz eines Satzes dieser Art häufig ein Indikator dafür, dass das Minimierungs- und das Maximierungsproblem „gut“ gelöst werden können. Für das Kürzeste-Wege-Problem gibt es verschiedene Min-Max-Sätze. Wir geben zwei Beispiele an und erinnern daran, dass ein  $(s, t)$ -Schnitt in einem Digraphen  $D = (V, A)$  eine Bogenmenge der Form  $\delta^+(W) = \{(i, j) \in A \mid i \in W, j \in V \setminus W\}$  ist mit der Eigenschaft  $s \in W, t \in W \setminus V$ .

**(5.28) Satz.** Sei  $D = (V, A)$  ein Digraph, und seien  $s, t \in V, s \neq t$ . Dann ist die minimale Länge (= Anzahl der Bögen) eines  $(s, t)$ -Weges gleich der maximalen Anzahl bogendisjunkter  $(s, t)$ -Schnitte.  $\triangle$

**Beweis.** Jeder  $(s, t)$ -Weg enthält aus jedem  $(s, t)$ -Schnitt mindestens einen Bogen. Gibt es also  $d$  bogendisjunkte  $(s, t)$ -Schnitte, so hat jeder  $(s, t)$ -Weg mindestens die Länge  $d$ . Daher ist das Minimum (d. h. die kürzeste Länge eines  $(s, t)$ -Weges) mindestens so groß wie das Maximum (gebildet über die Anzahl bogendisjunkter  $(s, t)$ -Schnitte).

Sei nun  $d$  die Länge eines kürzesten Weges, und sei  $V_i, i = 1, \dots, d$ , die Menge der Knoten  $v \in V$ , die von  $s$  aus auf einem Weg der Länge kleiner als  $i$  erreicht werden können. Dann sind die Schnitte  $\delta^+(V_i)$  genau  $d$  bogendisjunkte  $(s, t)$ -Schnitte.  $\square$

Eine Verallgemeinerung dieses Sachverhaltes auf gewichtete Digraphen ist das folgende Resultat.

**(5.29) Satz.** Seien  $D = (V, A)$  ein Digraph,  $s, t \in V, s \neq t$ , und  $c(a) \in \mathbb{Z}_+$  für alle  $a \in A$ . Dann ist die kürzeste Länge eines  $(s, t)$ -Weges gleich der maximalen Anzahl  $d$  von (nicht notwendig verschiedenen)  $(s, t)$ -Schnitten  $C_1, \dots, C_d$ , so dass jeder Bogen  $a \in A$  in höchstens  $c(a)$  Schnitten  $C_i$  liegt.  $\triangle$

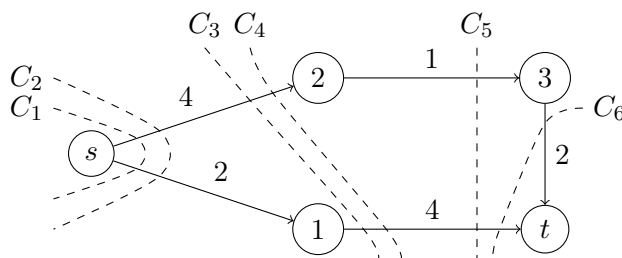


Abbildung 5.4: Beispiel für Satz (5.29).

**Beweis.** Sei  $P$  ein  $(s, t)$ -Weg und seien  $C_1, \dots, C_d$   $(s, t)$ -Schnitte wie im Satz gefordert, dann gilt

$$c(P) = \sum_{a \in P} c(a) \geq \sum_{a \in P} |\{i \mid a \in C_i\}| = \sum_{i=1}^d |C_i \cap P| \geq \sum_{i=1}^d 1 = d.$$

Also ist das Minimum nicht kleiner als das Maximum.

Wählen wir die  $(s, t)$ -Schnitte  $C_i := \delta^+(V_i)$ , mit  $V_i := \{v \in V \mid v \text{ kann von } s \text{ aus auf einem gerichteten Weg } P \text{ mit } c(P) \leq i - 1 \text{ erreicht werden}\}$ ,  $i = 1, \dots, d$ , dann sehen wir, dass Gleichheit gilt.  $\square$

Ein ganz einfaches Beispiel soll zur Veranschaulichung von Satz (5.29) dienen. Wir betrachten Abbildung 5.4. Der kürzeste Weg von  $s$  nach  $t$  hat offensichtlich Länge 6. Konstruieren wir die Knotenmengen  $V_i$  und die realisierenden  $(s, t)$ -Schnitte  $\delta^+(V_i)$ , so ergeben sich folgende Knotenmengen:  $V_1 = V_2 = \{s\}$ ,  $V_3 = V_4 = \{s, 1\}$ ,  $V_5 = \{s, 1, 2\}$ ,  $V_6 = \{s, 1, 2, 3\}$ . Die zugehörigen Schnitte  $C_i := \delta^+(V_i)$  sind in Abbildung 5.4 als gestrichelte Linien angedeutet.

Kommen wir zur Einführung von Min-Max-Sätzen zurück. In abstrakter kombinatorischer Notation kann man Satz (5.29) wie folgt notieren. Mit  $\mathcal{P}(D, c)$  bezeichnen wir die Menge aller  $(s, t)$ -Wege  $P$  in  $D$  und  $c(P) := \sum_{a \in P} c(a)$  sei die Länge eines Weges  $P \in \mathcal{P}(D, c)$ . Mit  $\mathcal{C}$  bezeichnen wir die Menge aller endlichen Folgen  $(C_1, C_2, \dots, C_k)$ ,  $k \geq 1$ , von  $(s, t)$ -Schnitten, d. h.  $C_i = \delta^+(W_i)$  mit  $s \in W_i$ ,  $t \notin W_i$ , wobei  $C_i$  und  $C_j$ ,  $i \neq j$ , nicht voneinander verschieden sein müssen. Sei

$$\mathcal{D}(D, c) := \{(C_1, C_2, \dots, C_k) \in \mathcal{C} \mid |\{i \in \{1, \dots, k\} : a \in C_i\}| \leq c(a) \forall a \in A\}$$

und der Wert  $b((C_1, C_2, \dots, C_k))$  einer Folge aus  $\mathcal{D}(D, c)$  sei die Anzahl der nicht notwendig verschiedenen Elemente der Folge, also  $b((C_1, C_2, \dots, C_k)) = k$ . Dann gilt:

$$\min\{c(P) \mid P \in \mathcal{P}(D, c)\} = \max\{b((C_1, C_2, \dots, C_k)) \mid (C_1, C_2, \dots, C_k) \in \mathcal{D}(D, c)\}.$$

**Kürzeste Wege in ungerichteten Graphen.** Transformieren wir einen ungerichteten Graphen  $G$  in einen gerichteten Graphen  $D$ , indem wir jeder Kante  $ij$  die beiden Bögen  $(i, j)$  und  $(j, i)$  mit dem Gewicht von  $ij$  zuordnen, so können wir natürlich durch Anwendung unserer Verfahren auf  $D$  auch kürzeste Wege bzw. Kreise in  $G$  bestimmen. Man

beachte jedoch, dass ein negatives Kantengewicht  $c(ij)$  in  $G$  automatisch zu einem negativen gerichteten Kreis  $(i, j)(j, i)$  in  $D$  führt. Mit unseren Methoden können wir also nur kürzeste Wege und Kreise in Graphen mit nichtnegativen Kantengewichten bestimmen.

Es sei an dieser Stelle jedoch darauf hingewiesen, dass auch in Graphen mit negativen Kantengewichten kürzeste Wege und Kreise bestimmt werden können, falls kein Kreis negativ ist. Dies geschieht mit Methoden der Matching-Theorie, auf die wir hier aus Zeitgründen nicht eingehen können.

**Laufzeiten.** Genaue Laufzeitanalysen von verschiedenen Varianten der hier vorgestellten Algorithmen zur Berechnung von kürzesten Wegen findet man z. B. in Ahuja et al. (1993) auf den Seiten 154–157 ebenso, wie einen kurzen geschichtlichen Überblick.

Umfangreiche historische Bemerkungen zur Theorie und Algorithmik von kürzesten Wegen bietet das Buch von Schrijver (2003). In den Abschnitten 7.5 und 8.6 sind z. B. Tabellen zu finden, die die Verbesserungen der Worst-Case-Laufzeiten von Kürzeste-Wege-Algorithmen dokumentieren.

Ein Algorithmus zur Bestimmung kürzester Wege muss jeden Bogen des gegebenen Digraphen  $D = (V, A)$  mindestens einmal „anfassen“. Eine untere Schranke für die Laufzeit eines jeden Algorithmus dieser Art ist somit  $O(m)$ ,  $m = |A|$ . Thorup (1997) hat gezeigt, dass man diese Laufzeit für ungerichtete Graphen mit nichtnegativen Kantengewichten tatsächlich erreichen kann. Er benutzt dazu sogenannte „Atomic Heaps“, deren Verwendung  $n = |V| \geq 2^{12^{20}}$  voraussetzt. Das bedeutet, dass diese Methode zwar theoretisch „gut“, aber für die Praxis ungeeignet ist. (Thorup diskutiert in seinem Aufsatz auch implementierbare Varianten, allerdings haben diese eine schlechtere Laufzeit, z. B.  $O(\log C_{\max} + m + n \log \log \log n)$ , wobei  $C_{\max}$  das größte Kantengewicht bezeichnet.)

Bei Routenplanern, wie sie z. B. im Internet oder in den Bordcomputern von Autos angeboten werden, treten Digraphen mit mehreren Millionen Knoten auf (die in der Regel nur einen kleinen Grad haben). Die Anbieter solcher Programme haben für derartige Probleme, bei denen ja der Grundgraph, der auf der CD gespeichert ist, fest bleibt, Spezialverfahren entwickelt (z. B. durch intensives Preprocessing und die Abspeicherung wichtiger kürzester Verbindungen), die Kürzeste-Wege-Probleme dieser Art sehr schnell lösen. Um (selbst gesetzte) Zeitschranken für den Nutzer einzuhalten, benutzen diese Algorithmen z. T. Heuristiken, und derartige Algorithmen liefern nicht notwendig immer einen beweisbaren kürzesten Weg. Einen Überblick über diesen Aspekt findet man in Goldberg (2007). Ein exakter Algorithmus zur Bestimmung kürzester Wege in sehr großen Straßennetzwerken ist in Geisberger et al. (2012) beschrieben.

Fast alle Navigationssysteme bieten mehrere Optimierungsmöglichkeiten zur Bestimmung eines „besten“ Weges an. Man kann z. B. den schnellsten oder den kürzesten Weg bestimmen lassen. Manche Navigationssysteme offerieren eine „Kombinationsoptimierung“, man kann dann etwa eingeben, dass Schnelligkeit mit 70% und Streckenkürze mit 30% berücksichtigt werden. Dies ist eine spezielle Version der Mehrzieloptimierung. Die zwei Zielfunktionen werden hierbei mit Parametern multipliziert und dann aufaddiert, so dass nur eine einzige Zielfunktion entsteht. Das nennt man *Skalierung (Scaling)*. Man könnte auch anders vorgehen: z. B. könnte man nach der schnellsten Route suchen, die eine

vorgegebene km-Zahl nicht überschreitet. Der Grund dafür, dass Navigationssysteme solche Optionen nicht anbieten, liegt darin, dass Kürzeste-Wege-Probleme mit Nebenbedingungen in der Regel  $\mathcal{NP}$ -schwer sind. Das ist z. B. so bei „schnellster Weg mit km-Beschränkung“ oder „kürzester Weg mit Zeitbeschränkung“.

Noch besser für den Autofahrer wäre die Angabe der *Pareto-Menge*. Im Falle der beiden Zielfunktionen „kürzester Weg“ und „schnellster Weg“ müsste das Navigationssystem alle Wege angeben, die „nicht dominiert“ sind. Solche Wege haben die Eigenschaft, dass beim Versuch, die Weglänge kürzer zu machen, die Fahrzeit erhöht wird oder umgekehrt. Das hierbei unüberwindliche Problem ist, dass die Kardinalität der Pareto-Menge exponentiell in der Kodierungslänge der Daten wachsen kann. Die Navigationssysteme würden „unendlich lange“ rechnen und der Autofahrer in der Informationsflut ertrinken. Aus diesem theoretischen Grund wird nur mit einer Skalierung gearbeitet, die durch den Nutzer (falls er das will) vorgegeben wird.

Man könnte glauben, dass Fahrzeugnavigation das größte Anwendungsfeld von Methoden zur Berechnung kürzester Wege sei, aber der Umfang der Verwendung dieser Methoden ist im Internet noch viel größer. Das derzeit am häufigsten verwendete Routing-Protokoll ist das „Open Shortest Path First-Protokoll“ (kurz: OSPF). Bei Verwendung dieses Protokolls wird für jedes Datenpaket ein kürzester Weg (u. U. mehrfach) bestimmt, und wenn man allein die Anzahl der E-Mails abschätzt, die weltweit täglich versandt werden, so sieht man sehr schnell, wie häufig die Bestimmung kürzester Wege zum Einsatz kommt. Ich kann hier das OSPF-Protokoll nicht im Detail erklären und verweise dazu auf Internetquellen, z. B. Wikipedia.

Die am Ende von Abschnitt 5.2 genannten Webseiten bieten auch verschiedene Algorithmen zur Berechnung kürzester Wege an.

## 5.4 LP/IP-Aspekte von Bäumen und Wegen

Wir haben nun gelernt, wie man minimale Bäume, maximale Wälder und kürzeste Wege bestimmen kann. Dies ist mit theoretisch und praktisch hocheffizienten Algorithmen möglich. Häufig ist das Auffinden von Bäumen und Wegen jedoch nur ein Teilproblem eines komplizierten mathematischen Modells, welches mit Methoden der gemischt-ganzzahligen Optimierung behandelt werden muss. Und ebenso häufig treten Zusatzanforderungen an die Bäume und Wege auf, die aus diesen einfachen Problemen  $\mathcal{NP}$ -schwere Aufgaben machen. In solchen Fällen (und nicht nur dabei) ist es erforderlich, Bäume und Wege-Probleme als ganzzahlige (oder wenn möglich lineare) Programme zu formulieren. Wir deuten nachfolgend an, wie das gemacht werden kann.

Wir beginnen mit Wäldern in einem gegebenen Graphen  $G = (V, E)$  mit Kantengewichten  $c_e$ ,  $e \in E$ . Wir wollen ein lineares Programm aufstellen, dessen ganzzahlige Lösungen den Wäldern in  $G$  entsprechen. Dazu müssen wir lernen, wie man eine Kantemenge  $F \subseteq E$  als Vektor darstellt. Wir betrachten hierzu den Vektorraum  $\mathbb{R}^E$ . Man kann diesen als die Menge der Abbildungen von  $E$  in  $\mathbb{R}$  auffassen. Die (für uns) angenehmste Sichtweise ist die folgende: Ein Vektor  $x \in \mathbb{R}^E$  hat für jedes Element  $e \in E$  eine Komponente  $x_e$ . Wir schreiben den Vektor dann wie folgt:  $x = (x_e)_{e \in E}$ . Für eine



beliebige Teilmenge  $F \subseteq E$  definieren wir ihren *Inzidenzvektor*  $\chi^F = (\chi_e^F)_{e \in E}$  dadurch, dass wir setzen:  $\chi_e^F = 1$ , falls  $e \in F$ , und  $\chi_e^F = 0$ , falls  $e \notin F$ . Definieren wir

$$\mathfrak{W} := \{F \subseteq E \mid (V, F) \text{ ist ein Wald in } G\},$$

so können wir die Aufgabe, einen gewichtsmaximalen Wald in  $\mathfrak{W}$  zu finden, nun auffassen als die Aufgabe, in der Menge aller Inzidenzvektoren

$$\mathfrak{F} := \{\chi^F \in \mathbb{R}^E \mid F \in \mathfrak{W}\}$$

von Wäldern einen Vektor mit höchstem Gewicht zu bestimmen. Die Kantengewichte  $c_e$  können wir analog zu einem Vektor  $c = (c_e)_{e \in E} \in \mathbb{R}^E$  zusammenfassen, und damit haben wir die Aufgabe, einen gewichtsmaximalen Wald zu finden, als Optimierungsproblem mit linearer Zielfunktion

$$\max \left\{ c^T x = \sum_{e \in E} c_e x_e \mid x \in \mathfrak{F} \right\}$$

über einer Menge von 0/1-Vektoren formuliert.

Um diese Aufgabe als lineares ganzzahliges Programm auffassen zu können, müssen wir nun noch lineare Ungleichungen finden, so dass die ganzzahligen Lösungen des Ungleichungssystems genau die Vektoren in  $\mathfrak{F}$  sind. Wir beginnen mit einer Trivialüberlegung. Die Komponenten der Vektoren in  $\mathfrak{F}$  nehmen nur die Werte 0 oder 1 an. Damit erfüllt jedes Element von  $\mathfrak{F}$  die Ungleichungen

$$0 \leq x_e \leq 1 \quad \forall e \in E.$$

Nun gilt es, ein System von Ungleichungen zu „erraten“, das von allen Elementen aus  $\mathfrak{F}$  erfüllt wird und das die Eigenschaft hat, dass jeder 0/1-Vektor, der nicht in  $\mathfrak{F}$  ist, mindestens eine der Ungleichungen des Systems verletzt. Die Idee dazu liefert die Definition. Ein Wald  $(V, F)$ ,  $F \subseteq E$  ist dadurch gekennzeichnet, dass  $(V, F)$  keinen Kreis enthält. Fordern wir also, dass für jeden Kreis  $C \subseteq E$  die Ungleichung

$$x(C) := \sum_{e \in C} x_e \leq |C| - 1$$

(genannt *Kreisungleichung*) erfüllt sein muss, so erfüllt nach Definition jeder Inzidenzvektor eines Waldes jede dieser Ungleichungen. Ist  $\chi^H$  der Inzidenzvektor einer Kantenmenge  $H \subseteq E$ , die einen Kreis  $C$  enthält, so gilt:

$$\sum_{e \in C} \chi_e^H = |C|$$

und damit verletzt  $\chi^H$  die zu  $C$  gehörige Kreisungleichung. Wir fassen zusammen: Das System von Ungleichungen

$$\begin{aligned} 0 \leq x_e \leq 1 & \quad \forall e \in E, \\ x(C) \leq |C| - 1 & \quad \forall C \subseteq E, C \text{ Kreis in } G \end{aligned}$$

## 5 Bäume und Wege

ist ein System von Ungleichungen, dessen 0/1-Lösungen genau der Menge der Inzidenzvektoren der Wälder in  $G$  entspricht. Daraus folgt, dass

$$\begin{aligned} \max \quad & c^T x \\ & 0 \leq x_e \leq 1 \quad \forall e \in E \\ & x(C) \leq |C| - 1 \quad \forall C \subseteq E, C \text{ Kreis in } G \\ & x_e \in \mathbb{Z} \quad \forall e \in E \end{aligned} \tag{5.30}$$

ein ganzzahliges lineares Programm ist, dessen ganzzahlige Lösungen die Inzidenzvektoren von Wäldern sind.

Es wird sich zeigen (und an dieser Stelle beweisen wir das noch nicht), dass das obige ganzzahlige Programm keine „gute Formulierung“ des Wald-Problems ist. Es gibt (mindestens) ein besseres Modell. Man kann beweisen, dass das folgende lineare Programm

$$\begin{aligned} \max \quad & c^T x \\ & 0 \leq x_e \leq 1 \quad \forall e \in E \\ & x(E(W)) \leq |W| - 1 \quad \forall W \subseteq V, (W, E(W)) \text{ ist 2-fach zusammenhängend} \end{aligned} \tag{5.31}$$

die Eigenschaft hat, dass die „Ecklösungen“ dieses LPs genau die Inzidenzvektoren der Wälder in  $G$  sind. Man kann also das Wald-Problem durch ein lineares Programm lösen (ohne Ganzzahligkeitsforderung). Man zahlt allerdings einen hohen Preis dafür. Die Anzahl der Nebenbedingungen ist exponentiell in  $|V|$ . Man kann sie z. B. für einen vollständigen Graphen mit 100 000 Knoten überhaupt nicht aufschreiben (es gibt rund  $2^{100000}$  Ungleichungen). Dennoch kann man derartige LPs effizient lösen, wie wir später sehen werden.

Wir wissen nun, wie man Wälder mit linearer Optimierung behandelt. Kann man auf ähnliche Weise auch minimale aufspannende Bäume bestimmen? Überlegen Sie sich das bitte!

Und nun zu  $(s, t)$ -Wegen in einem Digraphen  $D = (V, A)$  mit  $s, t \in V, s \neq t$ . Wir gehen hier analog vor. Jedem Bogen  $(u, v) \in A$  ordnen wir eine Variable  $x_{uv}$  zu. Diese Variablen fassen wir zu einem Vektor  $x = (x_{uv})_{(u,v) \in A} \in \mathbb{R}^A$  zusammen. Für jeden  $(s, t)$ -Weg  $P \subseteq A$  definieren wir den zugehörigen Inzidenzvektor  $\chi^P = (\chi_{uv}^P)_{(u,v) \in A} \in \mathbb{R}^A$  durch  $\chi_{uv}^P = 1$ , falls  $(u, v) \in P$ , und  $\chi_{uv}^P = 0$ , falls  $(u, v) \notin P$ . Kann man nun „kanonische“ Gleichungen und Ungleichungen finden, so dass das Kürzeste-Wege-Problem als ganzzahliges Programm oder gar als lineares Programm ohne Ganzzahligkeitsbedingung gelöst werden kann? Es folgen ein paar einfache Überlegungen hierzu.

Da alle Inzidenzvektoren 0/1-Vektoren sind, müssen die Ungleichungen

$$0 \leq x_{uv} \leq 1 \quad \forall (u, v) \in A \tag{5.32}$$

erfüllt sein. Aus dem Anfangsknoten  $s \in V$  führt ein Bogen hinaus, in den Endknoten  $t \in V$  des  $(s, t)$ -Weges führt ein Bogen hinein. Daraus folgt, dass die beiden folgenden

Gleichungen erfüllt sein müssen:

$$\begin{aligned} x(\delta^+(s)) &= \sum_{(s,v) \in A} x_{sv} = 1 \\ x(\delta^-(t)) &= \sum_{(u,t) \in A} x_{ut} = 1. \end{aligned} \tag{5.33}$$

Jeder andere Knoten  $v \in V \setminus \{s, t\}$  ist nur „Durchgangsknoten“, d. h. wenn ein Bogen des  $(s, t)$ -Weges in  $v$  hineinführt, dann geht auch ein Bogen dieses Weges hinaus. Geht kein Bogen in  $v$  hinein, dann geht auch keiner hinaus. Also müssen die folgenden Ungleichungen erfüllt sein:

$$x(\delta^-(v)) - x(\delta^+(v)) = \sum_{(u,v) \in A} x_{uv} - \sum_{(v,w) \in A} x_{vw} = 0 \quad \forall v \in V \setminus \{s, t\}. \tag{5.34}$$

Reicht das System von Gleichungen und Ungleichungen (5.32), (5.33), (5.34) zur Lösung des Kürzeste-Wege-Problems aus? Benötigt man zusätzlich Ganzzahligkeitsbedingungen? Braucht man noch mehr Ungleichungen oder Gleichungen? Denken Sie darüber nach!

## 5.5 Exkurs: Greedy-Heuristiken für das Rucksackproblem und deren Analyse

Betrachtet man das Kürzeste-Wege-Problem und die Berechnung minimaler aufspannender Bäume, dann könnte man meinen, dass der Greedy-Algorithmus immer eine gute Methode zur Lösung kombinatorischer Optimierungsprobleme ist. Das ist natürlich nicht so. Wir wollen deshalb Varianten dieses Algorithmus-Typs für das *Rucksackproblem* (engl.: *knapsack problem*) untersuchen. Das Rucksackproblem ist in einem gewissen Sinne eines der einfachsten (aber dennoch  $\mathcal{NP}$ -schweren) ganzzahligen Optimierungsprobleme.

Das Rucksackproblem kann folgendermaßen beschrieben werden. Gegeben seien  $n$  verschiedene Arten von Gegenständen, die in einen Rucksack gefüllt werden sollen. Der Rucksack hat ein beschränktes Fassungsvermögen  $b$ . Jeder Gegenstand einer Art  $j$  hat ein „Gewicht“  $a_j$  und einen „Wert“  $c_j$ . Der Rucksack soll so gefüllt werden, dass seine Kapazität nicht überschritten wird und der Gesamtwert so wertvoll wie möglich ist. Von diesem Problem gibt es viele Varianten. Hier betrachten wir zwei von diesen.

**(5.35) Definition.** Seien  $a_j, c_j \in \mathbb{Z}_+$ ,  $j = 1, 2, \dots, n$  und  $b \in \mathbb{N}$ .

(a) *Das Problem*

$$\begin{aligned} \max \quad & \sum_{j=1}^n c_j x_j \\ & \sum_{j=1}^n a_j x_j \leq b \\ & x_j \in \mathbb{Z}_+ \end{aligned} \tag{KP}$$

heißt (allgemeines) Rucksack-Problem *oder* Rucksackproblem.

- (b) Fordern wir zusätzlich, dass  $x_j \in \{0, 1\}$ ,  $j = 1, \dots, n$  gilt, so heißt (KP) binäres Rucksack-Problem oder 0/1-Rucksack-Problem.  $\triangle$

Zu dem (so simpel erscheinenden) Rucksackproblem gibt es eine außerordentlich umfangreiche Literatur. Wir erwähnen hier nur zwei Bücher. Der „Klassiker“ zum Thema ist Martello and Toth (1990), seine „Fortsetzung“ Kellerer et al. (2004). Dieses Buch mit 546 Seiten enthält so gut wie alles, was man über das Rucksack-Problem wissen will.

Wir untersuchen zwei Versionen des Greedy-Algorithmus für das Rucksack-Problem.

**(5.36) Zwei Greedy-Algorithmen für das allgemeine bzw. binäre Rucksack-Problem.**

**Eingabe:**  $c_j, a_j \in \mathbb{N}$ ,  $j = 1, \dots, n$  und  $b \in \mathbb{N}$ .

**Ausgabe:** Eine zulässige (approximative) Lösung für (KP) bzw. das binäre Rucksackproblem.

1. *Zielfunktionsgreedy:* Ordne die Indizes, so dass  $c_1 \geq c_2 \geq \dots \geq c_n$  gilt.
- 1'. *Gewichtsdichtengreedy:* Berechne die *Gewichtsdichten*  $\rho_j := c_j/a_j$ ,  $j = 1, 2, \dots, n$ , und ordne die Indizes so dass  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n$  gilt.
2. FOR  $j = 1$  TO  $n$  DO:
  - allgemeiner Rucksack:*
    - Setze  $x_j := \lfloor b/a_j \rfloor$  und  $b := b - a_j \lfloor b/a_j \rfloor$ .
  - 0/1-Rucksack:*
    - Falls  $a_j > b$ , setze  $x_j := 0$ .
    - Falls  $a_j \leq b$ , setze  $x_j := 1$  und  $b := b - a_j$ .
3. END.  $\triangle$

Die Laufzeit der in (5.36) beschriebenen Algorithmen ist offensichtlich  $O(n \log n)$ , wobei das Sortieren die Hauptarbeit erfordert.

Wir nennen einen Algorithmus  $A$   $\epsilon$ -*approximativ*,  $0 < \epsilon \leq 1$ , falls er für *jedes* Problembeispiel  $\mathcal{I}$  eines Maximierungsproblems einen Wert  $c_A(\mathcal{I})$  liefert, der mindestens einen  $\epsilon$ -Anteil des Optimalwerts  $c_{\text{opt}}(\mathcal{I})$  von  $\mathcal{I}$  beträgt, in Formeln, falls

$$c_A(\mathcal{I}) \geq \epsilon c_{\text{opt}}(\mathcal{I}). \tag{5.37}$$

**(5.38) Bemerkung.** Der Zielfunktionsgreedy kann für das allgemeine und auch für das binäre Rucksack-Problem beliebig schlechte Lösungen liefern, d. h. es gibt kein  $\epsilon$ ,  $0 < \epsilon \leq 1$ , so dass dieser Greedy-Algorithmus  $\epsilon$ -approximativ ist.  $\triangle$

**Beweis.** (a) Allgemeines Rucksack-Problem. Wir betrachten das folgende Beispiel mit  $\alpha \in \mathbb{Z}$ ,  $\alpha \geq 2$ :

$$\begin{aligned} \max \quad & \alpha x_1 + (\alpha - 1)x_2 \\ & \alpha x_1 + \quad \quad x_2 \leq \alpha \\ & \quad \quad \quad x_1, x_2 \in \mathbb{Z}_+ \end{aligned}$$

5.5 Exkurs: Greedy-Heuristiken für das Rucksackproblem und deren Analyse

Offenbar gilt  $c_{\text{greedy}} = \alpha$  und der Wert der Optimallösung ist  $c_{\text{opt}} = \alpha(\alpha - 1)$ . Aus der Forderung  $c_{\text{greedy}} \geq \epsilon c_{\text{opt}}$  folgt dann  $\epsilon \leq 1/(\alpha - 1)$ ; da wir  $\alpha$  beliebig wählen können, kann es kein festes  $\epsilon > 0$  mit  $c_{\text{greedy}} \geq \epsilon c_{\text{opt}}$  geben.

(b) 0/1-Rucksack-Problem. Wir betrachten für  $n \in \mathbb{Z}$ ,  $n \geq 2$ :

$$\begin{aligned} \max \quad & nx_1 + (n-1)x_2 + \dots + (n-1)x_n \\ & nx_1 + \quad \quad x_2 + \dots + \quad \quad x_n \leq n \\ & \quad \quad \quad \quad \quad \quad \quad \quad x_j \in \{0, 1\}, \quad j = 1, \dots, n. \end{aligned}$$

Trivialerweise gilt  $c_{\text{greedy}} = n$  und  $c_{\text{opt}} = (n-1)(n-1)$ , und wir können mit demselben Argument wie oben sehen, dass der Zielfunktionsgreedy nicht  $\epsilon$ -approximativ ist.  $\square$

**(5.39) Satz.** *Der Gewichts-dichten-Greedyalgorithmus ist für das allgemeine Rucksackproblem ein 1/2-approximativer Algorithmus.*  $\triangle$

**Beweis.** O. B. d. A. können wir annehmen, dass  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n$  und  $a_1 \leq b$  gilt. Es gilt offensichtlich für den Zielfunktionswert  $c_{\text{Ggreedy}}$  des Gewichts-dichten-Greedyalgorithmus

$$c_{\text{Ggreedy}} \geq c_1 x_1 = c_1 \lfloor b/a_1 \rfloor,$$

und ebenso

$$c_{\text{opt}} \leq c_1 b/a_1 \leq c_1 (\lfloor b/a_1 \rfloor + 1) \leq 2c_1 \lfloor b/a_1 \rfloor \leq 2c_{\text{Ggreedy}}$$

und somit

$$c_{\text{Ggreedy}} \geq \frac{1}{2} c_{\text{opt}}.$$

Wir zeigen nun, dass diese Schranke tatsächlich asymptotisch angenommen wird. Dazu betrachten wir das folgende Rucksackproblem:

$$\begin{aligned} \max \quad & 2\alpha x_1 + 2(\alpha - 1)x_2 \\ & \alpha x_1 + (\alpha - 1)x_2 \leq 2(\alpha - 1) \\ & x_1, x_2 \in \mathbb{Z}_+. \end{aligned}$$

Offensichtlich gilt  $\rho_1 \geq \rho_2$ ,  $c_{\text{Ggreedy}} = 2\alpha$ ,  $c_{\text{opt}} = 4(\alpha - 1)$  und somit

$$\frac{c_{\text{Ggreedy}}}{c_{\text{opt}}} = \frac{2\alpha}{4(\alpha - 1)} \longrightarrow \frac{1}{2}. \quad \square$$

Um lästige Trivialbemerkungen zu vermeiden, nehmen wir im Weiteren an, dass die Indizes so geordnet sind, dass für die Gewichts-dichten  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n$  und dass  $a_j \leq b$  gilt für  $j = 1, \dots, n$ . Leider gilt die schöne Schranke aus (5.39) nicht für das 0/1-Rucksack-Problem.

**(5.40) Bemerkung.** Gegeben sei ein 0/1-Rucksack-Problem. Dann gilt:

## 5 Bäume und Wege

(a) Für  $k = 0, 1, \dots, n-1$  gilt

$$c_{\text{opt}} \leq \sum_{j=1}^k c_j + \frac{c_{k+1}}{a_{k+1}} \cdot \left( b - \sum_{j=1}^k a_j \right).$$

(b)  $c_{\text{Greedy}} > c_{\text{opt}} - \max\{c_j \mid j = 1, \dots, n\}$ . △

**Beweis.** (a) Sei  $x_1^*, \dots, x_n^*$  eine optimale Lösung des 0/1-Rucksack-Problems und  $k \in \{0, 1, \dots, n-1\}$ . Dann gilt:

$$\begin{aligned} c_{\text{opt}} &= \sum_{j=1}^n c_j x_j^* \leq \sum_{j=1}^k c_j x_j^* + \sum_{j=k+1}^n \frac{a_j c_{k+1}}{a_{k+1}} x_j^* \\ &= \frac{c_{k+1}}{a_{k+1}} \sum_{j=1}^n a_j x_j^* + \sum_{j=1}^k \left( c_j - \frac{c_{k+1}}{a_{k+1}} a_j \right) x_j^* \\ &\leq \frac{c_{k+1}}{a_{k+1}} b + \sum_{j=1}^k \left( c_j - \frac{c_{k+1}}{a_{k+1}} a_j \right) \\ &= \sum_{j=1}^k c_j + \frac{c_{k+1}}{a_{k+1}} \left( b - \sum_{j=1}^k a_j \right). \end{aligned}$$

(b) Ist  $\sum_{j=1}^n a_j \leq b$ , so liefert der Greedy-Algorithmus offenbar die optimale Lösung und die Behauptung ist korrekt. Sei also  $k < n$  der größte Index, so dass  $\sum_{j=1}^k a_j \leq b$ . Dann gilt

$$0 \leq b - \sum_{j=1}^k a_j < a_{k+1},$$

und aus (a) folgt

$$c_{\text{opt}} < \sum_{j=1}^k c_j + \frac{c_{k+1}}{a_{k+1}} a_{k+1} \leq c_{\text{Greedy}} + c_{k+1},$$

woraus (b) folgt. □

### (5.41) Bemerkung.

- (a) Der Gewichtsichten-Greedyalgorithmus kann im Falle des 0/1-Rucksack-Problems beliebig schlechte Lösungen liefern.
- (b) Führen wir sowohl den Gewichtsichten- als auch den Zielfunktions-Greedyalgorithmus für ein 0/1-Rucksack-Problem aus, so ist dieses kombinierte Verfahren ein 1/2-approximativer Algorithmus. △

**Beweis.** (a) Wir betrachten das folgende Beispiel mit  $\rho_1 \geq \rho_2$ :

$$\begin{aligned} \max \quad & x_1 + \alpha x_2 \\ & x_1 + \alpha x_2 \leq \alpha \\ & x_1, x_2 \in \{0, 1\}. \end{aligned}$$

Es gilt  $c_{\text{opt}} = \alpha$  und  $c_{\text{Ggreedy}} = 1$ , also ist der Algorithmus für kein  $\epsilon$ ,  $0 < \epsilon \leq 1$ ,  $\epsilon$ -approximativ.

(b) Gilt nach Ausführung des Gewichtslichten-Greedyalgorithmus  $c_{\text{Ggreedy}} \geq \frac{1}{2}c_{\text{opt}}$ , so sind wir fertig. Andernfalls sei

$$c_k = \max\{c_j \mid j = 1, \dots, n\}$$

und mit (5.40)(b) gilt dann  $\frac{1}{2}c_{\text{opt}} > c_{\text{Ggreedy}} > c_{\text{opt}} - c_k$ , woraus  $c_k > \frac{1}{2}c_{\text{opt}}$  folgt. Für den Wert  $c_{\text{Ggreedy}}$  des Zielfunktions-Greedyalgorithmus gilt trivialerweise  $c_{\text{Ggreedy}} \geq c_k$ . Daraus ergibt sich die Behauptung.  $\square$

(5.41)(b) ist ein bemerkenswertes Ergebnis: Die Kombination zweier „schlechter“ Heuristiken ergibt eine „gute“ Heuristik (wenn man die theoretische Approximationsgüte als Qualitätsmaßstab wählt). Dies ist ein durchaus seltenes Ergebnis.

## Literaturverzeichnis

- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Theory, Algorithms and Applications*. Paramount Publishing International, Prentice Hall, New York, 1993.
- H. Bast, D. Delling, A. Goldberg, M. Müller-Hannemann, T. Pajor, P. Sanders, D. Wagner, and R. Werneck. Route planning in transportation networks. Technical report, Microsoft Research, Redmond, 2014. URL <http://research.microsoft.com/apps/pubs/default.aspx?id=207102>.
- R. E. Bellman. On a routing problem. *Quarterly of Applied Mathematics*, 16(1):87–90, 1958.
- E. W. Dijkstra. A note on two problems in connexion with graphs. *Numer. Math.*, 1: 269–271, 1959.
- W. Domschke. *Kürzeste Wege in Graphen*. Verlag A. Hain, Meisenheim am Glan, 1972.
- R. W. Floyd. Algorithm 97, shortest path. *Communications of the ACM*, 5(6):345, 1962.
- R. Geisberger, P. Sanders, D. Schultes, and C. Vetter. Exact routing in large road networks using contraction hierarchies. *Transportation Science*, 46(3):388–404, 2012. URL <http://pubsonline.informs.org/doi/abs/10.1287/trsc.1110.0401>.
- F. Glover, D. D. Klingman, and N. V. Phillips. A new polynomially bounded shortest path algorithm. *Operations Research*, 33(1):65–73, 1985.

- A. Goldberg. Point-to-point shortest path algorithms with preprocessing. In *SOFSE 2007: Theory and Practice of Computer Science*, pages 88–102, 2007.
- R. L. Graham and P. Hell. On the history of the minimum spanning tree problem. *Annals of the History of Computing*, 7:43–57, 1982.
- H. Kellerer, U. Pferschy, and D. Pisinger. *Knapsack Problems*. Springer, 2004.
- S. O. Krumke and H. Noltemeier. *Graphentheoretische Konzepte und Algorithmen*. Teubner, Wiesbaden, 2005. ISBN 3-519-00526-3.
- E. L. Lawler. *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart & Winston, New York, 1976.
- S. Martello and P. Toth. *Knapsack problems: Algorithms and computer implementations*. J. Wiley & Sons, 1990. Verfügbar als Download unter <http://www.or.deis.unibo.it/knapsack.html>.
- K. Mehlhorn. *Data Structures and Algorithms*, volume 1–3. Springer-Verlag, EATCS Monographie edition, 1984. (dreibändige Monographie, Band I liegt auch auf deutsch im Teubner-Verlag (1986) vor).
- E. F. Moore. The shortest path through a maze. In *Proc. Int. Symp. on Theory of Switching Part II*, pages 285–292. Harvard University Press, 1959.
- J. Nešetřil and H. Nešetřilová. The origins of minimal spanning tree algorithms – Borůvka and Jarník. In *Optimization Stories*, Documenta Mathematica, pages 127–141. DMV, 2012.
- A. Schrijver. *Combinatorial Optimization – Polyhedra and Efficiency*. Springer-Verlag, Berlin, 2003.
- M. M. Syslo, N. Deo, and J. S. Kowalik. *Discrete optimization algorithms*. Prentice Hall, Englewood Cliffs, N.J., 1983.
- M. Thorup. Undirected single shortest paths in linear time. In *Proceedings of the 38th IEE Symposium on Foundations of Comp. Sci. (FOCS)*, pages 12–21, 1997.



## 6 Maximale Flüsse in Netzwerken

In diesem Kapitel behandeln wir ein in sowohl theoretischer als auch praktischer Hinsicht außerordentlich interessantes Gebiet: Flüsse in Netzwerken. Es war früher üblich und wird aus Traditionsgründen häufig weiter so gehandhabt, einen Digraphen mit Bogen gewichten bzw. -kapazitäten ein *Netzwerk* zu nennen. Sind zusätzlich zwei Knoten  $s$  und  $t$  ausgezeichnet, so spricht man von einem  $(s, t)$ -*Netzwerk*. Wir wollen diese Bezeichnung hier nur gelegentlich übernehmen, mit dem Kapiteltitle aber den historischen Bezug herstellen. Es wird sich (später) zeigen, dass die Netzwerkflusstheorie als ein Bindeglied zwischen der linearen und der ganzzahligen Optimierung aufgefasst werden kann. Netzwerkflussprobleme sind (ganzzahlige) lineare Programme, für die sehr schnelle kombinatorische Lösungsmethoden existieren. Der Dualitätssatz der linearen Programmierung hat hier eine besonders schöne Form.

Netzwerkflüsse haben folgenden Anwendungshintergrund. Wir betrachten ein Rohrleitungssystem (Abwasserkanäle, Frischwasserversorgung), bei dem die Rohre gewisse Kapazitäten (z. B. maximale Durchflussmenge pro Minute) haben. Einige typische Fragen lauten: Was ist die maximale Durchflussmenge durch das Netzwerk? Welche Wassermenge kann man maximal pro Minute vom Speicher zu einem bestimmten Abnehmer pumpen? Wieviel Regenwasser kann das System maximal aufnehmen? Ähnliches gilt für Telefonnetze. Hier sind die „Rohre“ die Verbindungen zwischen zwei Knotenpunkten, die Kapazitäten die maximalen Anzahlen von Gesprächen bzw. die maximalen Datenmengen pro Zeiteinheit, die über eine Verbindung geführt werden können. Man interessiert sich z. B. für die maximale Zahl von Gesprächen, die parallel zwischen zwei Orten (Ländern) geführt werden können, oder den größtmöglichen Datentransfer pro Zeiteinheit zwischen zwei Teilnehmern.

Darüber hinaus treten Netzwerkflussprobleme in vielfältiger Weise als Unter- oder Hilfsprobleme bei der Lösung komplizierter Anwendungsprobleme auf, z. B. bei vielen Routenplanungs- und verschiedenen Logistikproblemen. Insbesondere werden Netzwerkflussalgorithmen sehr häufig als Separierungsroutinen bei Schnittebenenverfahren eingesetzt, ein Thema von ADM II und ADM III.

Das klassische Werk der Netzwerkflusstheorie ist das Buch Ford, Jr. and Fulkerson (1962). Es ist auch heute noch lesenswert. Es gibt unzählige Veröffentlichungen zur Theorie, Algorithmik und den Anwendungen der Netzwerkflusstheorie. Durch neue algorithmische Ansätze (Präfluss-Techniken, Skalierungsmethoden) und effiziente Datenstrukturen sind Ende der 80er und zu Beginn der 90er Jahre sehr viele Artikel zu diesem Thema erschienen. Gute Darstellungen hierzu sind in den umfangreichen und sehr informativen Übersichtsartikeln Ahuja et al. (1989), Goldberg et al. (1990) und Frank (1995) zu finden. Ein sehr empfehlenswertes Buch ist Ahuja et al. (1993). Die beiden Handbücher Ball et al. (1995a), Ball et al. (1995b) enthalten umfassende Informationen zur Theorie, Al-

gorithmik und zu den Anwendungen von Netzwerken. Die vierte Auflage des Buches von Jungnickel (2013) ist ebenfalls empfehlenswert. Und natürlich gibt es eine hoch kondensierte Zusammenfassung der Netzwerkflusstheorie und -algorithmik in Schrijver (2003).

## 6.1 Das Max-Flow-Min-Cut-Theorem

Im Folgenden sei  $D = (V, A)$  ein Digraph mit Bogenkapazitäten  $c(a) \in \mathbb{R}, c(a) \geq 0$  für alle  $a \in A$ . Ferner seien  $s$  und  $t$  zwei voneinander verschiedene Knoten aus  $V$ . Der Knoten  $s$  heißt *Quelle* (englisch: source), und  $t$  heißt *Senke* (englisch: sink). Eine Funktion  $x : A \rightarrow \mathbb{R}$  (bzw. ein Vektor  $x \in \mathbb{R}^A$ ) heißt  $(s, t)$ -*Fluss*, wenn die folgenden *Flusserhaltungsbedingungen* erfüllt sind:

$$\sum_{a \in \delta^-(v)} x_a = \sum_{a \in \delta^+(v)} x_a \quad \forall v \in V \setminus \{s, t\}. \quad (6.1)$$

Erfüllt  $x$  zusätzlich die *Kapazitätsbedingungen*

$$0 \leq x_a \leq c_a \quad \forall a \in A, \quad (6.2)$$

so ist  $x$  ein *zulässiger*  $(s, t)$ -*Fluss*. Für einen  $(s, t)$ -Fluss  $x \in \mathbb{R}^A$  heißt

$$\text{val}(x) := \sum_{a \in \delta^+(s)} x_a - \sum_{a \in \delta^-(s)} x_a \quad (6.3)$$

der *Wert* des  $(s, t)$ -Flusses  $x$ . Wenn nicht extra darauf hingewiesen wird, ist mit „Fluss“ stets ein zulässiger Fluss gemeint.

Wir werden uns in diesem Abschnitt damit beschäftigen, eine Charakterisierung des maximalen Wertes eines  $(s, t)$ -Flusses zu finden. In den nächsten beiden Abschnitten werden wir Algorithmen zur Bestimmung eines maximalen zulässigen Flusses angeben.

Wir erinnern daran, dass ein  $(s, t)$ -*Schnitt* in  $D$  eine Bogenmenge der Form  $\delta^+(W) = \delta^-(V \setminus W) = \{(i, j) \in A \mid i \in W, j \in V \setminus W\}$  mit  $s \in W \subseteq V$  und  $t \in V \setminus W$  ist. Die *Kapazität eines Schnittes*  $\delta^+(W)$  ist wie üblich mit  $c(\delta^+(W)) = \sum_{a \in \delta^+(W)} c_a$  definiert. Aus der „Rohrleitungsanwendung“ wird der Name „Schnitt“ klar. Durchschneiden wir alle Rohre irgendeines Schnittes, d. h. entfernen wir alle Bögen eines  $(s, t)$ -Schnittes aus dem Digraphen, dann kann kein Fluss mehr von  $s$  nach  $t$  „fließen“. Diese triviale Beobachtung liefert:

### (6.4) Lemma.

(a) Seien  $s \in W, t \notin W$ , dann gilt für jeden zulässigen  $(s, t)$ -Fluss  $x$ :

$$\text{val}(x) = \sum_{a \in \delta^+(W)} x_a - \sum_{a \in \delta^-(W)} x_a.$$

(b) Der maximale Wert eines zulässigen  $(s, t)$ -Flusses ist höchstens so groß wie die minimale Kapazität eines  $(s, t)$ -Schnittes.  $\triangle$

**Beweis.** (a) Aus der Flusserhaltungsbedingung (6.1) folgt

$$\begin{aligned} \text{val}(x) &= \sum_{a \in \delta^+(s)} x_a - \sum_{a \in \delta^-(s)} x_a = \sum_{v \in W} \left( \sum_{a \in \delta^+(v)} x_a - \sum_{a \in \delta^-(v)} x_a \right) \\ &= \sum_{a \in \delta^+(W)} x_a - \sum_{a \in \delta^-(W)} x_a. \end{aligned}$$

(b) Seien  $\delta^+(W)$  ein beliebiger  $(s, t)$ -Schnitt und  $x$  ein zulässiger  $(s, t)$ -Fluss, dann gilt wegen (a) und (6.2):

$$\text{val}(x) = \sum_{a \in \delta^+(W)} x_a - \sum_{a \in \delta^-(W)} x_a \leq \sum_{a \in \delta^+(W)} c_a = c(\delta^+(W)). \quad \square$$

Wir werden später einen kombinatorischen Beweis dafür angeben, dass der maximale Wert eines  $(s, t)$ -Flusses gleich der minimalen Kapazität eines  $(s, t)$ -Schnittes ist. Hier wollen wir jedoch bereits eine Vorschau auf die lineare Programmierung geben, die das Max-Flow-Min-Cut-Theorem in einen allgemeineren Kontext einbettet. Wir präsentieren dieses Resultat daher als Anwendung von Resultaten, die erst später in der Vorlesung im Kapitel über lineare Optimierung behandelt werden.

Wir schreiben zunächst die Aufgabe, einen maximalen  $(s, t)$ -Flusswert in  $D$  zu finden, als lineares Programm. Dieses lautet wie folgt:

$$\begin{aligned} \max \quad & \sum_{a \in \delta^+(s)} x_a - \sum_{a \in \delta^-(s)} x_a \quad [ = x(\delta^+(s)) - x(\delta^-(s)) ] \\ & \sum_{a \in \delta^-(v)} x_a - \sum_{a \in \delta^+(v)} x_a = x(\delta^-(v)) - x(\delta^+(v)) = 0 \quad \forall v \in V \setminus \{s, t\}, \\ & x_a \leq c_a \quad \forall a \in A, \\ & x_a \geq 0 \quad \forall a \in A. \end{aligned} \quad (6.5)$$

Jede zulässige Lösung von (6.5) ist also ein zulässiger  $(s, t)$ -Fluss, und jede optimale Lösung ein maximaler  $(s, t)$ -Fluss. Um das zu (6.5) duale lineare Programm aufschreiben zu können, führen wir für jeden Knoten  $v \in V \setminus \{s, t\}$  eine Dualvariable  $z_v$  und für jeden Bogen  $a \in A$  eine Dualvariable  $y_a$  ein. Das folgende lineare Programm ist dann (im Sinne

der Theorie der linearen Optimierung) zu (6.5) dual.

$$\begin{aligned} \min \quad & \sum_{a \in A} c_a y_a \\ & y_a + z_v - z_u \geq 0 \quad \text{falls } a = (u, v) \in A(V \setminus \{s, t\}), \\ & y_a - z_u \geq -1 \quad \text{falls } a = (u, s), u \neq t, \\ & y_a + z_v \geq 1 \quad \text{falls } a = (s, v), v \neq t, \\ & y_a - z_u \geq 0 \quad \text{falls } a = (u, t), u \neq s, \\ & y_a + z_v \geq 0 \quad \text{falls } a = (t, v), v \neq s, \\ & y_a \geq 1 \quad \text{falls } a = (s, t), \\ & y_a \geq -1 \quad \text{falls } a = (t, s), \\ & y_a \geq 0 \quad \text{für alle } a \in A. \end{aligned}$$

Führen wir zusätzlich (zur notationstechnischen Vereinfachung) die Variablen  $z_s$  und  $z_t$  ein und setzen sie mit 1 bzw. 0 fest, so kann man dieses LP äquivalent, aber etwas kompakter wie folgt schreiben:

$$\begin{aligned} \min \quad & \sum_{a \in A} c_a y_a \\ & y_a + z_v - z_u \geq 0 \quad \text{für alle } a = (u, v) \in A, \\ & z_s = 1 \\ & z_t = 0 \\ & y_a \geq 0 \quad \text{für alle } a \in A. \end{aligned} \tag{6.6}$$

Wir benutzen nun (6.5) und (6.6), um folgenden berühmten Satz, der auf Ford, Jr. and Fulkerson (1956) und Elias et al. (1956) zurückgeht, zu beweisen.

**(6.7) Satz (Max-Flow-Min-Cut-Theorem).** *Gegeben seien ein Digraph  $D = (V, A)$  mit Bogenkapazitäten  $c_a \in \mathbb{R}$ ,  $c_a \geq 0$  für alle  $a \in A$ , und zwei verschiedene Knoten  $s, t \in V$ . Dann ist der maximale Wert eines zulässigen  $(s, t)$ -Flusses gleich der minimalen Kapazität eines  $(s, t)$ -Schnittes.  $\triangle$*

**Beweis.** Aufgrund von Lemma (6.4) genügt es zu zeigen, dass es einen  $(s, t)$ -Schnitt gibt, dessen Kapazität gleich dem maximalen Flusswert ist. Da alle Variablen beschränkt sind und der Nullfluss zulässig ist, hat (6.5) eine optimale Lösung. Sei also  $x^*$  ein optimaler zulässiger  $(s, t)$ -Fluss mit Wert  $\text{val}(x^*)$ . Aufgrund des Dualitätssatzes der linearen Programmierung gibt es eine Lösung, sagen wir  $y_a^*$ ,  $a \in A$  und  $z_v^*$ ,  $v \in V$ , des zu (6.5) dualen Programms (6.6) mit  $\text{val}(x^*) = \sum_{a \in A} c_a y_a^*$ . Wir setzen:  $W := \{u \in V \mid z_u^* > 0\}$  und zeigen, dass  $\delta^+(W)$  ein  $(s, t)$ -Schnitt mit  $c(\delta^+(W)) = \text{val}(x^*)$  ist.

Offenbar gilt  $s \in W$ ,  $t \notin W$ , also ist  $\delta^+(W)$  ein  $(s, t)$ -Schnitt. Ist  $a = (u, v) \in \delta^+(W)$ , dann gilt  $z_u^* > 0$ ,  $z_v^* \leq 0$  und folglich  $y_a^* \geq z_u^* - z_v^* > 0$ . Aufgrund des Satzes vom schwachen komplementären Schlupf muss dann in der zu  $y_a^*$  gehörigen Ungleichung  $x_a \leq c_a$  des primalen Programms (6.5) Gleichheit gelten. Also erfüllt der optimale  $(s, t)$ -Fluss

$x^*$  die Gleichung  $x_a^* = c_a$ . Ist  $a = (u, v) \in \delta^-(W)$ , so gilt  $z_v^* > 0$ ,  $z_u^* \leq 0$  und somit (da  $y_a^* \geq 0$ )  $y_a^* - z_u^* + z_v^* \geq z_v^* - z_u^* > 0$ . Die Ungleichung ist also „locker“. Der Satz vom komplementären Schlupf impliziert nun, dass die zugehörige Primalvariable  $x_a \geq 0$  nicht positiv sein kann. Also gilt  $x_a^* = 0$ . Daraus folgt

$$c(\delta^+(W)) = x^*(\delta^+(W)) - x^*(\delta^-(W)) = x^*(\delta^+(s)) - x^*(\delta^-(s)) = \text{val}(x^*). \quad \square$$

Vielen Lesern des Manuskripts mag der obige Beweis noch unverständlich sein. Er wurde jedoch aufgenommen, um hier schon Beispielmateriale für die Theorie der linearen Programmierung vorzubereiten.

Man beachte, dass der obige Beweis des Max-Flow-Min-Cut Theorems konstruktiv ist. Aus jeder optimalen Lösung des dualen linearen Programms können wir in polynomialer Zeit einen  $(s, t)$ -Schnitt konstruieren, der den gleichen Wert wie die Optimallösung von (6.6) hat und somit ein  $(s, t)$ -Schnitt in  $D$  minimaler Kapazität ist. Aus jedem  $(s, t)$ -Schnitt  $\delta^+(W)$  können wir durch

$$\begin{aligned} y_a &:= 1 && \text{für alle } a \in \delta^+(W) \\ y_a &:= 0 && \text{für alle } a \in A \setminus \delta^+(W) \\ z_v &:= 1 && \text{für alle } v \in W \\ z_v &:= 0 && \text{für alle } v \in V \setminus W \end{aligned}$$

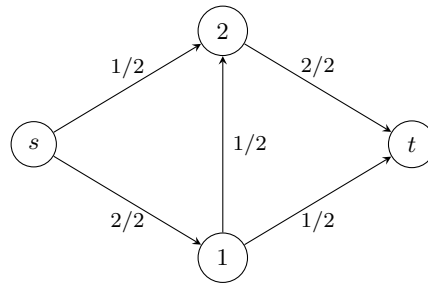
auch eine Lösung von (6.6) konstruieren, und daraus folgt, dass das lineare Programm (6.6) immer auch ganzzahlige optimale Lösungen hat. Wir können somit also das ganzzahlige Programm bestehend aus (6.6) plus Ganzzahligkeitsbedingungen für  $y_a$  und  $z_v$  durch das im Beweis von (6.7) angegebene Verfahren lösen. Wir werden im nächsten Abschnitt zeigen, dass auch das LP (6.5) immer ganzzahlige Optimallösungen hat, wenn alle Kapazitäten ganzzahlig sind. Die diesem Beweis unterliegende Konstruktion ist der Startpunkt für effiziente Algorithmen zur Lösung des Maximalflussproblems.

Das Max-Flow-Min-Cut-Theorem hat vielfältige Anwendungen in der Graphentheorie und kombinatorischen Optimierung. Aus Zeitgründen können wir an dieser Stelle nicht darauf eingehen. Wir verweisen u. a. auf Ahuja et al. (1993) und Schrijver (2003).

## 6.2 Der Ford-Fulkerson-Algorithmus

Wir haben gesehen, dass das Maximalflussproblem und das Minimalschnittproblem lineare Programme sind, folglich können wir sie effizient lösen. Das heißt, in der Praxis dürfte der Simplexalgorithmus für (6.5) und (6.6) in kurzer Zeit gute Lösungen liefern, während theoretisch die Ellipsoidmethode eine Laufzeit garantiert, die polynomial in der Inputlänge  $|V| + |A| + \sum_{a \in A} c_a$  ist. Für diese besonderen linearen Programme gibt es jedoch effiziente kombinatorische Algorithmen und Spezialversionen des Simplexalgorithmus, die außerordentlich schnell arbeiten.

Wir werden eines dieser Verfahren vorstellen, das auf Ford und Fulkerson zurückgeht. Die Idee hinter dieser Methode kann wie folgt erläutert werden. Man starte mit dem zulässigen  $(s, t)$ -Fluss, z. B. mit  $x_a = 0$  für alle  $a \in A$ . Hat man einen zulässigen  $(s, t)$ -Fluss,

Abbildung 6.1: Digraph ohne gerichteten  $(s, t)$ -Weg, der eine Flusserhöhung ermöglicht.

dann versuche man im gegebenen Digraphen einen gerichteten Weg von  $s$  nach  $t$  zu finden, auf dem zusätzlich ein positiver Wert durch das Netzwerk „geschoben“ werden kann. Geht dies, so erhöht man den gegenwärtigen Fluss und fährt fort. Die Suche nach einem gerichteten  $(s, t)$ -Weg, der die Erhöhung des Flusswertes erlaubt, führt allerdings nicht direkt zum gewünschten Erfolg. Betrachten wir z. B. den Digraphen  $D$  in Abbildung 6.1, bei dem die erste Zahl des zu einem Bogen gehörenden Zahlenpaares den gegenwärtigen Flusswert des Bogens anzeigt und die zweite Zahl die Kapazität des Bogens angibt. Der gegenwärtige Fluss hat den Wert 3, und offenbar hat der Maximalfluss den Wert 4. Es gibt im Digraphen von Abbildung 6.1 aber keinen gerichteten  $(s, t)$ -Weg auf dem der gegenwärtige Fluss verbessert werden könnte. Auf allen drei gerichteten  $(s, t)$ -Wegen ist mindestens ein Bogenfluss an seiner maximalen Kapazität. Eine Möglichkeit, den Fluss entlang eines *ungerichteten* Weges zu erhöhen, haben wir jedoch. Wir betrachten den  $[s, t]$ -Weg  $P$  mit den Bögen  $(s, 2)$ ,  $(1, 2)$ ,  $(1, t)$  und erhöhen den Fluss der Bögen  $(s, 2)$ ,  $(1, t)$  um jeweils 1, erniedrigen den Fluss durch  $(1, 2)$  um 1. Dadurch wird weder eine der Kapazitätsbedingungen (6.2) noch eine der Flusserhaltungsbedingungen verletzt, aber der Flusswert um 1 erhöht. Wir treffen daher folgende Definition.

**(6.8) Definition.** Sei  $D = (V, A)$  ein Digraph mit Bogenkapazitäten  $c_a$  für alle  $a \in A$ , seien  $s, t, v \in V$ ,  $s \neq t$ , und sei  $x$  ein zulässiger  $(s, t)$ -Fluss in  $D$ . In einem (ungerichteten)  $[s, v]$ -Weg  $P$  nennen wir einen Bogen  $(i, j)$ , der auf  $P$  in Richtung  $s$  nach  $v$  verläuft, Vorwärtsbogen, andernfalls heißt  $(i, j)$  Rückwärtsbogen.  $P$  heißt augmentierender  $[s, v]$ -Weg (bezüglich des  $(s, t)$ -Flusses  $x$ ), falls  $x_{ij} < c_{ij}$  für jeden Vorwärtsbogen  $(i, j)$  und  $x_{ij} > 0$  für jeden Rückwärtsbogen  $(i, j)$  gilt. Wenn wir nur augmentierender Weg sagen, so meinen wir immer einen augmentierenden  $[s, t]$ -Weg.  $\triangle$

Im oben angegebenen Weg  $P$  des in Abbildung 6.1 gezeigten Digraphen ist  $(1, 2)$  ein Rückwärtsbogen,  $(s, 2)$  und  $(1, t)$  sind Vorwärtsbögen.  $P$  selbst ist augmentierend bezüglich des gegebenen Flusses. Der folgende Satz liefert ein Optimalitätskriterium.

**(6.9) Satz.** Ein zulässiger  $(s, t)$ -Fluss  $x$  in einem Digraphen  $D$  mit Bogenkapazitäten ist genau dann maximal, wenn es in  $D$  keinen bezüglich  $x$  augmentierenden  $[s, t]$ -Weg gibt.  $\triangle$

**Beweis.** Ist  $P$  ein bezüglich  $x$  augmentierender  $[s, t]$ -Weg, dann sei

$$\varepsilon := \min \begin{cases} c_{ij} - x_{ij} & \text{falls } (i, j) \in P \text{ Vorwärtsbogen,} \\ x_{ij} & \text{falls } (i, j) \in P \text{ Rückwärtsbogen.} \end{cases} \quad (6.10)$$

Setzen wir

$$x'_{ij} := \begin{cases} x_{ij} + \varepsilon & \text{falls } (i, j) \in P \text{ Vorwärtsbogen,} \\ x_{ij} - \varepsilon & \text{falls } (i, j) \in P \text{ Rückwärtsbogen,} \\ x_{ij} & \text{falls } (i, j) \in A \setminus P, \end{cases} \quad (6.11)$$

dann ist offenbar  $x'_{ij}$  ein zulässiger  $(s, t)$ -Fluss mit  $\text{val}(x') = \text{val}(x) + \varepsilon$ . Also kann  $x$  nicht maximal sein.

Angenommen  $x$  besitzt keinen augmentierenden Weg. Dann sei  $W$  die Knotenmenge, die aus  $s$  und denjenigen Knoten  $v \in V$  besteht, die von  $s$  aus auf einem bezüglich  $x$  augmentierenden  $[s, v]$ -Weg erreicht werden können. Definition (6.8) impliziert  $x_a = c_a$  für alle  $a \in \delta^+(W)$  und  $x_a = 0$  für alle  $a \in \delta^-(W)$ . Daraus ergibt sich  $\text{val}(x) = x(\delta^+(W)) - x(\delta^-(W)) = x(\delta^+(W)) = c(\delta^+(W))$ . Aufgrund von Lemma (6.4)(b) ist somit  $x$  maximal.  $\square$

Der Beweis von Satz (6.9) liefert einen Schnitt  $\delta^+(W)$  mit  $\text{val}(x) = c(\delta^+(W))$ . Zusammen mit Lemma (6.4)(b) ergibt dies einen kombinatorischen Beweis des Max-Flow-Min-Cut-Theorems. Aus dem Beweis von Satz (6.9) folgt ebenfalls, dass das lineare Programm (6.5) ganzzahlige Optimallösungen hat, falls alle Kapazitäten ganzzahlig sind.

**(6.12) Satz.** Sei  $D = (V, A)$  ein Digraph mit ganzzahligen Bogenkapazitäten  $c_a \geq 0$ , und seien  $s, t \in V$ ,  $s \neq t$ . Dann gibt es einen maximalen zulässigen  $(s, t)$ -Fluss, der ganzzahlig ist.  $\triangle$

**Beweis.** Wir führen einen Induktionsbeweis über die Anzahl der „Additionen“ augmentierender Wege. Wir starten mit dem Nullfluss. Haben wir einen ganzzahligen Flussvektor und ist dieser nicht maximal, so bestimmen wir den Wert  $\varepsilon$  durch (6.10). Nach Voraussetzung ist  $\varepsilon$  ganzzahlig, und folglich ist der neue durch (6.11) festgelegte Flussvektor ebenfalls ganzzahlig. Bei jeder Augmentierung erhöhen wir den Flusswert um mindestens eins. Da der maximale Flusswert endlich ist, folgt die Behauptung aus Satz (6.9).  $\square$

Wir können nun den Ford-Fulkerson-Algorithmus angeben:

**(6.13) Ford-Fulkerson-Algorithmus.**

**Eingabe:** Digraph  $D = (V, A)$  mit Bogenkapazitäten  $c_a \in \mathbb{R}$ ,  $c_a \geq 0$  für alle Bögen  $a \in A$  und zwei Knoten  $s, t \in V$ ,  $s \neq t$ .

**Ausgabe:** Ein zulässiger  $(s, t)$ -Fluss  $x$  mit maximalem Wert  $\text{val}(x)$  und ein kapazitätsminimaler  $(s, t)$ -Schnitt  $\delta^+(W)$ .

1. (Initialisierung) Sei  $x = (x_{ij}) \in \mathbb{R}^A$  ein zulässiger  $(s, t)$ -Fluss. Hier verwendet man am besten eine schnelle Heuristik zur Auffindung eines „guten“ Flusses. Wenn einem nichts einfällt, setzt man z. B.  $x_{ij} = 0$  für alle  $(i, j) \in A$ .

Lege folgende Datenstrukturen an:

## 6 Maximale Flüsse in Netzwerken

$W$	Menge der markierten Knoten
$U$	Menge der markierten, aber noch nicht überprüften Knoten
VOR	$(n - 1)$ -Vektor, in dem der Vorgänger eines Knoten $v$ auf einem augmentierenden $[s, v]$ -Weg gespeichert wird
EPS	$(n - 1)$ -Vektor, zur sukzessiven Berechnung von (6.10)

### Markieren und Überprüfen

- Setze  $W := \{s\}$ ,  $U := \{s\}$ ,  $\text{EPS}(s) := +\infty$ .
- Ist  $U = \emptyset$ , dann gehe zu 9.
- Wähle einen Knoten  $i \in U$  aus und setze  $U := U \setminus \{i\}$ .
- Führe für alle Bögen  $(i, j) \in A$  mit  $j \notin W$  Folgendes aus:  
Ist  $x_{ij} < c_{ij}$ , dann setze

$$\begin{aligned}\text{EPS}(j) &:= \min\{c_{ij} - x_{ij}, \text{EPS}(i)\}, \\ \text{VOR}(j) &:= +i, \quad W := W \cup \{j\}, \quad U := U \cup \{j\}.\end{aligned}$$

- Führe für alle Bögen  $(j, i) \in A$  mit  $j \notin W$  Folgendes aus:  
Ist  $x_{ji} > 0$ , dann setze

$$\begin{aligned}\text{EPS}(j) &:= \min\{x_{ji}, \text{EPS}(i)\}, \\ \text{VOR}(j) &:= -i, \quad W := W \cup \{j\}, \quad U := U \cup \{j\}.\end{aligned}$$

- Gilt  $t \in W$ , gehe zu 8, andernfalls zu 3.

### Augmentierung

- Konstruiere einen augmentierenden Weg und erhöhe den gegenwärtigen Fluss um  $\text{EPS}(t)$ , d. h. bestimme  $j_1 = |\text{VOR}(t)|$ , falls  $\text{VOR}(t) > 0$ , setze  $x_{j_1 t} := x_{j_1 t} + \text{EPS}(t)$ , andernfalls setze  $x_{t j_1} := x_{t j_1} - \text{EPS}(t)$ . Dann bestimme  $j_2 = |\text{VOR}(j_1)|$ , falls  $\text{VOR}(j_1) > 0$ , setze  $x_{j_2 j_1} := x_{j_2 j_1} + \text{EPS}(t)$ , andernfalls  $x_{j_1 j_2} := x_{j_1 j_2} - \text{EPS}(t)$  usw. bis der Knoten  $s$  erreicht ist. Gehe zu 2.

### Bestimmung eines minimalen Schnittes

- Der gegenwärtige  $(s, t)$ -Fluss  $x$  ist maximal und  $\delta^+(W)$  ist ein  $(s, t)$ -Schnitt minimaler Kapazität. STOP. △

Aus den Sätzen (6.9) und (6.12) folgt, dass Algorithmus (6.13) für ganzzahlige Kapazitäten korrekt arbeitet und nach endlicher Zeit abbricht. Sind die Daten rational, so kann man (wie üblich) alle Kapazitäten mit dem kleinsten gemeinsamen Vielfachen ihrer Nenner multiplizieren. Man erhält so ein äquivalentes ganzzahliges Maximalflussproblem. Also funktioniert (6.13) auch bei rationalen Daten. Lässt man (zumindest theoretisch)



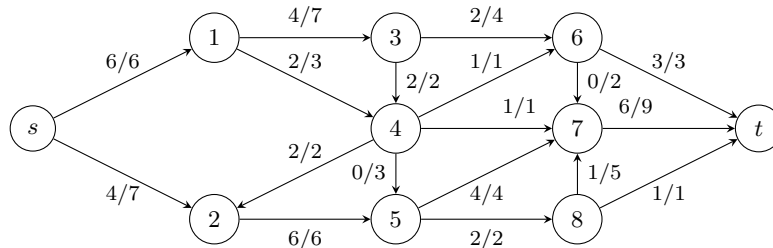


Abbildung 6.2: Beispiel-Digraph für den Ford-Fulkerson-Algorithmus.

auch irrationale Kapazitäten zu, so kann man Beispiele konstruieren, bei denen Algorithmus (6.13) nicht nach endlicher Zeit abbricht. Aber auch bei ganzzahligen Daten gibt es Probleme. Ein Durchlauf der Markierungs- und Überprüfungsphase und der Augmentierungsphase kann offenbar in  $O(m)$ ,  $m = |A|$ , Schritten durchgeführt werden. Nach jedem Durchlauf wird der Flusswert um mindestens 1 erhöht. Ist also  $v$  der Wert des maximalen  $(s, t)$ -Flusses, so ist die Laufzeit von (6.13)  $O(m \cdot v)$ . Diese Laufzeit ist nicht polynomial in  $n + m + \sum_{a \in A} \langle c_a \rangle$ , und wenn man die im Verfahren (6.13) noch nicht exakt spezifizierten Schritte ungeschickt ausführt, kann man tatsächlich zu exorbitanten Laufzeiten kommen. Allerdings haben Edmonds and Karp (1972) gezeigt:

**(6.14) Satz.** *Falls in Algorithmus (6.13) jeder Augmentierungsschritt entlang eines augmentierenden  $[s, t]$ -Weges mit minimaler Bogenzahl durchgeführt wird, dann erhält man einen Maximalfluss nach höchstens  $\frac{mn}{2}$  Augmentierungen. Also ist die Gesamtlaufzeit dieser Version des Verfahrens (6.13)  $O(m^2n)$ .  $\triangle$*

Satz (6.14) gilt für beliebige (auch irrationale) Bogenkapazitäten. Es ist in diesem Zusammenhang interessant zu bemerken, dass praktisch jeder, der Verfahren (6.13) implementiert, die Edmonds-Karp-Vorschrift einhält. Üblicherweise arbeitet man die Knoten in Breadth-First-Reihenfolge ab. Dies führt zu augmentierenden Wegen minimaler Bogenzahl. Das heißt, man implementiert die Menge  $U$  der markierten und noch nicht abgearbeiteten Knoten als Schlange. Wird ein Knoten in Schritt 5 oder 6 zu  $U$  hinzugefügt, so kommt er an das Ende der Schlange. In Schritt 4 wird immer der Knoten  $i \in U$  gewählt, der am Anfang der Schlange steht.

**(6.15) Beispiel.** Wir betrachten den in Abbildung 6.2 dargestellten Digraphen. Die erste Zahl des Zahlenpaares bei einem Bogen gibt den gegenwärtigen Fluss durch den Bogen an, die zweite die Kapazität des Bogens. In Abbildung 6.2 starten wir also mit einem Fluss des Wertes 10.

Wir führen einen Durchlauf der Markierungs- und Überprüfungsphase vor. Im weiteren sei

$$\begin{aligned} \text{VOR} &= (\text{VOR}(1), \text{VOR}(2), \dots, \text{VOR}(8), \text{VOR}(t)) \\ \text{EPS} &= (\text{EPS}(1), \text{EPS}(2), \dots, \text{EPS}(8), \text{EPS}(t)). \end{aligned}$$

Das Verfahren beginnt wie folgt:

## 6 Maximale Flüsse in Netzwerken

2.  $W := \{s\}, U := \{s\}$ .
3. —
4. Wir wählen  $s \in U$  und setzen  $U := \emptyset$ .
5.  $W := \{s, 2\}, U := \{2\}, \text{VOR} = (-, +s, -, -, -, -, -, -),$   
 $\text{EPS} = (-, 3, -, -, -, -, -, -).$
6. —
7. —
3. —
4. Wir wählen  $2 \in U, U := \emptyset$ .
5. —
6.  $W := \{s, 2, 4\}, U := \{4\}, \text{VOR} = (-, +s, -, -2, -, -, -, -),$   
 $\text{EPS} = (-, 3, -, 2, -, -, -, -).$
7. —
3. —
4. Wir wählen  $4 \in U, U := \emptyset$ .
5.  $W := \{s, 2, 4, 5\}, U := \{5\}, \text{VOR} = (-, +s, -, -2, +4, -, -, -),$   
 $\text{EPS} = (-, 3, -, 2, 2, -, -, -).$
6.  $W := \{s, 2, 4, 5, 1, 3\}, U := \{5, 1, 3\}, \text{VOR} = (-4, +s, -4, -2, +4, -, -, -),$   
 $\text{EPS} = (2, 3, 2, 2, 2, -, -, -).$
7. —
3. —
4. Wir wählen  $5 \in U, U := \{1, 3\}$ .
5. —
6. —
7. —
3. —
4. Wir wählen  $1 \in U, U := \{3\}$ .
5. —
6. —
7. —
3. —
4. Wir wählen  $3 \in U, U := \emptyset$ .
5.  $W := \{s, 2, 4, 5, 1, 3, 6\}, U := \{6\}, \text{VOR} = (-4, +s, -4, -2, +4, +3, -, -, -),$   
 $\text{EPS} = (2, 3, 2, 2, 2, 2, -, -, -).$
6. —
7. —
3. —
4. Wir wählen  $6 \in U, U := \emptyset$ .
5.  $W := \{s, 2, 4, 5, 1, 3, 6, 7\}, U := \{7\}, \text{VOR} = (-4, +s, -4, -2, +4, +3, +6, -, -),$   
 $\text{EPS} = (2, 3, 2, 2, 2, 2, 2, -, -).$
6. —
7. —
3. —
4. Wir wählen  $7 \in U, U := \emptyset$ .
5.  $W := \{s, 2, 4, 5, 1, 3, 6, 7, t\}, U := \{t\}, \text{VOR} = (-4, +s, -4, -2, +4, +3, +6, -, +7),$   
 $\text{EPS} = (2, 3, 2, 2, 2, 2, 2, -, 2).$
6. (Hier wird noch 8 markiert, das ist aber irrelevant, da  $t$  bereits markiert ist.)
7.  $t \in W$

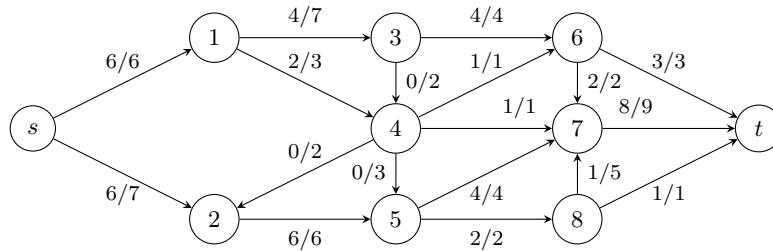


Abbildung 6.3: Maximaler Fluss für den Beispiel-Digraph aus Abbildung 6.2.

8. Es gilt

$$\text{VOR}(t) = +7$$

$$\text{VOR}(7) = +6$$

$$\text{VOR}(6) = +3$$

$$\text{VOR}(3) = -4$$

$$\text{VOR}(4) = -2$$

$$\text{VOR}(2) = +s$$

also ist der augmentierende Weg mit  $\text{EPS}(t) = 2$  der folgende:  $(s, 2)$ ,  $(4, 2)$ ,  $(3, 4)$ ,  $(3, 6)$ ,  $(6, 7)$ ,  $(7, t)$ . Der neue Fluss ist in Abbildung 6.3 dargestellt. Dieser  $(s, t)$ -Fluss ist maximal, ein  $(s, t)$ -Schnitt minimaler Kapazität ist  $\delta^+(\{s, 2\})$ , ein anderer  $\delta^+(\{s, 1, 2, 3, 4, 5\})$ .  $\triangle$

## 6.3 Einige Anwendungen

In diesem Abschnitt geht es nicht um praktische Anwendungen, sondern um Anwendungen der im Vorhergehenden angegebenen Sätze und Algorithmen zur Lösung anderer mathematischer (Optimierungs-)Probleme.

**Matchings maximaler Kardinalität in bipartiten Graphen** In (3.9) haben wir das bipartite Matchingproblem kennengelernt. Wir wollen nun zeigen, wie man die Kardinalitätsversion dieses Problems, d. h. alle Kantengewichte sind 1, mit Hilfe eines Maximalflussverfahrens lösen kann.

Ist also  $G = (V, E)$  ein bipartiter Graph mit Bipartition  $V_1, V_2$ , so definieren wir einen Digraphen  $D = (W, A)$  wie folgt. Wir wählen zwei neue Knoten, sagen wir  $s$  und  $t$ , und setzen  $W := V \cup \{s, t\}$ . Die Bögen von  $D$  seien die folgenden. Ist  $e = uv \in E$  eine Kante von  $G$ , so geben wir dieser die Richtung von  $V_1$  nach  $V_2$ . Ist also  $u \in V_1$  und  $v \in V_2$ , so wird aus  $uv \in E$  der Bogen  $(u, v)$  andernfalls der Bogen  $(v, u)$ . Ferner enthält  $D$  die Bögen  $(s, u)$  für alle  $u \in V_1$  und die Bögen  $(v, t)$  für alle  $v \in V_2$ . Alle Bögen von  $D$  erhalten die Kapazität 1. Die Konstruktion von  $D$  aus  $G$  ist in Abbildung 6.4 an einem Beispiel dargestellt.

## 6 Maximale Flüsse in Netzwerken

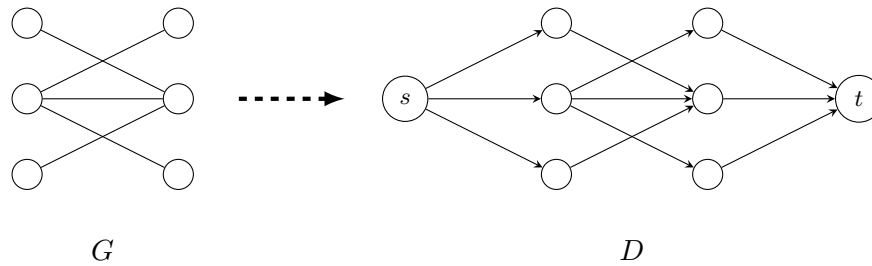


Abbildung 6.4: Transformation eines bipartiten Matchingproblems in ein Maximalflussproblem.

**(6.16) Satz.** *Ist  $G$  ein bipartiter Graph und  $D$  der wie oben angegeben aus  $G$  konstruierte Digraph, dann ist der Wert eines maximalen zulässigen  $(s, t)$ -Flusses  $x$  in  $D$  gleich dem Wert eines kardinalitätsmaximalen Matchings in  $D$ . Ferner kann ein kardinalitätsmaximales Matching  $M$  direkt aus  $x$  konstruiert werden.  $\triangle$*

**Beweis.** Hausaufgabe.  $\square$

**Zusammenhangsprobleme in Graphen und Digraphen** Mit Hilfe von Maximalflussalgorithmen können ferner für einen Digraphen die starke Zusammenhangszahl und die starke Bogenzusammenhangszahl in polynomialer Zeit bestimmt werden. Analog können in einem ungerichteten Graphen die Zusammenhangszahl und die Kantenzusammenhangszahl in polynomialer Zeit ermittelt werden.

**Mehrere Quellen und Senken** Die Festlegung, dass wir in einem Digraphen einen Fluss von nur einer Quelle zu nur einer Senke schicken wollen, scheint auf den ersten Blick eine starke Einschränkung zu sein. Jedoch können Maximalflussprobleme mit mehreren Quellen und Senken sehr einfach auf das von uns behandelte Problem zurückgeführt werden.

Gegeben sei ein Digraph  $D = (V, A)$  mit Bogenkapazitäten  $c(a) \geq 0$  für alle  $a \in A$ . Ferner seien  $S = \{s_1, \dots, s_p\} \subseteq V$  Quellen und  $T = \{t_1, \dots, t_q\} \subseteq V$  Senken. Es gelte  $S \cap T = \emptyset$ . Ein zulässiger  $(S, T)$ -Fluss in  $D$  ist ein Vektor  $x \in \mathbb{R}^A$  mit folgenden Eigenschaften

$$\begin{aligned} 0 \leq x_a \leq c_a & \quad \text{für alle } a \in A, \\ x(\delta^-(v)) = x(\delta^+(v)) & \quad \text{für alle } v \in V \setminus (S \cup T). \end{aligned}$$

Der Wert eines zulässigen  $(S, T)$ -Flusses  $x$  ist definiert als

$$\text{val}(x) := \sum_{s \in S} (x(\delta^+(s)) - x(\delta^-(s))).$$

Die Bestimmung eines maximalen  $(S, T)$ -Flusses in  $D$  kann wie folgt auf ein Maximalflussproblem in einem Digraphen  $D' = (V', A')$  mit einer Quelle und einer Senke zurück-

geführt werden. Wir wählen zwei neue Knoten  $s, t$  und setzen

$$V' := V \cup \{s, t\}.$$

Der Knoten  $s$  ist die Quelle,  $t$  ist die Senke von  $D'$ . Ferner sei

$$\begin{aligned} A' &:= A \cup \{(s, s_i) \mid i = 1, \dots, p\} \cup \{(t_i, t) \mid i = 1, \dots, q\}, \\ c'(a) &:= c(a) \text{ für alle } a \in A, \\ c(a) &:= M \text{ für alle } a \in A' \setminus A. \end{aligned}$$

Es reicht z.B.  $M := \sum_{a \in A} c(a) + 1$  zu wählen. Man überlegt sich sofort, dass jedem zulässigen  $(s, t)$ -Fluss in  $D'$  ein zulässiger  $(S, T)$ -Fluss in  $D$  mit gleichem Wert entspricht. Also liefert ein maximaler  $(s, t)$ -Fluss in  $D'$  einen maximalen  $(S, T)$ -Fluss in  $D$ .

**Separationsalgorithmen** Maximalfluss-Algorithmen spielen eine wichtige Rolle in sogenannten Schnittebenenverfahren für die ganzzahlige Optimierung. So treten etwa bei der Lösung von Travelling-Salesman-Problemen und Netzwerkentwurfsproblemen (Telekommunikation, Wasser- und Stromnetzwerke) Ungleichungen des Typs

$$\sum_{u \in W} \sum_{v \in V \setminus W} x_{uv} \geq f(w) \quad \forall W \subseteq V$$

auf, wobei  $f(w)$  eine problemspezifische Funktion ist. Die Anzahl dieser Ungleichungen ist exponentiell in  $|V|$ . Häufig kann man jedoch in einer Laufzeit, die polynomial in  $|V|$  ist, überprüfen, ob für einen gegebenen Vektor  $x^*$  alle Ungleichungen dieser Art erfüllt sind oder ob  $x^*$  eine der Ungleichungen verletzt. Algorithmen, die so etwas leisten, werden Separationsalgorithmen genannt. Beim TSP zum Beispiel können die „Schnittungleichungen“ durch Bestimmung eines kapazitätsminimalen Schnittes (mit Hilfe eines Maximalflussalgorithmus) überprüft werden.

## Literaturverzeichnis

- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Handbooks in Operations Research and Management Science*, volume 1, chapter Optimization, pages 211–360. Elsevier, North-Holland, Amsterdam, 1989.
- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Theory, Algorithms and Applications*. Paramount Publishing International, Prentice Hall, New York, 1993.
- M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors. *Handbooks in Operations Research and Management Science*, volume 7: Network Models. North-Holland, Amsterdam, 1995a.
- M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors. *Handbooks in Operations Research and Management Science*, volume 8: Network Routing. North-Holland, Amsterdam, 1995b.

## Literaturverzeichnis

- J. Edmonds and R. M. Karp. Theoretical improvement in algorithmic efficiency of network flow problems. *J. ACM*, (19):248–264, 1972.
- P. Elias, A. Feinstein, and C. E. Shannon. Note on maximum flow through a network. *IRE Trans. on Inform. Theory*, (2):117–119, 1956.
- L. R. Ford, Jr. and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.
- L. R. Ford, Jr. and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, 1962.
- A. Frank. Connectivity and network flows. In R. L. Graham et al., editors, *Handbook of Combinatorics*, chapter 2, pages 111–177. North-Holland, Amsterdam, 1995.
- V. Goldberg, É. Tardos, and R. E. Tarjan. Network Flow Algorithms. In B. Korte, L. Lovasz, H. J. Prömel, and A. Schrijver, editors, *Paths, Flows, and VLSI-Layout*, pages 101–164. Springer-Verlag, Berlin, 1990.
- D. Jungnickel. *Graphs, Networks and Algorithms*. Springer, 4. edition, 2013.
- A. Schrijver. *Combinatorial Optimization – Polyhedra and Efficiency*. Springer-Verlag, Berlin, 2003.

# 7 Flüsse mit minimalen Kosten

Das im vorhergehenden Kapitel behandelte Maximalflussproblem ist eines der Basisprobleme der Netzwerkflusstheorie. Es gibt noch weitere wichtige und anwendungsreiche Netzwerkflussprobleme. Wir können hier jedoch aus Zeitgründen nur wenige dieser Probleme darstellen und analysieren. Der Leser, der an einer vertieften Kenntnis der Netzwerkflusstheorie interessiert ist, sei auf die am Anfang von Kapitel 6 erwähnten Bücher und Übersichtsartikel verwiesen.

## 7.1 Flüsse mit minimalen Kosten

Häufig tritt das Problem auf, durch ein Netzwerk nicht einen maximalen Fluss senden zu wollen, sondern einen Fluss mit vorgegebenem Wert, der bezüglich eines Kostenkriteriums minimale Kosten verursacht. Wir wollen hier nur den Fall einer linearen Kostenfunktion behandeln, obwohl gerade auch konkave und stückweise lineare Kosten (bei Mengenrabatten) eine wichtige Rolle spielen.

Sind ein Digraph  $D = (V, A)$  mit Bogenkapazitäten  $c(a) \geq 0$  für alle  $a \in A$  und Kostenkoeffizienten  $w(a)$  für alle  $a \in A$  gegeben, sind  $s, t \in V$  zwei verschiedene Knoten, und ist  $f$  ein vorgegebener Flusswert, dann nennt man die Aufgabe, einen zulässigen  $(s, t)$ -Fluss  $x$  mit Wert  $f$  zu finden, dessen Kosten  $\sum_{a \in A} w_a x_a$  minimal sind, ein *Minimalkosten-Netzwerkflussproblem*. Analog zur LP-Formulierung (6.5) des Maximalflussproblems kann man ein Minimalkosten-Flussproblem als lineares Programm darstellen. Offenbar ist jede Optimallösung des linearen Programms (7.1) ein kostenminimaler  $c$ -kapazitierter  $(s, t)$ -Fluss mit Wert  $f$ .

$$\begin{aligned} \min \quad & \sum_{a \in A} w_a x_a \\ & x(\delta^-(v)) - x(\delta^+(v)) = 0 \quad \forall v \in V \setminus \{s, t\} \\ & x(\delta^-(t)) - x(\delta^+(t)) = f \\ & 0 \leq x_a \leq c_a \quad \forall a \in A \end{aligned} \tag{7.1}$$

Minimalkosten-Flussprobleme kann man daher mit Algorithmen der linearen Optimierung lösen. In der Tat gibt es besonders schnelle Spezialversionen des Simplexalgorithmus für Probleme des Typs (7.1). Sie werden *Netzwerk-Simplexalgorithmen* genannt. Wir werden in Abschnitt 7.3 genauer auf diesen Algorithmus eingehen.

Es gibt viele kombinatorische Spezialverfahren zur Lösung von Minimalkosten-Flussproblemen. Alle „Tricks“ der kombinatorischen Optimierung und Datenstrukturtechniken der Informatik werden benutzt, um schnelle Lösungsverfahren für (7.1) zu produzieren.

Ein Ende ist nicht abzusehen. Es gibt (zurzeit) kein global bestes Verfahren, weder bezüglich der beweisbaren Laufzeit, noch in Bezug auf Effizienz im praktischen Einsatz. Die Literatur ist allerdings voll mit Tabellen mit derzeitigen „Weltrekorden“ bezüglich der worst-case-Laufzeit unter speziellen Annahmen an die Daten. Alle derzeit gängigen Verfahren können – gut implementiert – Probleme des Typs (7.1) mit Zigtausenden von Knoten und Hunderttausenden oder gar Millionen von Bögen mühelos lösen.

Wir haben in dieser Vorlesung nicht genügend Zeit, um auf diese Details und Feinheiten einzugehen. Wir werden zunächst ein kombinatorisches Verfahren und die zugrundeliegende Theorie vorstellen. Um den Algorithmus und den Satz, auf dem seine Korrektheit beruht, darstellen zu können, führen wir einige neue Begriffe ein.

Sei  $x$  ein zulässiger  $(s, t)$ -Fluss in  $D$  und sei  $C$  ein (nicht notwendigerweise gerichteter) Kreis in  $D$ . Diesem Kreis  $C$  können wir offenbar zwei Orientierungen geben. Ist eine Orientierung von  $C$  gewählt, so nennen wir einen Bogen auf  $C$ , der in Richtung der Orientierung verläuft, *Vorwärtsbogen*, andernfalls nennen wir ihn *Rückwärtsbogen*. Ein Kreis  $C$  heißt *augmentierend* bezüglich  $x$ , wenn es eine Orientierung von  $C$  gibt, so dass  $x_a < c_a$  für alle Vorwärtsbögen  $a \in C$  und dass  $0 < x_a$  für alle Rückwärtsbögen  $a \in C$  gilt (vergleiche Definition (6.8)). Ein Kreis kann offenbar bezüglich beider, einer oder keiner Richtung augmentierend sein. Sprechen wir von einem augmentierenden Kreis  $C$ , so unterstellen wir fortan, dass eine Orientierung von  $C$  fest gewählt ist, bezüglich der  $C$  augmentierend ist.

Die Summe der Kostenkoeffizienten der Vorwärtsbögen minus der Summe der Kostenkoeffizienten der Rückwärtsbögen definieren wir als die *Kosten eines augmentierenden Kreises*. (Wenn ein Kreis in Bezug auf beide Orientierungen augmentierend ist, können die beiden Kosten verschieden sein!) Das zentrale Resultat dieses Abschnitts ist das Folgende.

**(7.2) Satz.** *Ein zulässiger  $(s, t)$ -Fluss  $x$  in  $D$  mit Wert  $f$  hat genau dann minimale Kosten, wenn es bezüglich  $x$  keinen augmentierenden Kreis mit negativen Kosten gibt.  $\Delta$*

**Beweis.** Wir beweisen zunächst nur die triviale Richtung. (Satz (7.6) liefert die Rückrichtung.) Gibt es einen augmentierenden Kreis  $C$  bezüglich  $x$ , so setzen wir:

$$\varepsilon := \min \begin{cases} c_{ij} - x_{ij} & (i, j) \in C \text{ Vorwärtsbogen,} \\ x_{ij} & (i, j) \in C \text{ Rückwärtsbogen.} \end{cases} \quad (7.3)$$

Definieren wir

$$x'_{ij} := \begin{cases} x_{ij} + \varepsilon & \text{falls } (i, j) \in C \text{ Vorwärtsbogen,} \\ x_{ij} - \varepsilon & \text{falls } (i, j) \in C \text{ Rückwärtsbogen,} \\ x_{ij} & \text{falls } (i, j) \in A \setminus C, \end{cases} \quad (7.4)$$

dann ist  $x' \in \mathbb{R}^A$  trivialerweise ein zulässiger  $(s, t)$ -Fluss mit Wert  $\text{val}(x') = f$ . Hat der augmentierende Kreis  $C$  negative Kosten  $\gamma < 0$ , dann gilt offenbar  $\sum_{(i,j) \in A} w_{ij} x'_{ij} = \sum_{(i,j) \in A} w_{ij} x_{ij} + \varepsilon \gamma$ . Gibt es also einen augmentierenden Kreis bezüglich  $x$  mit negativen Kosten, dann kann  $x$  nicht kostenminimal sein.  $\square$



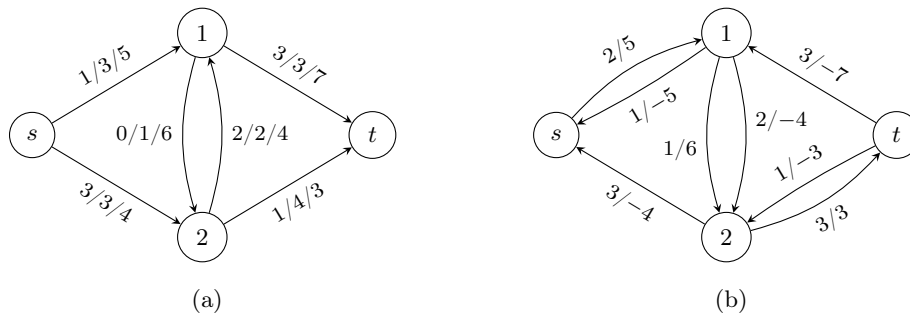


Abbildung 7.1: Ein Digraph  $D$  mit Fluss  $x$  und ein augmentierendes Netzwerk bzgl.  $x$ .

Um die umgekehrte Richtung zu beweisen, müssen wir etwas mehr Aufwand treiben, den wir allerdings direkt bei der Darstellung des Algorithmus benutzen können. Ist  $x$  ein zulässiger  $(s, t)$ -Fluss mit Wert  $f$ , dann definieren wir einen Digraphen (genannt *augmentierendes Netzwerk bezüglich  $x$* )  $N = (V, \bar{A}, \bar{c}, \bar{w})$  wie folgt: Es sei

$$A_1 := \{(u, v) \in A \mid x_{uv} < c_{uv}\}, \quad A_2 := \{(v, u) \mid (u, v) \in A \text{ und } x_{uv} > 0\}.$$

Ist  $a \in A$ , so schreiben wir  $a_1$  bzw.  $a_2$  um den zugehörigen Bogen aus  $A_1$  bzw.  $A_2$  zu bezeichnen. Schreiben wir für  $a \in A$  eine Formel wie etwa  $x'(a) = x(a) + \bar{x}(a_1) - \bar{x}(a_2)$  und ist einer der Bögen  $a_1$  bzw.  $a_2$  nicht definiert (d. h. es gilt entweder  $x_a = c_a$  oder  $x_a = 0$ ), dann ist der Wert  $\bar{x}(a_1)$  bzw.  $\bar{x}(a_2)$  als Null zu betrachten. Wir setzen  $\bar{A} := A_1 \dot{\cup} A_2$  (man beachte, dass  $\bar{A}$  parallele Bögen enthalten kann). Ferner sei für  $\bar{a} \in \bar{A}$

$$\bar{c}(\bar{a}) := \begin{cases} c(a) - x(a) & \text{falls } \bar{a} = a_1, \\ x(a) & \text{falls } \bar{a} = a_2, \end{cases}$$

$$\bar{w}(\bar{a}) := \begin{cases} w(a) & \text{falls } \bar{a} = a_1, \\ -w(a) & \text{falls } \bar{a} = a_2. \end{cases}$$

In Abbildung 7.1(a) ist ein Digraph mit einem  $(s, t)$ -Fluss  $x$  des Wertes 4 dargestellt. Die drei Zahlen bei einem Bogen  $a$  geben an: Fluss durch  $a$  / Kapazität von  $a$  / Kosten von  $a$ . Das augmentierende Netzwerk  $N$  bezüglich  $D$  und  $x$  ist in 7.1(b) gezeichnet. Die beiden Zahlen bei einem Bogen  $a$  in 7.1(b) geben an: Kapazität von  $a$  / Kosten von  $a$ .

Der (ungerichtete) Kreis  $\{(2, t), (1, t), (2, 1)\}$  in Abb. 7.1(a) ist ein augmentierender Kreis mit Kosten  $3 - 7 - 4 = -8$ . Dieser Kreis entspricht in (b) dem gerichteten Kreis  $\{(2, t), (t, 1), (1, 2)\}$ , wobei von den beiden zwischen 1 und 2 parallel verlaufenden Bögen natürlich der mit negativen Kosten zu wählen ist. Aufgrund der Konstruktion von  $N$  ist folgende Beziehung offensichtlich:

**(7.5) Lemma.** *Ist  $D = (V, A)$  ein Digraph mit Kapazitäten  $c \in \mathbb{R}_+^A$  und Kosten  $w \in \mathbb{R}^A$ , ist  $x$  ein zulässiger  $(s, t)$ -Fluss in  $D$ , und ist  $N = (V, \bar{A}, \bar{c}, \bar{w})$  das zu  $D$  und  $x$  gehörige augmentierende Netzwerk, dann entspricht jeder augmentierende Kreis in  $D$  genau einem gerichteten Kreis in  $N$ . Die Kosten eines augmentierenden Kreises in  $D$  stimmen überein mit der Summe der Kostenkoeffizienten des zugehörigen gerichteten Kreises in  $N$ .  $\triangle$*

## 7 Flüsse mit minimalen Kosten

Damit ist unser Exkurs zur Definition von augmentierenden Netzwerken beendet. Wir formulieren nun Theorem (7.2) unter Benutzung dieses neuen Konzepts um.

**(7.6) Satz.** *Sei  $x$  ein zulässiger  $(s, t)$ -Fluss in  $D$  mit Wert  $f$ , und  $N = (V, \bar{A}, \bar{c}, \bar{w})$  sei das bezüglich  $x$  und  $D$  augmentierende Netzwerk, dann gilt folgendes. Der Fluss  $x$  ist unter allen zulässigen  $(s, t)$ -Flüssen in  $D$  mit Wert  $f$  genau dann kostenminimal, wenn es in  $N$  keinen gerichteten Kreis mit negativen Kosten gibt.  $\triangle$*

**Beweis.** Gibt es in  $N$  einen gerichteten Kreis mit negativen Kosten, so folgt analog zum Beweis des einfachen Teils von (7.2), dass  $x$  nicht minimal ist.

Nehmen wir umgekehrt an, dass  $x$  nicht kostenminimal ist, dann müssen wir in  $N$  einen gerichteten Kreis mit negativen Kosten finden. Sei also  $x'$  ein zulässiger  $(s, t)$ -Fluss in  $D$  mit Wert  $f$  und  $w^T x' < w^T x$ . Für jeden Bogen  $\bar{a} \in \bar{A}$  setzen wir

$$\bar{x}(\bar{a}) := \begin{cases} \max\{0, x'(a) - x(a)\}, & \text{falls } \bar{a} = a_1 \in A_1 \\ \max\{0, x(a) - x'(a)\}, & \text{falls } \bar{a} = a_2 \in A_2 \end{cases}$$

Diese Definition ist genau so gewählt, dass gilt

$$\bar{x}(a_1) - \bar{x}(a_2) = x'(a) - x(a), \quad \forall a \in A.$$

Wir weisen zunächst einige Eigenschaften von  $\bar{x} \in \mathbb{R}^{\bar{A}}$  nach.

**Behauptung 1.**  $\bar{x}$  ist ein zulässiger  $(s, t)$ -Fluss in  $N$  mit Wert 0 und  $\bar{w}^T \bar{x} < 0$ .  $\triangle$

**Beweis.** Wir zeigen zunächst, dass  $\bar{x}$  negative Kosten hat. Für alle  $a \in A$  gilt offenbar

$$\bar{x}(a_1)\bar{w}(a_1) + \bar{x}(a_2)\bar{w}(a_2) = (x'(a) - x(a))w(a),$$

und daraus folgt:

$$\begin{aligned} \sum_{\bar{a} \in \bar{A}} \bar{w}(\bar{a})\bar{x}(\bar{a}) &= \sum_{a \in A} (\bar{w}(a_1)\bar{x}(a_1) + \bar{w}(a_2)\bar{x}(a_2)) \\ &= \sum_{a \in A} (x'(a) - x(a))w(a) \\ &= w^T x' - w^T x < 0. \end{aligned}$$

Der Vektor  $\bar{x}$  erfüllt nach Definition die Kapazitätsbedingungen  $0 \leq \bar{x}(\bar{a}) \leq \bar{c}(\bar{a})$ . Wir

zeigen nun, dass  $\bar{x}$  auch die Flusserhaltungsbedingungen für alle  $v \in V$  erfüllt.

$$\begin{aligned}
\bar{x}(\delta_N^+(v)) - \bar{x}(\delta_N^-(v)) &= \sum_{a \in \delta^+(v)} (\bar{x}(a_1) - \bar{x}(a_2)) - \sum_{a \in \delta^-(v)} (\bar{x}(a_1) - \bar{x}(a_2)) \\
&= \sum_{a \in \delta^+(v)} (x'(a) - x(a)) - \sum_{a \in \delta^-(v)} (x'(a) - x(a)) \\
&= \left( \sum_{a \in \delta^+(v)} x'(a) - \sum_{a \in \delta^-(v)} x'(a) \right) \\
&\quad - \left( \sum_{a \in \delta^+(v)} x(a) - \sum_{a \in \delta^-(v)} x(a) \right) \\
&= \begin{cases} 0 - 0 & \text{falls } v \in V \setminus \{s, t\} \\ \text{val}(x') - \text{val}(x) & \text{falls } v = s \\ -\text{val}(x') + \text{val}(x) & \text{falls } v = t \end{cases} \\
&= 0.
\end{aligned}$$

Daraus folgt, dass  $\bar{x}$  die Flusserhaltungsbedingungen erfüllt und dass  $\text{val}(\bar{x}) = 0$  gilt. Damit ist Behauptung 1 bewiesen.  $\square$

**Behauptung 2.** Zerlegung von  $x$  in Kreise. Genauer: Es gibt zulässige  $(s, t)$ -Flüsse  $x_1, \dots, x_k \in \mathbb{R}^{\bar{A}}$ ,  $k \leq |\bar{A}|$ , so dass folgendes gilt:

- (a)  $\bar{x}(\bar{a}) = \sum_{i=1}^k x_i(\bar{a})$  für alle  $\bar{a} \in \bar{A}$ .
- (b) Für jeden  $(s, t)$ -Fluss  $x_i$ ,  $i \in \{1, \dots, k\}$  gibt es einen gerichteten Kreis  $\bar{C}_i$  in  $N$  und eine positive Zahl  $\alpha_i$ , so dass  $x_i(\bar{a}) = \alpha_i$  für alle  $\bar{a} \in \bar{C}_i$  und  $x_i(\bar{a}) = 0$  für alle  $\bar{a} \in \bar{A} \setminus \bar{C}_i$ .  $\triangle$

**Beweis.** Sei  $p$  die Anzahl der Bögen  $\bar{a} \in \bar{A}$  mit  $\bar{x}(\bar{a}) \neq 0$ . Da  $x \neq x'$  gilt  $p \geq 1$ . Sei  $v_0$  ein Knoten, so dass ein Bogen  $(v_0, v_1) \in \bar{A}$  existiert mit  $\bar{x}((v_0, v_1)) \neq 0$ . Da in  $v_1$  die Flusserhaltungsbedingung gilt, muss es einen Bogen  $(v_1, v_2) \in \bar{A}$  geben mit  $\bar{x}((v_1, v_2)) \neq 0$ . Fahren wir so weiter fort, so erhalten wir einen gerichteten Weg  $v_0, v_1, v_2, \dots$ . Da  $N$  endlich ist, muss irgendwann ein Knoten auftreten, der schon im bisher konstruierten Weg enthalten ist. Damit haben wir einen gerichteten Kreis  $\bar{C}_p$  gefunden. Sei  $\alpha_p$  der kleinste Wert  $\bar{x}(\bar{a})$  der unter den Bögen  $\bar{a}$  des Kreises  $\bar{C}_p$  auftritt. Definieren wir

$$x_p(\bar{a}) := \begin{cases} \alpha_p & \text{für alle } \bar{a} \in \bar{C}_p, \\ 0 & \text{sonst,} \end{cases}$$

so ist  $x_p \in \mathbb{R}^{\bar{A}}$  ein  $(s, t)$ -Fluss mit Wert 0. Setzen wir nun  $\bar{x}_p := \bar{x} - x_p$ , so ist  $\bar{x}_p$  ein  $(s, t)$ -Fluss mit Wert 0, und die Zahl der Bögen  $\bar{a} \in \bar{A}$  mit  $\bar{x}_p(\bar{a}) \neq 0$  ist kleiner als  $p$ . Führen wir diese Konstruktion iterativ fort bis  $\bar{x}_p = 0$  ist, so haben wir die gesuchten Kreise gefunden.  $\square$

## 7 Flüsse mit minimalen Kosten

Damit können wir den Beweis von (7.6) beenden. Für die nach Behauptung 2 existierenden  $(s, t)$ -Flüsse  $x_i$  gilt offenbar

$$\bar{w}^T \bar{x} = \sum_{i=1}^k \bar{w}^T x_i.$$

Da  $\bar{w}^T \bar{x} < 0$  nach Behauptung 1 ist, muss einer der Werte  $\bar{w}^T x_i$  kleiner als Null sein, dass heißt, wir haben in  $N$  einen gerichteten Kreis mit negativen Kosten gefunden.  $\square$

Satz (7.2) bzw. Satz (7.6) sind Optimalitätskriterien für zulässige  $(s, t)$ -Flüsse. Man kann beide Aussagen – was algorithmisch noch wichtiger ist – benutzen, um zu zeigen, dass Kostenminimalität erhalten bleibt, wenn man entlang Wegen minimaler Kosten augmentiert.

**(7.7) Satz.** Sei  $D = (V, A)$  ein Digraph mit gegebenen Knoten  $s, t \in V$ , Kapazitäten  $c \in \mathbb{R}_+^A$  und Kosten  $w \in \mathbb{R}^A$ . Sei  $x$  ein zulässiger  $(s, t)$ -Fluss in  $D$  mit Wert  $f$ , der kostenminimal unter allen  $(s, t)$ -Flüssen mit Wert  $f$  ist, und sei  $N = (V, \bar{A}, \bar{c}, \bar{w})$  das zugehörige augmentierende Netzwerk. Sei  $P$  ein  $(s, t)$ -Weg in  $N$  mit minimalen Kosten  $\bar{w}(P)$ , und sei  $\bar{x}$  ein zulässiger  $(s, t)$ -Fluss in  $N$ , so dass  $\bar{x}(\bar{a}) > 0$  für alle  $\bar{a} \in P$  und  $\bar{x}(\bar{a}) = 0$  für alle  $\bar{a} \in \bar{A} \setminus P$ , dann ist der Vektor  $x' \in \mathbb{R}^A$ , definiert durch

$$x'(a) := x(a) + \bar{x}(a_1) - \bar{x}(a_2) \text{ für alle } a \in A,$$

ein zulässiger  $(s, t)$ -Fluss in  $D$  mit Wert  $f + \text{val}(\bar{x})$ , der kostenminimal unter allen Flüssen dieses Wertes in  $D$  ist.  $\triangle$

**Beweis.** Trivialerweise ist  $x' \in \mathbb{R}^A$  ein zulässiger  $(s, t)$ -Fluss mit Wert  $f + \text{val}(\bar{x})$ . Wir zeigen, dass  $x'$  kostenminimal ist. Angenommen, dies ist nicht der Fall, dann gibt es nach Satz (7.6) einen negativen gerichteten Kreis  $C'$  im bezüglich  $x'$  augmentierenden Netzwerk  $N' = (V, A', c', w')$ . Wir beweisen, dass dann auch ein negativer gerichteter Kreis in  $N$  bezüglich  $x$  existiert.

Wir bemerken zunächst, dass das augmentierende Netzwerk  $N'$  aus  $N$  dadurch hervorgeht, dass die Bögen aus  $P \subseteq \bar{A}$  neue Kapazitäten erhalten und möglicherweise ihre Richtung und damit ihre Kosten ändern. Sei  $\bar{B} \subseteq P$  die Menge der Bögen aus  $\bar{A}$ , die in  $N'$  eine andere Richtung als in  $N$  haben, und sei  $B' := \{(v, u) \in A' \mid (u, v) \in \bar{B}\}$ .

Wir untersuchen nun den gerichteten Kreis  $C' \subseteq A'$  in  $N'$  mit negativen Kosten. Gilt  $C' \cap B' = \emptyset$ , so ist  $C'$  in  $\bar{A}$  enthalten und somit ein negativer Kreis in  $N$ . Dann wäre  $x$  nach Satz (7.6) nicht kostenoptimal, was unserer Voraussetzung widerspricht. Wir können daher annehmen, dass  $C' \cap B' \neq \emptyset$  gilt.

Der Beweis verläuft nun wie folgt. Wir konstruieren aus dem  $(s, t)$ -Weg  $P$  und dem gerichteten Kreis  $C'$  einen  $(s, t)$ -Weg  $Q \subseteq \bar{A}$  und einen gerichteten Kreis  $K' \subseteq A'$  mit den Eigenschaften

$$\begin{aligned} \bar{w}(P) + w'(C') &\geq \bar{w}(Q) + w'(K'), \\ |C' \cap B'| &> |K' \cap B'|. \end{aligned}$$

Durch iterative Wiederholung dieser Konstruktion erhalten wir nach höchstens  $|B'|$  Schritten einen  $(s, t)$ -Weg in  $N$  und einen gerichteten Kreis in  $N'$ , dessen Kosten negativ sind und der keinen Bogen aus  $B'$  enthält. Folglich ist dieser Kreis ein negativer Kreis in  $N$ . Widerspruch!

Die Konstruktion verläuft wie folgt. Da  $C' \cap B' \neq \emptyset$ , gibt es einen Bogen  $(u, v) \in P$  mit  $(v, u) \in C'$ . Wir wählen denjenigen Bogen  $(u, v) \in P$ , der auf dem Weg von  $s$  nach  $t$  entlang  $P$  der erste Bogen mit  $(v, u) \in C'$  ist. Wir unterscheiden zwei Fälle.

**Fall 1:**  $C' \cap B'$  enthält mindestens zwei Bögen.

Sei  $(y, x)$  der nächste Bogen auf dem gerichteten Kreis  $C'$  nach  $(v, u)$ , der in  $B'$  ist. Wir konstruieren einen  $(s, t)$ -Pfad  $\bar{P} \subseteq \bar{A}$  wie folgt. Wir gehen von  $s$  aus entlang  $P$  nach  $u$ , dann von  $u$  entlang  $C'$  nach  $y$  und von  $y$  entlang  $P$  nach  $t$ . Starten wir nun in  $v$ , gehen entlang  $P$  zu  $x$  und dann entlang  $C'$  nach  $v$ , so erhalten wir eine geschlossene gerichtete Kette  $P' \subseteq A'$ . Aus  $\bar{w}_{uv} = -w'_{vu}$  und  $\bar{w}_{xy} = -w'_{yx}$  folgt, dass  $\bar{w}(P) + w'(C') = \bar{w}(\bar{P}) + w'(P')$  gilt.  $\bar{P}$  ist die Vereinigung von gerichteten Kreisen  $C'_1, \dots, C'_k \subseteq \bar{A}$  mit einem gerichteten  $(s, t)$ -Weg  $Q \subseteq \bar{A}$ , und  $P'$  ist die Vereinigung von gerichteten Kreisen  $C'_{k+1}, \dots, C'_r \subseteq A'$ . Da  $P$  kostenminimal in  $N$  ist, gilt  $\bar{w}(P) \leq \bar{w}(Q)$ , und somit gibt es wegen  $\bar{w}(\bar{P}) + w'(P') = \bar{w}(Q) + \sum_{i=1}^k \bar{w}(C'_i) + \sum_{i=k+1}^r w'(C'_i)$  mindestens einen gerichteten Kreis in  $A'$ , sagen wir  $K'$ , der negative Kosten hat. Nach Konstruktion gilt  $|K' \cap B'| < |C' \cap B'|$ .

**Fall 2:** Der Bogen  $(u, v)$  ist der einzige Bogen auf  $P$  mit  $(v, u) \in C' \cap B'$ .

Wir konstruieren einen gerichteten  $(s, t)$ -Pfad  $\bar{P} \subseteq \bar{A}$  wie folgt. Wir starten in  $s$  und folgen  $P$  bis  $u$ , dann folgen wir dem gerichteten Weg von  $u$  entlang  $C'$  bis  $v$  und dann wieder dem gerichteten Weg von  $v$  entlang  $P$  bis  $t$ . Offenbar ist  $\bar{P}$  in  $\bar{A}$  enthalten und ein gerichteter  $(s, t)$ -Pfad in  $N$ . Aus  $\bar{w}_{uv} = -w'_{vu}$  folgt direkt  $\bar{w}(P) + w'(C') = \bar{w}(\bar{P})$ . Der gerichtete  $(s, t)$ -Pfad  $\bar{P}$  ist die Vereinigung eines  $(s, t)$ -Weges  $Q$  und einiger gerichteter Kreise  $C'_1, \dots, C'_k$ . Da  $P$  ein  $(s, t)$ -Weg in  $N$  mit minimalen Kosten ist, gilt  $\bar{w}(Q) \geq \bar{w}(P)$ , und aus  $\bar{w}(Q) = \bar{w}(\bar{P}) - \sum_{i=1}^k \bar{w}(C'_i)$  folgt, dass mindestens einer der Kreise  $C'_i$  negativ ist. Da alle  $C'_i$  in  $\bar{A}$  enthalten sind, enthält  $\bar{A}$  einen negativen Kreis. Widerspruch!  $\square$

Damit können wir nun einen Algorithmus zur Lösung des Minimalkosten-Flussproblems angeben.

### (7.8) Algorithmus.

**Eingabe:** Digraph  $D = (V, A)$ , mit Kapazitäten  $c \in \mathbb{R}_+^A$  und Kosten  $w \in \mathbb{R}^A$ , zwei verschiedene Knoten  $s, t \in V$  und ein Flusswert  $f$ .

**Ausgabe:** Ein zulässiger  $(s, t)$ -Fluss  $x$  mit Wert  $f$ , der kostenminimal unter allen zulässigen  $(s, t)$ -Flüssen mit Wert  $f$  ist, oder die Aussage, dass kein zulässiger  $(s, t)$ -Fluss mit Wert  $f$  existiert.

1. Setze  $x(a) = 0$  für alle  $a \in A$  (bzw. starte mit einem zulässigen  $(s, t)$ -Fluss mit Wert nicht größer als  $f$ ).

## 7 Flüsse mit minimalen Kosten

2. Konstruiere das augmentierende Netzwerk  $N = (V, \bar{A}, \bar{c}, \bar{w})$  bezüglich  $D$  und  $x$ .
3. Wende einen Kürzeste-Wege-Algorithmus (z. B. den Floyd-Algorithmus (5.25)) an, um im Digraphen  $N = (V, \bar{A})$  mit den „Bogenlängen“  $\bar{w}(\bar{a})$ ,  $\bar{a} \in \bar{A}$ , einen negativen gerichteten Kreis  $C$  zu finden. Gibt es keinen, dann gehe zu 5.
4. (Augmentierung entlang  $C$ )  
Bestimme  $\varepsilon := \min\{\bar{c}(\bar{a}) \mid \bar{a} \in C\}$ , setze für  $a \in A$

$$x(a) := \begin{cases} x(a) + \varepsilon & \text{falls } a_1 \in C \\ x(a) - \varepsilon & \text{falls } a_2 \in C \\ x(a) & \text{andernfalls} \end{cases}$$

und gehe zu 2. (Hier erhalten wir einen Fluss mit gleichem Wert und geringeren Kosten.)

5. Ist  $\text{val}(x) = f$ , STOP, gib  $x$  aus.
6. Bestimme mit einem Kürzeste-Wege-Algorithmus (z. B. einer der Varianten des Moore-Bellman-Verfahrens (5.20), es gibt keine negativen Kreise!) einen  $(s, t)$ -Weg  $P$  in  $N$  mit minimalen Kosten  $\bar{w}(P)$ .
7. Gibt es in  $N$  keinen  $(s, t)$ -Weg, dann gibt es in  $D$  keinen zulässigen  $(s, t)$ -Fluss mit Wert  $f$ , STOP.
8. (Augmentierung entlang  $P$ )  
Bestimme  $\varepsilon' := \min\{\bar{c}(\bar{a}) \mid \bar{a} \in P\}$ ,  $\varepsilon := \min\{\varepsilon', f - \text{val}(x)\}$ , setze für  $a \in A$

$$x(a) := \begin{cases} x(a) + \varepsilon & \text{falls } a_1 \in P \\ x(a) - \varepsilon & \text{falls } a_2 \in P \\ x(a) & \text{andernfalls,} \end{cases}$$

konstruiere das bzgl.  $x$  und  $D$  augmentierende Netzwerk  $N = (V, \bar{A}, \bar{c}, \bar{w})$  und gehe zu 5. △

Die Korrektheit des Algorithmus folgt unmittelbar aus (7.6) und (7.7). Wir wollen nun die Laufzeit abschätzen. Hat  $D$  nur nichtnegative Kosten  $w(a)$  bzw. enthält  $D$  keinen augmentierenden Kreis mit negativen Kosten, so ist der Nullfluss ein kostenoptimaler Fluss mit Wert Null, und die Schleife über die Schritte 2, 3 und 4 braucht nicht durchlaufen zu werden. Sind alle Kapazitäten ganzzahlig, so wird der Flusswert in Schritt 8 um jeweils mindestens eine Einheit erhöht. Also sind höchstens  $f$  Aufrufe eines Kürzesten-Wege-Algorithmus erforderlich.

**(7.9) Satz.** *Ist  $D = (V, A)$  ein Digraph mit ganzzahligen Kapazitäten  $c(a)$  und nichtnegativen Kosten  $w(a)$ , und sind  $s, t$  zwei verschiedene Knoten und  $f \in \mathbb{Z}_+$  ein vorgegebener Flusswert, so findet Algorithmus (7.8) in  $\mathcal{O}(f|V|^3)$  Schritten einen kostenminimalen zulässigen  $(s, t)$ -Fluss mit Wert  $f$ , falls ein solcher existiert. △*

Der Algorithmus ist in dieser Form nicht polynomial, da seine Laufzeit polynomial in der Kodierungslänge  $\langle f \rangle$  sein müsste. Ferner ist nicht unmittelbar klar, wie lange er läuft, wenn negative Kosten erlaubt sind, da die Anzahl der Kreise mit negativen Kosten, auf denen der Fluss verändert werden muß, nicht ohne weiteres abgeschätzt werden kann. Diese Schwierigkeiten können durch neue Ideen (Augmentierung entlang Kreisen mit minimalen durchschnittlichen Kosten  $\bar{w}(C)/|C|$ , Skalierungstechniken) überwunden werden, so dass Versionen von Algorithmus (7.8) existieren, die polynomiale Laufzeit haben. Aus Zeitgründen können diese Techniken hier nicht dargestellt werden. Es sei hierzu wiederum auf die schon mehrfach erwähnten Übersichtsartikel und das Buch von Ahuja et al. (1993) verwiesen, die auch ausführlich auf die historische Entwicklung eingehen. Der Aufsatz von M. Shigeno and McCormick (2000) präsentiert zwei Skalierungsmethoden und gibt dabei eine gute Vergleichsübersicht über viele der bekannten Min-Cost-Flow-Algorithmen.

## 7.2 Transshipment-, Transport- u. Zuordnungsprobleme

Wie beim Maximalflussproblem ist es natürlich möglich, Minimalkosten-Flussprobleme, bei denen mehrere Quellen und Senken vorkommen und bei denen von Null verschiedene untere Kapazitätsschranken für die Bögen auftreten, zu lösen. Sie können auf einfache Weise auf das Standardproblem (7.1) reduziert werden.

Es gibt noch eine Reihe von Varianten des Minimalkosten-Flussproblems, die in der Literatur große Beachtung gefunden und viele Anwendungen haben. Ferner gibt es für alle dieser Probleme Spezialverfahren zu ihrer Lösung. Aus Zeitgründen können wir diese nicht behandeln. Wir wollen diese Probleme jedoch zumindest aus „Bildungsgründen“ erwähnen und zeigen, wie sie in Minimalkosten-Flussprobleme transformiert werden können.

**(7.10) Transshipment-Probleme (Umladeprobleme).** Gegeben sei ein Digraph  $D = (V, A)$ , dessen Knotenmenge zerlegt sei in drei disjunkte Teilmengen  $V_a$ ,  $V_n$  und  $V_u$ . Die Knoten aus  $V_a$  bezeichnen wir als *Angebotsknoten*. (Bei ihnen fließt ein Strom in das Netzwerk ein.) Die Knoten  $V_n$  bezeichnen wir als *Nachfrageknoten* (bei ihnen verlässt der Strom das Netz), und die Knoten  $V_u$  werden als *Umladeknoten* bezeichnet (hier wird der Fluss erhalten). Jedem Bogen  $a \in A$  sind eine Kapazität  $c(a)$  und ein Kostenkoeffizient  $w(a)$  zugeordnet. Ferner sei bei jedem Angebotsknoten  $v \in V_a$  die Menge  $a(v)$  verfügbar, und bei jedem Nachfrageknoten die Menge  $b(v)$  erwünscht. Die Aufgabe, einen Plan zu ermitteln, der Auskunft darüber gibt, von welchen Anbietern aus über welche Transportwege der Bedarf der Nachfrager zu decken ist, damit die Kosten für alle durchzuführenden Transporte minimiert werden, heißt *Umladeproblem*.  $\triangle$

Offenbar kann man ein Umladeproblem wie in (7.10) angegeben als lineares Programm

schreiben.

$$\begin{aligned} \min \sum_{a \in A} w(a)x(a) \\ x(\delta^-(v)) - x(\delta^+(v)) &= \begin{cases} 0 & \text{falls } v \in V_u \\ b(v) & \text{falls } v \in V_n \\ -a(v) & \text{falls } v \in V_a \end{cases} \quad (7.11) \\ 0 \leq x(a) \leq c(a) & \quad \forall a \in A. \end{aligned}$$

Problem (7.10) ist offensichtlich höchstens dann lösbar, wenn  $\sum_{v \in V_n} b(v) = \sum_{v \in V_a} a(v)$  gilt. Das lineare Programm (7.11) kann in ein Minimalkosten-Flussproblem (7.1) wie folgt transformiert werden. Wir führen eine (künstliche) Quelle  $s$  und eine künstliche Senke  $t$  ein. Die Quelle  $s$  verbinden wir mit jedem Knoten  $v \in V_a$  durch einen Bogen  $(s, v)$  mit Kosten Null und Kapazität  $a(v)$ . Jeden Knoten  $v \in V_n$  verbinden wir mit der Senke  $t$  durch einen Bogen  $(v, t)$  mit Kosten Null und Kapazität  $b(v)$ . Offenbar liefert der kostenminimale  $(s, t)$ -Fluss in diesem neuen Digraphen mit Wert  $\sum_{v \in V_n} b(v)$  eine optimale Lösung des Umladeproblems.

**(7.12) Transportprobleme.** Ein Transshipment-Problem, bei dem  $V_u = \emptyset$  gilt, heißt Transportproblem. Hier wird also von Erzeugern direkt zu den Kunden geliefert, ohne den Zwischenhandel einzuschalten.  $\triangle$

**(7.13) Zuordnungsproblem.** Ein Transportproblem, bei dem  $|V_a| = |V_n|$ ,  $b(v) = 1$  für alle  $v \in V_n$  und  $a(v) = 1 \forall v \in V_a$  gilt, heißt Zuordnungsproblem (vergleiche (3.9)).  $\triangle$

Zur Lösung von Zuordnungs- und Transportproblemen gibt es besonders schnelle Algorithmen, die erheblich effizienter als die Algorithmen zur Bestimmung kostenminimaler Flüsse sind. Näheres hierzu wird in den Übungen behandelt. Aber auch in diesem Bereich kann man nicht – wie bei Algorithmen zur Bestimmung von Flüssen mit Minimalkosten – davon sprechen, dass irgendein Verfahren das schnellste (in der Praxis) ist. Immer wieder gibt es neue Implementationstechniken, die bislang unterlegene Algorithmen erheblich beschleunigen und anderen Verfahren überlegen erscheinen lassen. Siehe hierzu (Ahuja et al., 1993).

Wir behandeln hier das Zuordnungsproblem (englisch: assignment problem) stiefmütterlich als Spezialfall des Minimalkosten-Flussproblems. Seine Bedeutung für die Entwicklung der Algorithmen zur Lösung kombinatorischer Optimierungsprobleme ist jedoch erheblich. Im Jahr 1955 veröffentlichte Harold Kuhn (Kuhn, 1955) einen Algorithmus, den er, um die Bedeutung von Ideen zweier ungarischer Mathematiker (D. König und J. Egerváry) hervorzuheben, „ungarische Methode“ nannte. Dieser Algorithmus ist ein Vorläufer der heute so genannten „Primal-Dual-Verfahren“ und erwies sich (nach einer Modifikation von Munkres) als polynomialer Algorithmus zur Lösung von Zuordnungsproblemen. Der ungarische Algorithmus war Vorbild für viele andere Methoden zur Lösung kombinatorischer Optimierungsprobleme, und er wurde so oft zitiert, dass der Aufsatz (Kuhn, 1955) von der Zeitschrift *Naval Logistics Quarterly* im Jahre 2004 als wichtigstes Paper gewählt wurde, das seit Bestehen der Zeitschrift in dieser erschienen ist.



Vollkommen überraschend war die Entdeckung eines französischen Mathematikers im Jahre 2006, dass die ungarische Methode bereits um 1850 von Carl Gustav Jacob Jacobi (1804 in Potsdam – 1851 in Berlin) beschrieben wurde. Siehe hierzu Abschnitt 2 in (Grötschel, 2008) sowie Kuhn (2012) und die URLs:

<http://www.lix.polytechnique.fr/~ollivier/JACOBI/jacobiEngl.htm>

[http://en.wikipedia.org/wiki/Hungarian\\_method](http://en.wikipedia.org/wiki/Hungarian_method)

[http://www.zib.de/groetschel/pubnew/paper/groetschel2008\\_pp.pdf](http://www.zib.de/groetschel/pubnew/paper/groetschel2008_pp.pdf)

## 7.3 Der Netzwerk-Simplex-Algorithmus

Wie schon in der Einführung erwähnt kann der Simplexalgorithmus zur Lösung allgemeiner linearer Optimierungsprobleme zur Lösung des Minimalkosten-Flussproblems spezialisiert werden. Diese *Netzwerk-Simplex-Algorithmen* nutzen die spezielle Struktur des Minimalkosten-Flussproblems aus, um die Operationen des Simplexalgorithmus besonders effizient umzusetzen. Die Grundstruktur eines Netzwerk-Simplex-Algorithmus kann jedoch auch ohne Kenntnis des Simplexalgorithmus verstanden werden. Betrachten wir dazu das Minimalkosten-Flussproblem in der allgemeinen Form

$$\min c^T x$$

$$x(\delta^-(i)) - x(\delta^+(i)) = b_i \quad \forall i \in V, \quad (7.14)$$

$$l_a \leq x_a \leq u_a \quad \forall a \in A, \quad (7.15)$$

für einen Digraphen  $D = (V = \{1, \dots, n\}, A)$  mit  $n$  Knoten und  $m$  Bögen. Die Bogenbewertung  $c : A \rightarrow \mathbb{R}$  definiert „Kosten“, während die Bogenbewertungen  $l, u : A \rightarrow \mathbb{R}_+$  untere bzw. obere Schranken für den Fluss auf einem Bogen angeben und damit die Kapazität des Bogens definieren. Die Knotenbewertung  $b : V \rightarrow \mathbb{R}$  beschreibt die Einspeisung ( $b_i \leq 0, i \in V$ ) bzw. Ausspeisung ( $b_i \geq 0, i \in V$ ) an den Knoten; Ein- und Ausspeisung werden zusammen auch als *Bedarfe* bezeichnet. Wir nehmen in diesem Abschnitt an, dass  $D$  zusammenhängend ist und die Beziehung  $\sum_{i \in V} b_i = 0$  gilt. In Analogie zur Terminologie für  $(s, t)$ -Flüsse bezeichnen wir einen Vektor  $x \in \mathbb{R}^A$ , der (7.14) erfüllt, als *Fluss*. Erfüllt  $x$  zusätzlich (7.15), so sprechen wir von einem *zulässigen Fluss*.

(Optimale) zulässige Flüsse des Minimalkosten-Flussproblems lassen sich kombinatorisch über aufspannende Bäume beschreiben, die den Fluss definieren.

**(7.16) Definition (Baum-Lösung).** Ein Flussvektor  $x$  zusammen mit einem aufspannenden Baum  $T$  für  $D = (V, A)$  und Bogenmengen  $L, U \subseteq A$  heißt Baum-Lösung, wenn gilt:

- $T, L$  und  $U$  sind eine Partition von  $A$ ,
- $x_a = l_a$  für alle  $a \in L$ ,
- $x_a = u_a$  für alle  $a \in U$ .

Erfüllt  $(x, T, L, U)$  außerdem alle Bedingungen  $l \leq x \leq u$  (d. h. auch für die Bögen in  $T$ ), so heißt  $(x, T, L, U)$  zulässige Baum-Lösung.  $\triangle$

## 7 Flüsse mit minimalen Kosten

Wie wir später sehen werden (Satz (7.24)), genügt es, sich bei der Suche nach kostenminimalen Flüssen auf Baum-Lösungen zu beschränken. Baum-Lösungen haben die Eigenschaft, dass sich die Flusswerte auf den Baum-Kanten über die Flusserhaltungsbedingungen aus den Flusswerten der Nicht-Baum-Kanten eindeutig ergeben.

**(7.17) Lemma.** *Seien  $(T, L, U)$  eine Partition von  $A$  und  $T$  ein aufspannender Baum für  $D = (V, A)$ . Dann existiert ein (nicht unbedingt zulässiger) eindeutiger Fluss  $x$  mit  $x_a = l_a$  für alle  $a \in L$  und  $x_a = u_a$  für alle  $a \in U$ .  $\triangle$*

**Beweis.** Sei  $i$  ein Blatt von  $T$  und  $a \in T$  der mit  $i$  inzidente Bogen sowie  $M := L \cup U$  die Menge der Bögen mit festem Flusswert. Da  $i$  ein Blatt ist, sind alle anderen mit  $i$  inzidenten Bögen in  $M$  enthalten. Auf der linken Seite der Flusserhaltungsbedingung  $x(\delta^-(i)) - x(\delta^+(i)) = b_i$  sind daher alle Terme außer  $x_a$  konstant, d. h. der Wert von  $x_a$  ist durch diese Gleichung eindeutig bestimmt. Wenn wir nun  $x_a$  auf diesen Wert festsetzen und  $a$  aus  $T$  entfernen und in die Menge  $M$  aufnehmen, können wir dieses Verfahren mit dem übrigen Baum  $T$  und der aktualisierten Menge  $M$  fortsetzen. Schließlich ergibt sich ein (nicht notwendig zulässiger) Fluss für jeden Bogen des ursprünglichen Baumes  $T$ .  $\square$

Die Grundidee des Netzwerk-Simplex-Algorithmus ist, eine bekannte zulässige Baum-Lösung in eine andere, kostengünstigere Baum-Lösung „umzubauen“. Konkret passiert dies, indem wir einen Bogen  $e \in A \setminus T$  (engl. „entering arc“) zu  $T$  hinzunehmen und einen Bogen  $f \in T$  (engl. „leaving arc“) aus  $T$  entfernen. Damit die Baum-Eigenschaft erhalten bleibt, sollte  $f$  natürlich auf dem eindeutigen Kreis  $C$  in  $T+e$  liegen. Wir wählen die Orientierung von  $C$  so, dass  $e$  ein Vorwärtsbogen auf  $C$  ist;  $F$  und  $B$  bezeichnen die Mengen der Vorwärts- bzw. Rückwärtsbögen in  $C$ . Damit die Aufnahme von  $e$  den Zielfunktionswert verbessert, müssen wir den Fluss auf  $e$  erhöhen ( $e \in L$ ) oder verringern ( $e \in U$ ). Wenn wir den Fluss auf  $e$  um  $\varepsilon$  verändern, erhalten wir einen neuen Fluss  $x'$ , indem wir alle Flüsse auf Bögen von  $C$  gemäß Gleichung (7.4) anpassen, sodass die Flusserhaltung weiter gewährleistet ist. Damit  $x'$  auch die Kapazitätsschranken einhält, dürfen wir  $\varepsilon$  nur so groß wählen, dass  $x'$  auf  $C$  eine Kapazitätsschranke erreicht, d. h.  $\varepsilon$  ist

$$\varepsilon := \min\{\min\{x_a - l_a \mid a \in B\}, \min\{u_a - x_a \mid a \in F\}\}. \quad (7.18)$$

Es ist durchaus möglich, dass  $\varepsilon = 0$  auftritt. Sei  $f$  einer der Bögen in  $C$ , der durch die Änderung des Flusses um  $\varepsilon$  seine Kapazitätsschranke erreicht (Achtung: hier kann auch  $e = f$  gelten). Die Baum-Lösung für den neuen Fluss  $x'$  ergibt sich als

$$\begin{aligned} T &:= (T - f) + e, \\ L &:= \begin{cases} (L - e) + f & \text{falls } x'_f = l_f, \\ (L - e) & \text{sonst,} \end{cases} \\ U &:= \begin{cases} (U - e) + f & \text{falls } x'_f = u_f, \\ (U - e) & \text{sonst.} \end{cases} \end{aligned} \quad (7.19)$$

Natürlich sollte  $e$  so gewählt werden, dass sich die Kosten verringern. Die Kosten von  $x$  und  $x'$  unterscheiden sich um

$$\varepsilon \left( \sum_{a \in F} c_a - \sum_{a \in B} c_a \right), \quad (7.20)$$

d. h. die Kosten des Kreises  $C$  (die Differenz der Summen in den Klammern) sollten negativ sein. Die Berechnung der Kosten von  $C$  kann durch die Einführung von *Knotenpreisen*  $y = (y_i)_{i \in V}$  vereinfacht werden. Die Idee (und ökonomische Interpretation) der Knotenpreise ist, dass  $y_i$  den Wert des transportierten Gutes (das hier abstrakt als Fluss behandelt wird) an Knoten  $i \in V$  angibt. Diese Knotenpreise können lokal unterschiedlich sein, da die Transportkosten zu einem Knoten (die indirekt auch durch die Bogenkapazitäten bestimmt werden) sich von Knoten zu Knoten unterscheiden können. Sie sind allerdings nicht völlig unabhängig voneinander: Wenn das Gut zum Preis  $y_i$  an Knoten  $i$  bereitgestellt werden kann und für den Transport zu Knoten  $j$  die Kosten  $c_{ij}$  anfallen, so sollte der Preis an Knoten  $j$  *genau*  $y_i + c_{ij}$  betragen. Ist der Preis geringer, ist er nicht kostendeckend, ist er höher, würde jemand anderes an Knoten  $i$  kaufen und den Transport selbst durchführen. Wir verlangen daher von einem Knotenpreisvektor  $y$ , dass er das Gleichungssystem

$$y_i + c_{ij} = y_j \quad \forall (i, j) \in T \quad (7.21)$$

erfüllt. Dieses Gleichungssystem hat  $n$  Variablen, jedoch nur  $(n - 1)$  Gleichungen, die Lösung ist also nicht eindeutig. Allerdings ist die *Differenz*  $y_i - y_j$  für  $(i, j) \in T$  eindeutig und mit der Setzung  $y_n = 0$  ergibt sich eine eindeutige Lösung für jeden aufspannenden Baum  $T$ . Wir definieren die *reduzierten Kosten*  $\bar{c}_{ij}$  eines Bogens  $(i, j) \in A$  als

$$\bar{c}_{ij} := c_{ij} + y_i - y_j \quad \forall (i, j) \in A. \quad (7.22)$$

Offensichtlich gilt  $\bar{c}_{ij} = 0$  für  $(i, j) \in T$ . Es zeigt sich, dass die reduzierten Kosten von  $e$  genau den Kosten des von  $e$  induzierten Kreises  $C$  entsprechen.

**(7.23) Lemma.** *Seien  $e \in A \setminus T$ ,  $C$  der eindeutige Kreis in  $T + e$  sowie  $F$  und  $B$  die Menge der Vorwärts- und Rückwärtsbögen in  $C$  mit  $e \in F$ . Dann gilt*

$$\bar{c}_e = \sum_{a \in F} c_a - \sum_{a \in B} c_a. \quad \triangle$$

**Beweis.** Es gilt

$$\begin{aligned} \sum_{a \in F} c_a - \sum_{a \in B} c_a &= \sum_{(i,j) \in F} (\bar{c}_{ij} - y_i + y_j) - \sum_{(i,j) \in B} (\bar{c}_{ij} - y_i + y_j) \\ &= \bar{c}_e + \underbrace{\sum_{(i,j) \in F} (-y_i + y_j) - \sum_{(i,j) \in B} (-y_i + y_j)}_{=0} \end{aligned}$$

wegen  $\bar{c}_a = 0$  für  $a \in T$  und weil die beiden Summen zusammen jeden Knoten des Kreises  $C$  jeweils einmal mit positivem und negativem Vorzeichen enthalten.  $\square$

## 7 Flüsse mit minimalen Kosten

Die reduzierten Kosten lassen sich, die Knotenpreise  $y$  vorausgesetzt, leichter berechnen als die Kosten eines Kreises. Intuitiv geben die reduzierten Kosten eines Bogens an, wie sich die Kosten des Gesamtflusses ändern, wenn man den Fluss auf dem Bogen um eine Einheit erhöht. Sie liefern daher das folgende Optimalitätskriterium.

**(7.24) Satz.** *Eine zulässige Baum-Lösung  $(x, T, L, U)$  ist optimal, wenn ein Knotenpreisvektor  $y$  existiert mit*

$$\bar{c}(a) \begin{cases} = 0 & a \in T, \\ \geq 0 & a \in L, \\ \leq 0 & a \in U. \end{cases} \quad \triangle$$

**Beweis.** Seien  $x$  ein Fluss und  $y$  ein Knotenpreisvektor. Dann gilt

$$\begin{aligned} \bar{c}^T x &= \sum_{(i,j) \in A} x_{ij} (c_{ij} + y_i - y_j) \\ &= \sum_{(i,j) \in A} x_{ij} c_{ij} + \sum_{i \in V} \sum_{(i,j) \in A} x_{ij} y_i - \sum_{j \in V} \sum_{(i,j) \in A} x_{ij} y_j \\ &= c^T x + \sum_{i \in V} y_i \left( \sum_{(i,j) \in A} x_{ij} - \sum_{(j,i) \in A} x_{ji} \right) \\ &= c^T x - \sum_{i \in V} y_i b_i, \end{aligned}$$

d. h. die reduzierten Kosten  $\bar{c}^T x$  eines Flusses  $x$  unterscheiden sich nur um eine Konstante von den Kosten  $c^T x$ . Ist  $x$  optimal bzgl. der reduzierten Kosten, dann ist  $x$  also auch optimal bzgl. der ursprünglichen Kosten.

Sei nun  $(x, T, L, U)$  eine zulässige Baum-Lösung mit zugehörigem Knotenpreisvektor  $y$  und darüber definierten reduzierten Kosten  $\bar{c}$ , die die im Satz genannten Voraussetzungen erfüllen. Betrachten wir einen weiteren zulässigen Fluss  $x'$ . Es gilt dann:

$$\begin{aligned} \bar{c}^T x' &= \sum_{a \in T} x'_a \bar{c}_a + \sum_{a \in L} x'_a \bar{c}_a + \sum_{a \in U} x'_a \bar{c}_a \\ &= \sum_{a \in L} x'_a \bar{c}_a + \sum_{a \in U} x'_a \bar{c}_a \\ &\geq \sum_{a \in L} l_a \bar{c}_a + \sum_{a \in U} u_a \bar{c}_a \\ &\geq \sum_{a \in A} x_a \bar{c}_a = \bar{c}^T x. \end{aligned}$$

Also ist  $x$  optimal bezüglich  $\bar{c}$  und damit auch bezüglich  $c$ . □

**(7.25) Folgerung.** *Falls für ein Minimalkosten-Flussproblem ein zulässiger Fluss existiert, so existiert auch eine optimale zulässige Baum-Lösung.* △

**Beweis.** Folgt aus dem zweiten Teil des Beweises von Satz (7.24).  $\square$

Der Netzwerk-Simplex-Algorithmus nutzt Folgerung (7.25) aus, indem ausschließlich zulässige Baum-Lösungen erzeugt werden.

Noch offen ist die Konstruktion einer ersten zulässigen Baum-Lösung. Anders als beim Maximalflussproblem ist dies für das Minimalkosten-Flussproblems nicht trivial, da der Nullfluss nicht notwendig zulässig ist. Wir können jedoch aus jeder Instanz des Minimalkosten-Flussproblems  $\mathcal{I} = (D = (V, A), c, l, u)$  mit Bedarfen  $b$  eine abgeleitete Instanz  $\mathcal{I}' = (D' = (V', A'), c', l', u')$  mit Bedarfen  $b'$  konstruieren,

- für die wir einen zulässigen Fluss angeben können und
- deren optimale Lösungen im Falle der Zulässigkeit von  $\mathcal{I}$  denen von  $\mathcal{I}'$  entsprechen.

Dazu definieren wir den „Nettobedarf“  $\tilde{b}(i)$  für Knoten in  $V$  als

$$\tilde{b}_i := b_i + l(\delta^+(i)) - l(\delta^-(i)), \quad (7.26)$$

also als Bedarfsplus der Differenz der Mindestabgabe und der Mindestzulieferung. Ein Knoten  $i \in V$  ist „unterversorgt“, wenn sein Bedarf  $b_i$  nicht durch die eingehenden Bögen gedeckt werden kann, d. h. wenn  $\tilde{b}_i > 0$  und „überversorgt“ sonst. Wir setzen

$$\begin{aligned} V^+ &:= \{i \in V \mid \tilde{b}_i \leq 0\}, \\ V^- &:= \{i \in V \mid \tilde{b}_i > 0\}. \end{aligned}$$

Der Digraph  $D' = (V', A')$  enthält einen zusätzlichen Knoten  $k$  mit Bedarf  $b'_k = 0$  sowie einen Bogen  $(k, i)$  für jeden Knoten  $i \in V^-$  und einen Bogen  $(i, k)$  für jeden Knoten  $i \in V^+$ . Die Kosten für diese zusätzlichen Bögen werden so hoch gesetzt, dass sie in keiner Optimallösung für  $\mathcal{I}'$  gewählt werden, falls  $\mathcal{I}$  zulässig ist, also z. B.

$$M := 1 + \frac{1}{2}|V| \max\{|c_a| \mid a \in A\}.$$

Die erweiterte Instanz  $\mathcal{I}'$  ist dann gegeben durch:

$$\begin{aligned} V' &:= V \cup \{k\} \\ A' &:= A \cup \{(k, i) \mid i \in V^-\} \cup \{(i, k) \mid i \in V^+\} \\ c'_a &:= \begin{cases} c_a & a \in A \\ M & a \in A' \setminus A \end{cases} \\ b'_i &:= \begin{cases} b_i & i \in V \\ 0 & i \in V' \setminus V \end{cases} \\ l'_a &:= \begin{cases} l_a & a \in A \\ 0 & a \in A' \setminus A \end{cases} \\ u'_a &:= \begin{cases} u_a & a \in A \\ \infty & a \in A' \setminus A \end{cases} \end{aligned} \quad (7.27)$$

**(7.28) Satz.** Sei  $\mathcal{I}$  eine Instanz eines Minimalkosten-Flussproblems mit  $l_a < u_a$  für alle  $a \in A$  und  $\mathcal{I}'$  eine weitere, gemäß Gleichung (7.27) definierte Instanz. Dann gilt:

(a) Es existiert eine zulässige Lösung für  $\mathcal{I}'$ . Genauer: Es bilden

$$T := A' \setminus A, \quad L := A, \quad U := \emptyset$$

und der durch  $T$  definierte Fluss  $x$  eine zulässige Baum-Lösung von  $\mathcal{I}'$ .

(b) Gilt in einer Optimallösung  $x$  von  $\mathcal{I}'$   $x_a > 0$  für einen Bogen  $a \in A' \setminus A$ , so existiert kein zulässiger Fluss für  $\mathcal{I}$ .

(c) Ist  $x$  ein optimaler zulässiger Fluss für  $\mathcal{I}'$  mit  $x_a = 0$  für  $a \in A' \setminus A$ , so ist  $x$  eingeschränkt auf  $A$  ein optimaler zulässiger Fluss für  $\mathcal{I}$ .  $\triangle$

**Beweis.** Hausaufgabe.  $\square$

**(7.29) Algorithmus Netzwerk-Simplex-Algorithmus.**

**Eingabe:** Zusammenhängender Digraph  $D = (V, A)$ , mit Kapazitäten  $l, u : A \rightarrow \mathbb{R}_+^A$  mit  $l_a < u_a$  und Kosten  $w \in \mathbb{R}^A$  sowie Bedarfe an den Knoten  $b : V \rightarrow \mathbb{R}$  mit  $\sum_{i \in V} b_i = 0$ .

**Ausgabe:** Ein zulässiger Fluss  $x$ , der kostenminimal unter allen zulässigen Flüssen ist, oder die Aussage, dass kein zulässiger Fluss existiert.

1. *Initialisierung:* Erstelle  $\mathcal{I}'$  aus  $\mathcal{I}$ . Alle weiteren Schritte beziehen sich auf  $\mathcal{I}'$ . Setze

$$T := A' \setminus A, \quad L := A, \quad U := \emptyset$$

und bestimme den entsprechenden Fluss  $x$ .

2. *Berechnung Knotenpreise:* Bestimme neue Knotenpreise  $y$  durch Lösen des Gleichungssystems (7.21) und  $y_k = 0$ .

3. *Optimalitätstest:* Wenn es kein  $a \in L \cup U$  gibt mit

$$a \in L \text{ und } \bar{c}_a < 0 \quad \text{oder} \quad a \in U \text{ und } \bar{c}_a > 0, \quad (7.30)$$

dann STOP: Wenn es ein  $a \in A' \setminus A$  gibt mit  $x_a > 0$ , dann existiert kein zulässiger Fluss für  $\mathcal{I}$ . Sonst gib die Einschränkung von  $x$  auf  $A$  als optimalen Fluss aus.

4. *Pricing:* Wähle einen Bogen  $e \in L \cup U$ , der (7.30) erfüllt.

5. *Augmentierung:* Finde Kreis  $C$  in  $T + e$  und bestimme die Flussänderung  $\varepsilon$  gemäß (7.18). Aktualisiere den Fluss  $x$  gemäß (7.4) und bestimme einen Bogen  $f$ , der seine untere oder obere Kapazitätsschranke erreicht hat.

6. *Update:* Aktualisiere die Baum-Lösung gemäß (7.19) und gehe zu Schritt 2.  $\triangle$

Algorithmus (7.29) muss nicht terminieren, da es möglich ist, dass Bögen in  $T$  existieren, die bereits ihre untere oder obere Kapazitätsschranke erreicht haben. Es gilt dann  $\varepsilon = 0$  in Gleichung (7.18) und durch den Tausch der Bögen  $e$  und  $f$  ändert sich zwar die Baumlösung, jedoch weder der Fluss  $x$  noch der Zielfunktionswert. Ohne entsprechende Maßnahmen kann es tatsächlich zu *Cycling* kommen, d. h. nach einer endlichen Zahl solcher Schritte werden dieselben Bögen erneut getauscht und der Algorithmus gerät in eine Endlosschleife.

Um Cycling zu vermeiden, muss man den Bogen  $f$ , der den Baum  $T$  verlässt, sorgfältig auswählen. Wir nennen eine Baumlösung  $(x, T, L, U)$  *degeneriert*, falls ein Bogen  $a \in T$  mit  $x_a = l_a$  oder  $x_a = u_a$  existiert. Außerdem wählen wir einen festen Knoten  $r \in V$ , der im Folgenden als Wurzel aller Spannbäume  $T$  fungieren wird.

**(7.31) Definition.** Eine zulässige Baumlösung  $(x, T, L, U)$  heißt stark zulässig, wenn für jeden Knoten  $i \neq r$  der eindeutige  $[i, r]$ -Weg  $P$  in  $T$  ein augmentierender Weg für  $x$  ist, d. h. wenn für alle Bögen  $a \in P$  gilt:  $l_a < x_a$ , falls  $a$  ein Rückwärtsbogen und  $x_a < u_a$ , falls  $a$  ein Vorwärtsbogen ist. Insbesondere ist jede nicht degenerierte zulässige Baumlösung stark zulässig.  $\triangle$

Es ist leicht zu sehen, dass bei der Wahl  $r = k$  die Start-Baumlösung mit  $(T, L, U) = (A' \setminus A, A, \emptyset)$  für  $\mathcal{T}'$  nicht degeneriert und daher stark zulässig ist.

Sei  $e$  der im Pricing-Schritt gewählte Bogen und  $C$  der Kreis in  $T + e$ . Als Orientierung für  $C$  wählen wir die Richtung von  $e$ , falls  $e \in L$  und die entgegengesetzte Richtung, falls  $e \in U$  (der Grund für diese Abweichung von Algorithmus (7.29) wird im Beweis von Satz (7.37) klar). Ein Bogen in  $C - e$  heißt *blockierend*, wenn er nach dem Augmentieren eine Kapazitätsgrenze erreicht. Außerdem nennen wir den eindeutigen Knoten von  $C$ , der  $r$  in  $T$  am nächsten ist, den *Scheitel* von  $C$ . Wenn nun der Bogen  $f \in C$  im Augmentierungsschritt nach der folgenden Regel gewählt wird, ist die resultierende Baumlösung wieder stark zulässig.

**(7.32) Regel des letzten blockierenden Bogens.** Durchlaufe  $C$  beginnend beim Scheitel  $s$  in der Orientierung von  $C$  und wähle  $f$  als letzten blockierenden Bogen.  $\triangle$

**(7.33) Satz.** Sei  $(x, T, L, U)$  eine stark zulässige Baumlösung und  $e \notin T$  der im Pricing-Schritt gewählte Bogen. Wenn im Augmentierungsschritt der Bogen  $f$  gemäß (7.32) gewählt wird, so ist die neue Baumlösung nach dem Update-Schritt stark zulässig.  $\triangle$

**Beweis.** Sei  $x'$  der Fluss nach dem Augmentierungsschritt. Wir müssen zeigen, dass für jeden Knoten  $i \neq r$  der  $[i, r]$ -Weg in  $T' := (T - f) + e$  bezüglich  $x'$  augmentierend ist. Dazu zerlegen wir den Kreis  $C$  in  $T + e$  in einen Weg  $W$ , der beim Scheitel  $s$  beginnt und  $C$  in Orientierungsrichtung durchläuft, den Bogen  $f$ , und einen Weg  $W'$ , der in  $s$  endet, also  $C = (W, f, W')$  (siehe Abbildung 7.2).

**1. Fall:**  $i$  ist der Scheitel  $s$  von  $C$ . In  $T$  und  $T'$  ist  $i$  auf demselben  $[i, r]$ -Weg  $P$  mit  $r$  verbunden. Für einen Bogen  $a \in P$  ändert sich weder der Fluss noch die Richtung zu  $r$ . Da  $P$  in  $T$  augmentierend ist, ist  $P$  daher auch in  $T'$  augmentierend. Für die folgenden Fälle genügt es also zu zeigen, dass der  $[i, s]$ -Weg augmentierend ist.

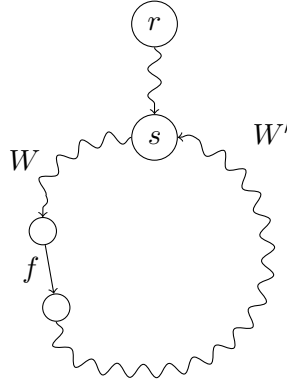


Abbildung 7.2: Situation im Beweis von Satz (7.33).

2. **Fall:**  $i$  ist auf dem Weg  $W'$ . Weil  $f$  der letzte blockierende Bogen ist, wenn  $C$  in Orientierungsrichtung durchlaufen wird, ist kein Bogen in  $W'$  blockierend in der Richtung von  $C$ . Zusammen mit Fall 1 folgt, dass kein Bogen auf dem  $[i, r]$ -Weg blockierend ist.
3. **Fall:**  $i$  ist auf dem Weg  $W$ . Sei  $\varepsilon \geq 0$  der Augmentierungswert im Augmentierungsschritt. Ist  $\varepsilon > 0$ , so ist  $x'$  auf  $W$  um  $\varepsilon$  Einheiten größer als  $x$ , also ist kein Bogen von  $W$  in Richtung  $s$  blockierend. Gilt hingegen  $\varepsilon = 0$ , so impliziert  $f \notin W$ , dass der  $[i, s]$ -Weg  $P$  in  $T$  und  $T'$  übereinstimmt und sich darauf weder Fluss noch Richtung zu  $r$  ändert. Wie im 1. Fall ist  $P$  daher auch in  $T'$  bezüglich  $x'$  augmentierend.
4. **Fall:**  $i$  ist nicht in  $C$ . Da  $(x, T, L, U)$  stark zulässig ist, ist der  $[i, r]$ -Weg augmentierend bezüglich  $x$ . Sind  $P$  und  $C$  bogendisjunkt, so ist  $P$  auch der  $[i, r]$ -Weg in  $T'$  und daher augmentierend bezüglich  $x'$ . Sonst sei  $j$  der erste Knoten auf  $P$ , der zu  $C$  gehört. In  $T'$  ist der  $[i, j]$ -Weg augmentierend bezüglich  $x'$ , weil in  $T$  augmentierend bezüglich  $x$ . Da für den Knoten  $j$  einer der ersten drei Fälle gilt, ist auch der  $[i, r]$ -Weg augmentierend in  $T'$ .  $\square$

Das folgende Ergebnis benötigen wir für den Terminierungs-Beweis. Es ermöglicht außerdem eine effizientere Berechnung des Knotenpreisvektors in Schritt 2 von Algorithmus (7.29).

**(7.34) Lemma.** Sei  $(x, T, L, U)$  eine Baumlösung in Algorithmus (7.29) und sei  $y$  der zugehörige Knotenpreisvektor. Seien weiterhin die Bögen  $e = (u, v)$  und  $f$  die im Pricing- bzw. Augmentierungsschritt gewählten Bögen und  $y'$  der neue Knotenpreisvektor. Schließlich seien  $T_1$  die Zusammenhangskomponente von  $T - f$ , die  $r$  enthält, und  $T_2 := V \setminus T_1$ . Dann gilt:

$$y'_i = \begin{cases} y_i & i \in T_1, \\ y_i + \bar{c}_e & i \in T_2, u \in T_1, \\ y_i - \bar{c}_e & i \in T_2, u \in T_2. \end{cases} \quad (7.35) \quad \triangle$$



**Beweis.** Setze  $T' := (T + e) - f$ . Nach Definition von  $y'$  und der reduzierten Kosten bezüglich  $y$  gelten für  $e$  die Gleichungen

$$\begin{aligned} c_{uv} + y'_u - y'_v &= 0, \\ \bar{c}_e &= \bar{c}_{uv} = c_{uv} + y_u - y_v. \end{aligned} \tag{7.36}$$

Betrachte einen Knoten  $i \in T_1$ . Da der  $[i, r]$ -Weg in  $T$  mit dem in  $T'$  übereinstimmt und sich darauf die Flüsse nicht ändern, folgt aus  $y_r = y'_r$  (wir haben  $r = k$  und  $y_k = 0$  per Definition) direkt  $y'_i = y_i$ .

Im Fall  $u \in T_1, v \in T_2$  gilt  $y'_u = y_u$  und mit (7.36) folgt

$$y'_v = c_{uv} + y'_u = c_{uv} + y_u = y_v + \bar{c}_e.$$

Außerdem stimmen die  $[v, i]$ -Pfade in  $T$  und  $T'$  für alle  $i \in T_2$  überein, sodass  $y'_i = y_i + \bar{c}_e$  gilt.

Der Fall  $u \in T_2, v \in T_1$  ist analog: Wegen  $y'_v = y_v$  und (7.36) gilt

$$y'_u = -c_{uv} + y'_v = -c_{uv} + y_v = y_u - \bar{c}_e,$$

woraus sich dann  $y'_i = y_i - \bar{c}_e$  für alle  $i \in T_2$  ergibt.  $\square$

**(7.37) Satz.** Sei  $\mathcal{I}$  eine Instanz des Minimalkosten-Flussproblems mit rationalen Daten  $c, l, u, b$ . Dann terminiert Algorithmus (7.29) nach endlich vielen Schritten wenn Regel (7.32) zur Auswahl des Bogens  $f$  im Augmentierungsschritt verwendet wird.  $\triangle$

**Beweis.** Wir können o. B. d. A. annehmen, dass die Eingabedaten ganzzahlig sind. Dann verringert jede Augmentierung mit positivem  $\varepsilon$  die Kosten um mindestens 1. Da die optimalen Kosten nach unten beschränkt sind, kann dies nur endlich oft auftreten. Zu zeigen ist daher, dass jeweils nur endlich viele aufeinanderfolgende Augmentierungen mit  $\varepsilon = 0$  durchgeführt werden. Dafür verwenden wir ein ähnliches Argument wie für positives  $\varepsilon$ , allerdings bezogen auf eine andere Funktion, die „den Fortschritt misst“; statt der Kosten betrachten wir die Summe der Knotenpreise  $\sum_{i \in V} y_i$ . Die maximale Differenz

$$\max_{i, j \in V, i \neq j} |y_i - y_j|$$

zweier Knotenpreise ist beschränkt durch  $|V| \max_{a \in A'} |c_a| =: M' < \infty$ . Da  $y_k = 0$  ist, gilt  $y_i \leq M'$  für alle  $i \in V$ . Daher ist die maximale Summe aller Knotenpreise nach oben beschränkt. Wenn wir nun zeigen können, dass die Summe der Knotenpreise bei jedem Augmentierungsschritt mit  $\varepsilon = 0$  zunimmt, d. h.  $\sum_{i \in V} y'_i > \sum_{i \in V} y_i$  gilt, so folgt, dass es nur endlich viele solche aufeinanderfolgende Augmentierungsschritte geben kann.

Betrachten wir einen Augmentierungsschritt mit  $\varepsilon = 0$ , mit stark zulässiger Baumlösung  $(x, T, L, U)$  und  $C, e = (u, v), f, y, y'$  wie bisher. Wir unterscheiden zwei Fälle gemäß des Pricing-Schritts (siehe Abbildung 7.3):

$e \in L, \bar{c}_e < 0$ : Der Bogen  $e$  und  $C$  sind gleich orientiert. Dann liegt  $f$  zwischen dem Scheitel  $s$  von  $C$  und  $u$ : Weil nämlich  $(x, T, L, U)$  stark zulässig ist, ist der  $[v, s]$ -Weg  $P$  als Teil des  $[v, r]$ -Weges augmentierend bezüglich  $x$ . Wegen  $\varepsilon = 0$  kann dann kein Bogen auf  $P$  blockierend sein; insbesondere kann  $f$  nicht auf  $P$  liegen. Daher ist  $u \in T_2$  und wir haben wegen Lemma (7.34)  $y'_i = y_i - \bar{c}_e > y_i$  für alle  $i \in T_2$ .

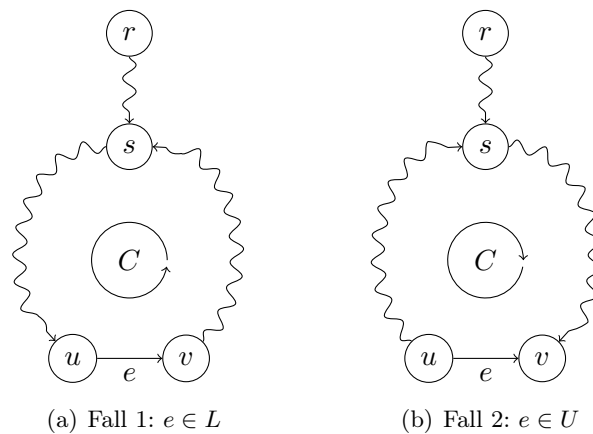


Abbildung 7.3: Fälle im Beweis von Satz (7.37).

$e \in U, \bar{c}_e > 0$ : Der Bogen  $e$  und  $C$  sind entgegengesetzt orientiert. Nun liegt  $f$  zwischen dem Scheitel  $s$  von  $C$  und  $v$ , weil der  $[u, s]$ -Weg augmentierend bezüglich  $x$  ist und daher wie im vorherigen Fall kein Bogen auf diesem Weg blockierend sein kann. Also ist  $u \in T_1$  und wir erhalten mit Lemma (7.34)  $y'_i = y_i + \bar{c}_e > y_i$  für alle  $i \in T_2$ .

Wegen Lemma (7.34) gilt  $y'_i = y_i$  für  $i \in T_1$  und in beiden Fällen haben wir für  $i \in T_2$  gezeigt, dass  $y'_i > y_i$  ist. Daraus folgt  $\sum_{i \in V} y'_i > \sum_{i \in V} y_i$  und der Satz ist bewiesen.  $\square$

Selbst nach genauerer Spezifikation der Auswahl des Bogens  $f$  bietet Algorithmus (7.29) noch viel Raum für Varianten. Auf spezielle Datenstrukturen für eine sehr effiziente Implementierung werden wir noch in den Übungen eingehen. Sehr wichtig für die praktische Laufzeit ist die genaue Regel für die Auswahl des Bogens  $e$  im Pricing-Schritt. Obwohl für Netzwerksimplex-Algorithmen Problembeispiele mit exponentieller Laufzeit bekannt sind, ist diese Algorithmengruppe in der Praxis oft effizienter als die bekannten kombinatorischen Algorithmen mit polynomialer Laufzeit.

Es gibt viele Implementierungen des Netzwerksimplex-Algorithmus. Ein von A. Löbel (Konrad-Zuse-Zentrum, Berlin) implementierter Code dieser Art, es handelt sich um einen sogenannten primal-dualen Netzwerk-Simplex-Algorithmus, ist auf dem ZIB-Server für akademische Nutzung verfügbar, siehe URL:

[http://typo.zib.de/opt-long\\_projects/Software/Mcf/](http://typo.zib.de/opt-long_projects/Software/Mcf/)

Mit diesem Code namens MCF können Minimalkosten-Flussprobleme mit Zigtausenden Knoten und Hundertmillionen Bögen in wenigen Minuten gelöst werden. MCF findet derzeit u. a. in verschiedenen Planungssystemen für den öffentlichen Nahverkehr Anwendung. MCF ist als einer der Integer-Benchmark Codes in die SPEC CPU2006-Suite aufgenommen worden, mit der Leistungsevaluierungen moderner Computersysteme vorgenommen werden, siehe URL:

<http://www.spec.org>

## Literaturverzeichnis

- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Handbooks in Operations Research and Management Science*, volume 1, chapter Optimization, pages 211–360. Elsevier, North-Holland, Amsterdam, 1989.
- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Theory, Algorithms and Applications*. Paramount Publishing International, Prentice Hall, New York, 1993.
- L. R. Ford, Jr. and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, 1962.
- V. Goldberg, É. Tardos, and R. E. Tarjan. Network Flow Algorithms. In B. Korte, L. Lovasz, H. J. Prömel, and A. Schrijver, editors, *Paths, Flows, and VLSI-Layout*, pages 101–164. Springer-Verlag, Berlin, 1990.
- M. Grötschel. Tiefensuche: Bemerkungen zur Algorithmengeschichte. In H. Hecht, R. Mikosch, I. Schwarz, H. Siebert, and R. Werther, editors, *Kosmos und Zahl - Beiträge zur Mathematik- und Astronomiegeschichte, zu Alexander von Humboldt und Leibniz*, volume 58 of *Boethius*, pages 331–346. Franz Steiner Verlag, 2008. URL <http://www.steiner-verlag.de/titel/56734.html>.
- T. C. Hu. *Integer Programming and Network Flows*. Addison-Wesley, Reading, Massachusetts, 1969.
- P. A. Jensen and J. W. Barnes. *Network Flow Programming*. Wiley & Sons, New York, 1980.
- D. Jungnickel. *Graphs, Networks and Algorithms*. Springer, 4. edition, 2013.
- J. Kennington and R. Helgason. *Algorithms for Network Programming*. Wiley & Sons, New York, 1980.
- H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Logistics Quarterly*, 2:83–97, 1955.
- H. W. Kuhn. A tale of three eras: The discovery and rediscovery of the Hungarian Method. *European Journal of Operational Research*, 219:641–651, 2012.
- E. L. Lawler. *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart & Winston, New York, 1976.
- S. I. M. Shigeno and S. T. McCormick. Relaxed most negative cycle and most positive cut canceling algorithms for minimum cost flow. *Mathematics of Operations Research*, 25:76–104, 2000.



## 8 Grundlagen der Polyedertheorie

In diesem Kapitel beschäftigen wir uns mit „Begrenzungsflächen“ von Polyedern  $P$  und erarbeiten für die lineare Optimierung grundlegende (algebraische) Charakterisierungen.

Wir erinnern an die Einführung der Grundbegriffe der Polyedertheorie und an die Festlegung einiger Bezeichnungen in Abschnitt 2.3. Insbesondere sei noch einmal erwähnt, dass wir ein Polyeder, das als Lösungsmenge eines linearen Ungleichungssystems  $Ax \leq b$  definiert ist, mit  $P(A, b)$  bezeichnen. Ferner wollen wir nochmals auf eine die Schreibtechnik vereinfachende Konvention hinweisen. Wir betrachten Matrizen und endliche Mengen als (im Wesentlichen) ein und dieselben Objekte. Ist z. B.  $A$  eine  $(m, n)$ -Matrix, so schreiben wir  $\text{conv}(A)$  und meinen damit die konvexe Hülle der Spaltenvektoren von  $A$ . Ist umgekehrt z. B.  $V \subseteq \mathbb{K}^n$  eine endliche Menge, so fassen wir  $V$  auch als eine  $(n, |V|)$ -Matrix auf und schreiben  $V^T$ , um die Matrix zu bezeichnen, deren Zeilen die Vektoren aus  $V$  sind.  $V^T$  ist natürlich nicht eindeutig definiert, da wir keine Reihenfolge der Vektoren aus  $V$  angegeben haben. Wir werden diese Schreibweise jedoch nur dann benutzen, wenn es auf die Reihenfolge der Zeilen nicht ankommt.

**(8.1) Definition.** Es seien  $S \subseteq \mathbb{K}^n$ ,  $a \in \mathbb{K}^n$ ,  $\alpha \in \mathbb{K}$ .

- (a) Die Ungleichung  $a^T x \leq \alpha$  heißt gültig bezüglich  $S$ , falls  $S \subseteq \{x \in \mathbb{K}^n \mid a^T x \leq \alpha\}$ .
- (b) Eine Hyperebene  $H = \{x \mid a^T x = \alpha\}$ ,  $a \neq 0$ , heißt Stützhyperebene von  $S$ , falls  $a^T x \leq \alpha$  gültig bezüglich  $S$  und  $S \cap H \neq \emptyset$  ist.  $\triangle$

Zu sagen, dass  $a^T x \leq \alpha$ ,  $a \neq 0$ , gültig bezüglich  $S$  ist, heißt nichts anderes als:  $S$  ist im Halbraum  $\{x \in \mathbb{K}^n \mid a^T x \leq \alpha\}$  enthalten.

Wir werden nun diejenigen Teilmengen von Polyedern untersuchen, die als Durchschnitte mit Stützhyperebenen entstehen.

**(8.2) Definition.**  $P \subseteq \mathbb{K}^n$  sei ein Polyeder. Eine Menge  $F \subseteq P$  heißt Seitenfläche von  $P$ , wenn es eine gültige Ungleichung  $c^T x \leq \gamma$  bezüglich  $P$  gibt mit

$$F = P \cap \{x \mid c^T x = \gamma\}.$$

Eine Seitenfläche  $F$  von  $P$  heißt echt, wenn  $F \neq P$  gilt.  $F$  heißt nichttrivial, wenn  $\emptyset \neq F \neq P$  gilt. Ist  $c^T x \leq \gamma$  gültig bezüglich  $P$ , dann heißt

$$F := P \cap \{x \mid c^T x = \gamma\}$$

die von  $c^T x \leq \gamma$  induzierte oder definierte Seitenfläche.  $\triangle$

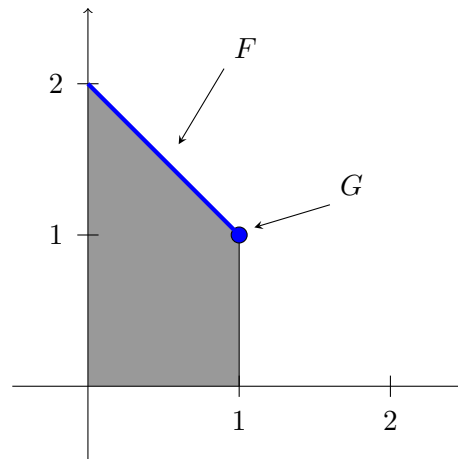


Abbildung 8.1: Polyeder aus Beispiel (8.3).

Offenbar sind Seitenflächen von Polyedern wiederum Polyeder. Eine Seitenfläche  $F$  kann durchaus von verschiedenen Ungleichungen definiert werden. Trivialerweise gilt  $F = P \cap \{x \mid c^T x = \gamma\} = P \cap \{x \mid \lambda c^T x = \lambda \gamma\}$  für alle  $\lambda > 0$ , aber auch zwei Ungleichungen, die nicht durch Skalierung auseinander hervorgehen, können  $F$  definieren.

**(8.3) Beispiel.** Sei  $P(A, b) \subseteq \mathbb{K}^2$  das durch

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

definierte Polyeder (siehe Abbildung 8.1), dann ist das Geradenstück zwischen  $(1, 1)^T$  und  $(0, 2)^T$  eine Seitenfläche  $F$ , definiert durch  $x_1 + x_2 \leq 2$ . Die Menge  $G = \{(1, 1)^T\}$  ist ebenfalls eine Seitenfläche von  $P$ , denn

$$\begin{aligned} G &= P(A, b) \cap \{x \mid 2x_1 + x_2 = 3\} \\ &= P(A, b) \cap \{x \mid 3x_1 + x_2 = 4\}, \end{aligned}$$

und, wie man sieht, kann  $\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$  durch zwei völlig unterschiedliche Ungleichungen definiert werden.  $\triangle$

**(8.4) Folgerung.** Sei  $P \subseteq \mathbb{K}^n$  ein Polyeder, dann gilt:

- (a)  $P$  ist eine Seitenfläche von sich selbst.
- (b)  $\emptyset$  ist eine Seitenfläche von  $P$ .
- (c) Ist  $F = \{x \in P \mid c^T x = \gamma\}$  eine nichttriviale Seitenfläche von  $P$ , dann gilt  $c \neq 0$ .  $\triangle$

**Beweis.** (a)  $P = \{x \in P \mid 0^T x = 0\}$ .

(b)  $\emptyset = \{x \in P \mid 0^T x = 1\}$ .

(c)  $\{x \mid 0^T x = \alpha\}$  ist entweder leer ( $\alpha \neq 0$ ) oder  $\mathbb{K}^n$  ( $\alpha = 0$ ). Falls  $F$  nichttrivial ist, muss also  $c \neq 0$  gelten.  $\square$

**(8.5) Satz.** Seien  $P = P(A, b) \neq \emptyset$  ein Polyeder,  $c \in \mathbb{K}^n$ ,  $\gamma \in \mathbb{K}$  und

$$z^* = \begin{cases} +\infty & \text{falls } c^T x \text{ auf } P \text{ unbeschränkt,} \\ \max\{c^T x \mid x \in P\} & \text{andernfalls.} \end{cases}$$

(a)  $c^T x \leq \gamma$  ist gültig bezüglich  $P \iff \gamma \geq z^*$ .

(b)  $z^* < +\infty \implies F = \{x \in P \mid c^T x = z^*\}$  ist eine nichtleere Seitenfläche von  $P$ , und  $\{x \mid c^T x = z^*\}$  ist eine Stützhyperebene von  $P$ , falls  $c \neq 0$ .

(c) Die Menge der Optimallösungen eines linearen Programms ist eine Seitenfläche des durch die Nebenbedingungen definierten Polyeders.  $\triangle$

**Beweis.** (a) Ist  $\gamma < z^*$ , dann existiert ein  $x^* \in P$  mit  $c^T x^* > \gamma$ , also ist  $c^T x \leq \gamma$  nicht gültig. Ist  $\gamma \geq z^*$ , so gilt  $c^T x \leq z^* \leq \gamma \forall x \in P$ , also ist  $c^T x \leq \gamma$  gültig.

(b) Nach (a) ist  $c^T x \leq z^*$  gültig bezüglich  $P$  und nach Definition existiert ein  $x^* \in P$  mit  $c^T x^* = z^*$ , also ist  $F$  eine nichtleere Seitenfläche von  $P$ . Offenbar ist  $\{x \mid c^T x = z^*\}$  eine Stützhyperebene, falls  $c \neq 0$ .

(c) folgt aus (b).  $\square$

**(8.6) Definition.** Sei  $P = P(A, b) \subseteq \mathbb{K}^n$  ein Polyeder und  $M$  die Zeilenindexmenge von  $A$ . Für  $F \subseteq P$  sei

$$\text{eq}(F) := \{i \in M \mid A_i x = b_i \forall x \in F\},$$

d. h.  $\text{eq}(F)$  (genannt Gleichheitsmenge oder equality set von  $F$ ) ist die Menge der für alle  $x \in F$  bindenden Restriktionen. Für  $I \subseteq M$  sei

$$\text{fa}(I) := \{x \in P \mid A_I x = b_I\}.$$

Wir nennen  $\text{fa}(I)$  die von  $I$  induzierte Seitenfläche.  $\triangle$

Offenbar ist  $\text{fa}(I)$  tatsächlich eine Seitenfläche von  $P = P(A, b)$ , denn setzen wir  $c^T := \sum_{i \in I} A_i$ ,  $\gamma := \sum_{i \in I} b_i$ , so ist  $c^T x \leq \gamma$  gültig bezüglich  $P(A, b)$ , und, wie man leicht sieht, gilt

$$\text{fa}(I) = \{x \in P \mid c^T x = \gamma\}.$$

Zur Veranschaulichung der oben definierten Abbildungen  $\text{fa}$  und  $\text{eq}$  betrachten wir Beispiel (8.3). Für das in (8.3) definierte Polyeder  $P(A, b)$  gilt  $M = \{1, 2, 3, 4\}$  und

$$\begin{aligned} \text{fa}(\{1, 2\}) &= G, \\ \text{eq}\left(\left\{\left\{\begin{pmatrix} 0 \\ 2 \end{pmatrix}\right\}\right\}\right) &= \{1, 3\}. \end{aligned}$$

Für die lineare Optimierung sind vor allem niedrigdimensionale Seitenflächen relevant, über die gewissermaßen der „Rand“ aller höherdimensionalen Seitenflächen und damit auch des Polyeders  $P$  beschrieben werden kann.

**(8.7) Definition.** Es seien  $P \subseteq \mathbb{K}^n$  ein Polyeder und  $F$  eine Seitenfläche von  $P$ . Gibt es  $x \in \mathbb{K}^n$  und  $0 \neq z \in \mathbb{K}^n$  mit

$$\left\{ \begin{array}{l} F = \{x\} \\ F = x + \text{lin}(\{z\}) \\ F = x + \text{cone}(\{z\}) \end{array} \right\}, \text{ so heißt } F \left\{ \begin{array}{l} \text{Ecke} \\ \text{Extremallinie} \\ \text{Extremalstrahl} \end{array} \right\}. \quad \triangle$$

Ist  $F = \{x\}$  eine Ecke von  $P$ , so sagen wir einfach „ $x$  ist eine Ecke von  $P$ “. Wir werden uns im weiteren insbesondere für Ecken interessieren und sie charakterisieren. Offenbar sind Ecken Seitenflächen der Dimension 0, während Extremallinien und Extremalstrahlen Seitenflächen der Dimension 1 sind. Allgemein heißen Seitenflächen der Dimension 1 *Kanten*. Kanten sind entweder Extremallinien, Extremalstrahlen oder Verbindungsstrecken zwischen zwei Ecken. (Nicht jede Verbindungslinie zwischen zwei Ecken ist eine Kante!) Sind zwei Ecken  $x, y$  eines Polyeders  $P$  durch eine Kante verbunden, d. h.  $\text{conv}(\{x, y\})$  ist eine Seitenfläche von  $P$ , so nennt man  $x$  und  $y$  *adjazent* auf  $P$ .

**(8.8) Satz.** Seien  $P = P(A, b) \subseteq \mathbb{K}^n$  ein Polyeder und  $x \in P$ . Dann sind die folgenden Aussagen äquivalent:

- (a)  $x$  ist eine Ecke von  $P$ .
- (b)  $\{x\}$  ist eine nulldimensionale Seitenfläche von  $P$ .
- (c)  $\text{rang}(A_{\text{eq}(\{x\})}) = n$ .
- (d)  $\exists c \in \mathbb{K}^n \setminus \{0\}$ , so dass  $x$  die eindeutig bestimmte Optimallösung des linearen Programms  $\max c^T y, y \in P$  ist. △

**Beweis.** (a)  $\iff$  (b): Definition!

- (a)  $\implies$  (d): Nach Definition ist  $\{x\}$  eine Seitenfläche, also existiert eine bezüglich  $P$  gültige Ungleichung  $c^T y \leq \gamma$ , so dass  $\{y \in P \mid c^T y = \gamma\} = \{x\}$  gilt. Folglich ist  $x$  die eindeutig bestimmte Optimallösung von  $\max c^T y, y \in P$ . Ist  $P \neq \{x\}$ , so ist nach (8.4)(c)  $c \neq 0$ , andernfalls kann  $c \neq 0$  gewählt werden.



- (d)  $\implies$  (c): Sei  $x$  die eindeutige Optimallösung des linearen Programms  $\max c^T y$ ,  $y \in P$  für ein  $c \in \mathbb{K}^n \setminus \{0\}$ . Wir nehmen  $\text{rang}(A_{\text{eq}(\{x\})}) < n$  an und betrachten die durch  $I := \text{eq}(\{x\})$  induzierte Seitenfläche  $F = \text{fa}(I)$ . Dann gibt es ein  $y \neq 0$  im Kern der Untermatrix  $A_I$  von  $A$ . Für  $J := \{1, \dots, m\} \setminus I$  gilt nach Definition von  $\text{eq}(\{x\})$  und wegen der Zulässigkeit von  $x$  die strikte Ungleichung  $A_J x < b_J$ . Für genügend kleines  $\epsilon > 0$  folgt daher  $x \pm \epsilon y \in F$ , weil einerseits  $A_I(x \pm \epsilon y) = b_I$  und andererseits  $A_J(x \pm \epsilon y) < b_J$  gilt. Nun hat entweder einer der beiden Punkte  $x \pm \epsilon y$  einen besseren Zielfunktionswert als  $x$  oder beide Punkte haben denselben Zielfunktionswert wie  $x$ . In beiden Fällen ergibt sich ein Widerspruch, es muss also  $\text{rang}(A_{\text{eq}(\{x\})}) = n$  gelten.
- (c)  $\implies$  (b): Sei  $I = \text{eq}(\{x\})$ , dann hat  $A_I y = b_I$  nach Voraussetzung eine eindeutig bestimmte Lösung, nämlich  $x$ . Daraus folgt  $F = \{y \in P \mid A_I y = b_I\} = \{x\}$ , also ist  $\{x\}$  eine nulldimensionale Seitenfläche.  $\square$

Für Polyeder der Form  $P^=(A, b) = \{x \in \mathbb{K}^n \mid Ax = b, x \geq 0\}$  gibt es eine weitere nützliche Kennzeichnung der Ecken. Ist  $x \in \mathbb{K}^n$ , so setzen wir

$$\text{supp}(x) := \{i \in \{1, \dots, n\} \mid x_i \neq 0\}.$$

Die Indexmenge  $\text{supp}(x)$  heißt *Träger* von  $x$ .

**(8.9) Satz.** Für  $x \in P^=(A, b) \subseteq \mathbb{K}^n$  sind folgende Aussagen äquivalent:

- (a)  $x$  ist Ecke von  $P^=(A, b)$ .  
 (b)  $\text{rang}(A_{\text{supp}(x)}) = |\text{supp}(x)|$ .  
 (c) Die Spaltenvektoren  $A_{\cdot j}$ ,  $j \in \text{supp}(x)$ , sind linear unabhängig.  $\triangle$

**Beweis.** (b)  $\iff$  (c): Trivial.

- (a)  $\iff$  (b): Sei  $D = \begin{pmatrix} A \\ -A \\ -I \end{pmatrix}$ ,  $d = \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}$ , dann gilt  $P(D, d) = P^=(A, b)$ . Mit Satz (8.8)

gilt nun:

$$\begin{aligned} x \text{ Ecke von } P^=(A, b) &\iff x \text{ Ecke von } P(D, d) \\ &\iff \text{rang}(D_{\text{eq}(\{x\})}) = n \\ &\iff \text{rang} \begin{pmatrix} A \\ -A \\ I_J \end{pmatrix} = n \text{ mit } J = \{j \mid x_j = 0\} \\ &\hspace{15em} = \{1, \dots, n\} \setminus \text{supp}(x) \\ (\text{Streichen der Spalten } j \in J) &\iff \text{rang} \begin{pmatrix} A_{\cdot \text{supp}(x)} \\ -A_{\cdot \text{supp}(x)} \end{pmatrix} = n - |J| = |\text{supp}(x)| \\ &\iff \text{rang}(A_{\text{supp}(x)}) = |\text{supp}(x)|. \quad \square \end{aligned}$$

Nicht alle Polyeder haben Ecken; z. B. haben die Polyeder, die als Durchschnitt von weniger als  $n$  Halbräumen des  $\mathbb{K}^n$  entstehen, keine Ecken. Polyeder mit Ecken sind besonders für die lineare Optimierung wichtig.

**(8.10) Definition.** Ein Polyeder heißt spitz, wenn es eine Ecke besitzt.  $\triangle$

Der folgende Satz liefert eine algebraische Charakterisierung spitzer Polyeder.

**(8.11) Satz.** Sei  $P = P(A, b) \subseteq \mathbb{K}^n$  ein nichtleeres Polyeder, dann sind äquivalent:

(a)  $P$  ist spitz.

(b)  $\text{rang}(A) = n$ .

(c) Jede nichtleere Seitenfläche von  $P$  ist spitz.  $\triangle$

**Beweis.** (a)  $\implies$  (b): Ist  $x$  eine Ecke von  $P$ , so gilt nach (8.8)  $n = \text{rang}(A_{\text{eq}(\{x\})}) \leq \text{rang}(A) \leq n$ , also  $\text{rang}(A) = n$ .

(b)  $\implies$  (a): Sei  $x \in P$  so gewählt, dass  $I := \text{eq}(\{x\})$  maximal bezüglich Mengeninklusion ist. Sei  $F = \{y \in P \mid A_I \cdot y = b_I\}$ . Ist  $\text{rang}(A_I) < n$ , dann enthält der Kern von  $A_I$  einen Vektor  $y \neq 0$ . Wie im Beweis von Satz (8.8) gibt es dann ein  $\varepsilon > 0$ , so dass  $x \pm \varepsilon y \in P$ . Die Gerade  $G = \{x + \lambda y \mid \lambda \in \mathbb{K}\}$  trifft mindestens eine der Hyperebenen  $H_j = \{z \mid A_j \cdot z = b_j\}$ ,  $j \notin I$ , da  $\text{rang}(A) > \text{rang}(A_I)$ . Also muss es ein  $\delta \in \mathbb{K}$  geben, so dass  $x + \delta y \in P$  und  $\text{eq}(\{x + \delta y\}) \supset I$ . Dies widerspricht der Maximalität von  $I$ .

(c)  $\implies$  (a): Trivial, da  $P$  eine Seitenfläche von sich selbst ist.

(b)  $\implies$  (c): Für jede nichtleere Seitenfläche  $F$  von  $P$  gilt

$$F = \{x \in \mathbb{K}^n \mid Ax \leq b, A_{\text{eq}(F)} \cdot x \leq b_{\text{eq}(F)}, -A_{\text{eq}(F)} \cdot x \leq -b_{\text{eq}(F)}\}.$$

Aus (b) und der Äquivalenz von (b) und (a), angewandt auf  $F$ , folgt damit direkt, dass  $F$  spitz ist.  $\square$

**(8.12) Folgerung.** Sei  $P = P^-(A, b) \subseteq \mathbb{K}^n$ , dann gilt

$$P \neq \emptyset \iff P \text{ spitz.} \quad \triangle$$

**Beweis.** Es ist  $P = P(D, d)$  mit  $D = \begin{pmatrix} A \\ -A \\ -I \end{pmatrix}$ ,  $d = \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}$ , und offenbar hat  $D$  den Rang  $n$ . Aus Satz (8.11) folgt dann die Behauptung.  $\square$

**(8.13) Folgerung.** Sei  $P \subseteq \mathbb{K}^n$  ein Polytop, dann gilt

$$P \neq \emptyset \iff P \text{ spitz.} \quad \triangle$$

**Beweis.** Da  $P$  beschränkt ist, gibt es einen Vektor  $u$  mit  $P \subseteq \{x \mid x \leq u\}$ . Gilt  $P = P(A, b)$ , so folgt daraus  $P = P(D, d)$  mit  $D = \begin{pmatrix} A \\ I \end{pmatrix}$ ,  $d = \begin{pmatrix} b \\ u \end{pmatrix}$ .  $D$  hat den Rang  $n$ , daraus folgt mit (8.11) die Behauptung.  $\square$

**(8.14) Folgerung.** Das Polyeder  $P$  sei spitz, und das lineare Programm  $\max c^T x$ ,  $x \in P$  habe eine Optimallösung, dann hat dieses lineare Programm auch eine optimale Lösung, die eine Ecke von  $P$  ist (eine sogenannte optimale Ecklösung).  $\triangle$

**Beweis.**  $F = \{x \in P \mid c^T x = \max\{c^T y \mid y \in P\}\}$  ist eine nichtleere Seitenfläche von  $P$ , die nach Satz (8.11) eine Ecke hat.  $\square$

**(8.15) Folgerung.** Ist  $P$  ein nichtleeres Polytop, dann hat jedes lineare Programm  $\max c^T x$ ,  $x \in P$  eine optimale Ecklösung.  $\triangle$

Das folgende Korollar wird sich als sehr wichtig erweisen.

**(8.16) Folgerung.** Lineare Programme der Form

$$\begin{array}{ccc} \max c^T x & & \max c^T x \\ Ax = b & \text{oder} & Ax \leq b \\ x \geq 0 & & x \geq 0 \end{array}$$

besitzen genau dann eine endliche Optimallösung, wenn sie eine optimale Ecklösung besitzen.  $\triangle$

**Beweis.** Nach (8.12) ist  $P = P(A, b)$  spitz (falls nicht leer) und somit hat das erste der obigen LPs nach (8.14) eine optimale Ecklösung, falls es überhaupt eine optimale Lösung hat. Analog folgt die Behauptung für das zweite LP.  $\square$

Wir beenden diesen kurzen Einblick in die Polyedertheorie mit einer interessanten Beobachtung. Wir haben bereits gesehen (Satz (8.8)), dass sich Ecken eines Polyeders  $P = P(A, b) \subseteq \mathbb{K}^n$  als Lösungsmenge eines Systems von  $n$  Gleichungen des Systems  $Ax = b$  darstellen lassen. Abstrakter gesagt, jede Ecke eines spitzen Polyeders  $P$  ist ein affiner Raum. Dies gilt allgemein für die „kleinsten“ Seitenflächen eines (nicht notwendig spitzen) Polyeders.

**(8.17) Definition.** Eine nichtleere Seitenfläche  $F$  eines Polyeders  $P$  heißt minimal, wenn sie inklusionsweise minimal ist, d. h. wenn keine Seitenfläche  $F' \neq F$  mit  $F' \subseteq F$  existiert.  $\triangle$

**(8.18) Satz.** Für jede nichtleere minimale Seitenfläche  $F$  eines Polyeders  $P = P(A, B)$  gilt

$$F = \{x \in P \mid A_{\text{eq}(F)} \cdot x = b_{\text{eq}(F)}\} = \{x \in \mathbb{K}^n \mid A_{\text{eq}(F)} \cdot x = b_{\text{eq}(F)}\} =: \overline{F},$$

d. h.  $F$  ist ein affiner Raum. △

**Beweis.**  $F \subseteq \overline{F}$  ist trivial.

Angenommen, es existiert ein  $y \in \overline{F} \setminus F$ . Für ein fest gewähltes  $z \in F$  erfüllt jeder Punkt des Strahls  $p(\lambda) := \lambda z + (1 - \lambda)y$ ,  $\lambda \in \mathbb{K}$ , das Gleichungssystem  $A_{\text{eq}(F)} \cdot x = b_{\text{eq}(F)}$ . Wegen  $y \notin P$  gibt es eine Ungleichung in  $Ax \leq b$ , die von  $y$  verletzt wird. Für jede solche Ungleichung  $A_i \cdot x \leq b_i$  existiert ein  $\lambda_i$ , sodass der Punkt  $p(\lambda_i)$  diese Ungleichung mit Gleichheit erfüllt. Sei  $i^*$  ein Index, für den  $|\lambda_i|$  minimal ist. Dann ist  $F'$  mit  $\text{eq}(F') = \text{eq}(F) \cup \{i^*\}$  eine Seitenfläche von  $P$  mit  $F' \subseteq F$ .  $F'$  ist nicht leer, weil es mindestens den Punkt  $p(\lambda_{i^*})$  enthält. Außerdem ist  $F'$  echt kleiner als  $F$ , weil die Gleichung  $A_{i^*} \cdot x \leq b_{i^*}$  nicht von allen Punkten von  $F$  erfüllt wird (sonst wäre  $i^* \in \text{eq}(F)$ ), also ist  $F$  nicht minimal. Widerspruch. □

## Literaturverzeichnis

B. Grünbaum. *Convex Polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 2003.

A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, New York, 1998.

G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer, 1995.

## 9 Die Grundversion des Simplex-Algorithmus

In Satz (8.5) haben wir gezeigt, dass die Menge der Optimallösungen eines linearen Programms eine Seitenfläche des durch die Restriktionen definierten Polyeders ist. Dieses Ergebnis haben wir in Folgerung (8.14) dahingehend verschärft, dass wir nachweisen konnten, dass bei spitzen Polyedern das Optimum, falls es existiert, stets in einer Ecke angenommen wird. Wollen wir also ein Verfahren entwerfen, das lineare Programme über spitzen Polyedern löst, so genügt es, ein Verfahren anzugeben, das eine optimale Ecke derartiger Polyeder findet.

Aus Abschnitt 2.3 wissen wir, dass wir uns auf lineare Programme über Polyedern der Form  $P^=(A, b) = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  beschränken können. In Folgerung (8.12) haben wir gezeigt, dass ein solches Polyeder stets spitz ist, falls es nicht leer ist. Wenn wir also einen Algorithmus angeben können, der die optimale Ecke eines linearen Programms der Form  $\max c^T x, x \in P^=(A, b)$  findet, so können wir mit diesem alle linearen Programmierungsprobleme lösen.

**(9.1) Definition.** *Ein lineares Programm der Form*

$$\begin{aligned} \max c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

*heißt ein lineares Programm in Standardform. Wenn nichts anderes gesagt wird, nehmen wir immer an, dass  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$ , und  $c \in \mathbb{K}^n$  gilt.  $\triangle$*

Ist der Rang von  $A$  gleich  $n$ , so wissen wir, dass das Gleichungssystem  $Ax = b$  entweder gar keine oder eine eindeutig bestimmte Lösung hat, die mit dem Gauß'schen Eliminationsverfahren berechnet werden kann. Da also entweder keine Lösung existiert oder der einzige zulässige Punkt leicht bestimmt werden kann, liegt kein wirkliches Optimierungsproblem vor. Wir wollen daher voraussetzen, dass  $\text{rang}(A) < n$  gilt. Ferner nehmen wir an, um die Darstellung einfacher zu machen, dass die Zeilen von  $A$  linear unabhängig sind und dass  $P^=(A, b) \neq \emptyset$  ist. Wir treffen also folgende Verabredung:

**(9.2) Generalvoraussetzungen für Kapitel 9.**

(a)  $A$  ist eine  $(m, n)$ -Matrix mit  $m < n$ ,

(b)  $\text{rang}(A) = m$ ,

(c)  $P^=(A, b) \neq \emptyset$ .  $\triangle$

## 9 Die Grundversion des Simplex-Algorithmus

Wir werden später zeigen, dass lineare Programme, bei denen die Voraussetzungen (9.2) nicht gegeben sind, auf solche zurückgeführt werden können, die diese erfüllen, dass also alle Voraussetzungen (9.2) o. B. d. A. gemacht werden können.

Um die weitere Darstellung schreibtechnisch zu vereinfachen, legen wir für dieses Kapitel folgendes fest (vergleiche Abschnitt 2.2):

**(9.3) Konventionen in Kapitel 9.** Es sei  $A \in \mathbb{K}^{(m,n)}$ , wobei nach (9.2)  $m < n$  gelten soll.

- (a) Mit  $\{1, \dots, m\}$  bezeichnen wir die Zeilenindexmenge von  $A$ , mit  $\{1, \dots, n\}$  die Spaltenindexmenge.
- (b)  $B$  und  $N$  bezeichnen stets Spaltenindexvektoren von  $A$ , wobei  $B = (p_1, \dots, p_m) \in \{1, \dots, n\}^m$  und  $N = (q_1, \dots, q_{n-m}) \in \{1, \dots, n\}^{n-m}$  gilt. Ferner kommt kein Spaltenindex, der in  $B$  vorkommt, auch in  $N$  vor, und umgekehrt. (Um Transpositionszeichen zu sparen, schreiben wir  $B$  und  $N$  immer als Zeilenvektoren.)
- (c) Wir werden aber  $B$  und  $N$  auch einfach als Mengen auffassen (wenn die Anordnung der Indizes keine Rolle spielt), dann gilt also nach  $B \cap N = \emptyset$ ,  $B \cup N = \{1, \dots, n\}$ . Insbesondere können wir dann  $i \in B$  und  $j \in N$  schreiben.
- (d) Zur schreibtechnischen Vereinfachung schreiben wir im folgenden  $A_B$  und  $A_N$  statt  $A_{.B}$  und  $A_{.N}$ .
- (e) Ist ein Spaltenindexvektor  $B$  wie oben angegeben definiert, so bezeichnet, falls nicht anders spezifiziert,  $N = (q_1, \dots, q_{n-m})$  immer den Spaltenindexvektor, der alle übrigen Spaltenindizes enthält, wobei  $q_i < q_j$  für  $i < j$  gilt.  $\triangle$

### 9.1 Basen, Basislösungen, Entartung

**(9.4) Definition.** Gegeben sei ein Gleichungssystem  $Ax = b$  mit  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$  und  $\text{rang}(A) = m$ . Spaltenindexvektoren  $B$  und  $N$  seien entsprechend Konvention (9.3) gegeben.

- (a) Ist  $A_B$  regulär, so heißt  $A_B$  Basismatrix oder kurz Basis von  $A$ , und  $A_N$  heißt Nichtbasismatrix oder Nichtbasis von  $A$ . Der Vektor  $x \in \mathbb{K}^n$  mit  $x_N = 0$  und  $x_B = A_B^{-1}b$  heißt Basislösung von  $Ax = b$  zur Basis  $A_B$  oder die zu  $A_B$  gehörige Basislösung oder kurz Basislösung.
- (b) Ist  $A_B$  eine Basis, dann heißt jede Variable  $x_j$ ,  $j \in B$ , Basisvariable und  $x_j$ ,  $j \in N$ , heißt Nichtbasisvariable.
- (c) Ist  $A_B$  eine Basis, dann heißen  $A_B$  und die zugehörige Basislösung  $x_B$  zulässig (bezüglich  $P^=(A, b)$ ), wenn  $x_B = A_B^{-1}b \geq 0$  gilt.
- (d) Eine Basislösung  $x$  zur Basis  $A_B$  heißt nichtentartet (oder nichtdegeneriert), falls  $x_B = A_B^{-1}b > 0$ , andernfalls heißt sie entartet (oder degeneriert).  $\triangle$

Die oben eingeführten Begriffe gehören zur Standardterminologie der linearen Programmierung. Eine erste begriffliche Brücke zur Polyedertheorie schlägt der nächste Satz.

**(9.5) Satz.** Seien  $P = P^=(A, b) \subseteq \mathbb{K}^n$  ein Polyeder mit  $\text{rang}(A) = m < n$  und  $x \in P$ . Dann sind äquivalent

- (1)  $x$  ist eine Ecke von  $P^=(A, b)$ .
- (2)  $x$  ist eine zulässige Basislösung (d. h. es gibt eine Basis  $A_B$  von  $A$  mit der Eigenschaft  $x_B = A_B^{-1}b \geq 0$  und  $x_N = 0$ ). △

**Beweis.** (1)  $\implies$  (2) Sei  $I := \text{supp}(x)$ . Ist  $x$  eine Ecke von  $P^=(A, b)$ , so sind die Spaltenvektoren  $A_{\cdot j}$ ,  $j \in I$ , nach (8.9) linear unabhängig. Wegen  $\text{rang}(A) = m$  gibt es eine Menge  $J \subseteq \{1, \dots, n\} \setminus I$ ,  $|J| = m - |I|$ , so dass die Spalten  $A_{\cdot j}$ ,  $j \in I \cup J$ , linear unabhängig sind. Folglich ist  $A_B$  mit  $B = I \cup J$  eine Basis von  $A$ . Nehmen wir o. B. d. A. an, dass  $B$  aus den ersten  $m$  Spalten von  $A$  besteht, also  $A = (A_B, A_N)$  mit  $N = \{m + 1, \dots, n\}$  gilt, dann erhalten wir aus  $x^T = (x_B^T, x_N^T) = (x_B^T, 0)$  folgendes:  $A_B^{-1}b = A_B^{-1}(Ax) = A_B^{-1}(A_B, A_N) \begin{pmatrix} x_B \\ 0 \end{pmatrix} = (I, A_B^{-1}A_N) \begin{pmatrix} x_B \\ 0 \end{pmatrix} = x_B$ . Da  $x \geq 0$  gilt, ist  $x$  eine zulässige Basislösung.

(2)  $\implies$  (1) folgt direkt aus (8.9). □

Damit sehen wir, dass Ecken von  $P^=(A, b)$  zulässigen Basislösungen von  $Ax = b$ ,  $x \geq 0$  entsprechen. Also gibt es zu jeder Ecke eine zulässige Basis von  $A$ . Sind zwei Ecken verschieden, so sind natürlich die zugehörigen Basen verschieden. Dies gilt aber nicht umgekehrt!

$$\text{Ecke} \xleftrightarrow{\text{eindeutig}} \text{zulässige Basislösung} \xleftrightarrow{\text{nicht eindeutig}} \text{zulässige Basis.}$$

Wir wollen nun Entartung und ihre Ursachen untersuchen und betrachten zunächst drei Beispiele.

**(9.6) Beispiel (Degeneration).**

- (a) Im  $\mathbb{K}^2$  gibt es keine „richtigen“ degenerierten Probleme, da man solche Probleme durch Entfernung von redundanten Ungleichungen nichtdegeneriert machen kann. Im folgenden Beispiel ist die Ungleichung  $2x_1 + x_2 \leq 2$  redundant.

$$\begin{aligned} x_1 + x_2 &\leq 1 \\ 2x_1 + x_2 &\leq 2 \\ x_1, x_2 &\geq 0 \end{aligned} \tag{9.7}$$

Das durch das Ungleichungssystem definierte Polyeder (siehe Abbildung 9.1) hat die drei Ecken  $E_1, E_2, E_3$ . Wir formen durch Einführung von Schlupfvariablen um:

$$\begin{aligned} x_1 + x_2 + s_1 &= 1 \\ 2x_1 + x_2 + s_2 &= 2 \\ x_1, x_2, s_1, s_2 &\geq 0. \end{aligned}$$

## 9 Die Grundversion des Simplex-Algorithmus

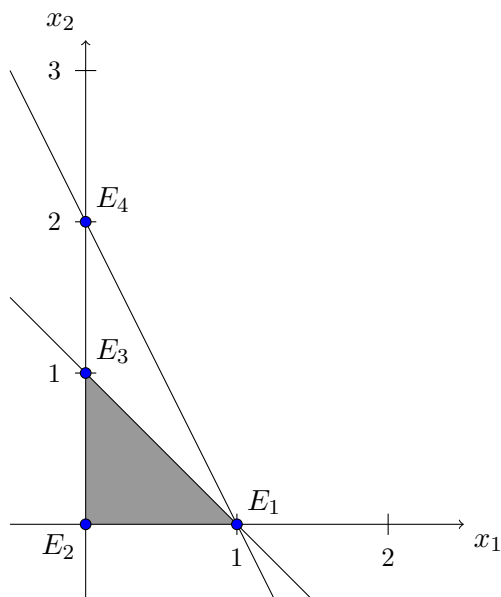


Abbildung 9.1: Grafische Darstellung der Ungleichungen (9.7).

Also ist  $P^=(A, b)$  gegeben durch:

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \quad (9.8)$$

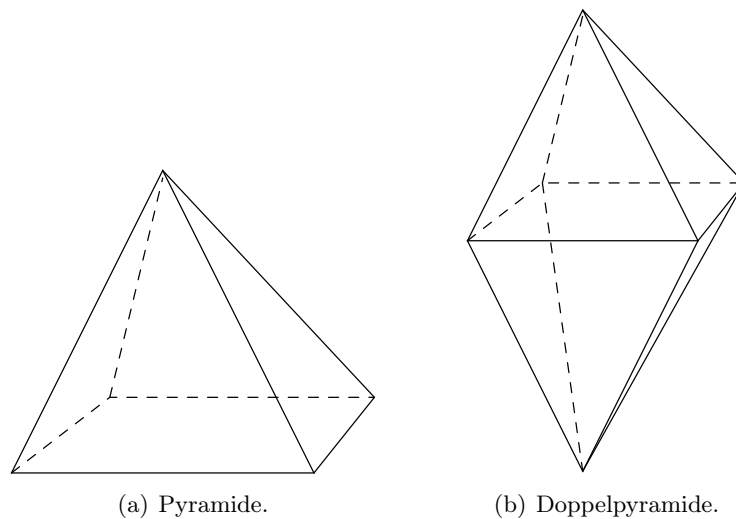
Tabelle 9.1 enthält die Basen von  $A$  und die zugehörigen Basislösungen.

Im obigen Falle ist jede  $(2, 2)$ -Untermatrix von  $A$  eine Basis. Dies ist natürlich nicht immer so! Zur Ecke  $E_1$  gehören drei verschiedene Basen. Hier liegt also Entartung vor.

Spalten in $B$	$A_B$	$x_B$	$x_N$	$x$	Ecke
1,2	$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$	$(x_1, x_2) = (1, 0)$	$(s_1, s_2)$	$(1, 0, 0, 0)$	$E_1$
1,3	$\begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}$	$(x_1, s_1) = (1, 0)$	$(x_2, s_2)$	$(1, 0, 0, 0)$	$E_1$
1,4	$\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$	$(x_1, s_2) = (1, 0)$	$(x_2, s_1)$	$(1, 0, 0, 0)$	$E_1$
2,3	$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$	$(x_2, s_1) = (2, -1)$	$(x_1, s_2)$	$(0, 2, -1, 0)$	$E_4$
2,4	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$	$(x_2, s_2) = (1, 1)$	$(x_1, s_1)$	$(0, 1, 0, 1)$	$E_3$
3,4	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$(s_1, s_2) = (1, 2)$	$(x_1, x_2)$	$(0, 0, 1, 2)$	$E_2$

Tabelle 9.1: Übersicht der Basen des Gleichungssystems (9.8).



Abbildung 9.2: Beispiele für Degeneration im  $\mathbb{R}^3$ .

Alle anderen Basislösungen sind nichtdegeneriert. Die zu  $B = (2, 3)$  gehörige Basis ist unzulässig. Zu unzulässigen Basislösungen sagt man manchmal auch *unzulässige Ecke* von  $P^=(A, b)$ ,  $E_4$  ist also eine unzulässige Ecke.

- (b) Bei dreidimensionalen Polyedern im  $\mathbb{K}^3$  tritt echte Entartung auf. So ist z. B. die Spitze der Pyramide in Abbildung 9.2(a) entartet, während alle übrigen vier Ecken nichtentartet sind.
- (c) Dagegen sind alle Ecken der Doppelpyramide in Abbildung 9.2(b) entartet.  $\triangle$

Die Sätze (9.5) und (8.9) (und ihre Beweise) sagen uns, wie Ecken von  $P^=(A, b)$  mit den Basen von  $A_B$  zusammenhängen. Aus der linearen Algebra wissen wir zunächst, dass jeder Punkt  $x$  im  $\mathbb{K}^n$  als Durchschnitt von  $n$  Hyperebenen dargestellt werden kann, deren Normalenvektoren linear unabhängig sind. Algebraisch ausgedrückt, jeder Punkt  $x \in \mathbb{K}^n$  ist eindeutig bestimmte Lösung eines regulären  $(n, n)$ -Gleichungssystems  $Dy = d$  (mit  $\text{rang}(D) = n$ ).

Ecken von  $P^=(A, b)$  erhalten wir dadurch, dass wir nur Lösungen von regulären Gleichungssystemen

$$\begin{aligned} Ay &= b \\ e_j^T y &= 0, \quad j \in J \end{aligned}$$

mit  $J \subseteq \{1, \dots, n\}$  betrachten (und prüfen, ob die Lösungen in  $P^=(A, b)$  enthalten sind). Dabei kann es zu folgenden Situationen kommen: Ist  $A_B$  eine Basis von  $A$  mit  $(A_B^{-1}b)_i \neq 0$  für alle  $i \in B$ , so gibt es nur eine einzige Möglichkeit das System  $Ay = b$  durch ein System  $e_j^T y = 0$ ,  $j \in J$ , so zu ergänzen, dass das Gesamtsystem regulär wird. Man muss  $J = \{1, \dots, n\} \setminus B$  wählen, andernfalls sind nicht genügend Gleichungen vorhanden. Daraus folgt speziell:

**(9.9) Bemerkung.** Ist  $x$  eine nichtdegenerierte Basislösung von  $Ay = b$ ,  $y \geq 0$ , dann gibt es eine eindeutig bestimmte zu  $x$  gehörige Basis  $A_B$  von  $A$  ( $x_B = A_B^{-1}b$ ,  $x_N = 0$ ). $\triangle$

Die Umkehrung gilt i. A. nicht, wie das folgende Beispiel zeigt.

**(9.10) Beispiel.**  $P = (A, b) \subseteq \mathbb{R}^3$  sei gegeben durch

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

Die Matrix  $A$  besitzt zwei Basen  $A_{B_1}$ ,  $A_{B_2}$  mit  $B_1 = (1, 3)$ ,  $B_2 = (2, 3)$ . Die zu  $A_{B_1}$  gehörige Basislösung ist  $x_1^T = (1, 0, 0)$ , die zu  $A_{B_2}$  gehörige Basislösung ist  $x_2^T = (0, 1, 0)$ . Beide Basislösungen sind entartet, aber die zugehörigen Basen sind eindeutig bestimmt. $\triangle$

Ist  $A_B$  eine Basis von  $A$  und  $x$  die zugehörige Basislösung, so dass einige Komponenten von  $A_B^{-1}b$  Null sind (d. h.  $x$  ist entartet), so enthält das Gleichungssystem  $Ay = b$ ,  $e_j^T y = 0$  für alle  $j \in J = \{j \mid x_j = 0\}$  mehr als  $n$  Gleichungen (also mehr als notwendig). I. A. ist dann  $x$  Lösung von mehreren regulären  $(n, n)$ -Untersystemen dieses Gleichungssystems. Man hat also keine eindeutige Zuordnung mehr von der Ecke  $x$  zu einem Gleichungssystem der Form  $Ay = b$ ,  $e_j^T y = 0$ ,  $j \in J$ . Wir werden sehen, dass Entartung zu rechentechnischen Problemen führt. Entartung kann drei Ursachen haben:

- überflüssige Variable (im Beispiel (9.10) kann  $x_3$  weggelassen werden, dann sind alle Basislösungen nichtentartet),
- redundante Ungleichungen (Beispiel (9.6)(a)),
- geometrische Gründe (Beispiele (9.6)(b) und (9.6)(c)).

Die durch überflüssige Variablen oder redundante Ungleichungen hervorgerufene Entartung kann i. A. beseitigt werden durch Weglassen der überflüssigen Variablen bzw. der redundanten Ungleichungen. Es gibt jedoch Polyeder – wie z. B. die Doppelpyramide in (9.6)(c) – die genuin entartet sind. Der Grund liegt hier darin, dass die Ecken überbestimmt sind, d. h. es gibt Ecken  $x$  von  $P$ , in denen mehr als  $\dim(P)$  Facetten zusammentreffen. Dann bestimmen je  $\dim(P)$  der diese Facetten definierenden Ungleichungen eine Basis von  $A$ , die  $x$  als zugehörige Basislösung liefert. Die linearen Programme, die als Relaxierungen von kombinatorischen Optimierungsproblemen entstehen, haben sehr häufig stark entartete Ecken, d. h. Basislösungen mit sehr vielen verschiedenen Basen, die diese Ecke bestimmen.

## 9.2 Basisaustausch (Pivoting), Simplexkriterium

Da jede zulässige Basis der Matrix  $A$  eine Ecke von  $P = (A, b)$  definiert, kann man ein lineares Programm der Form (9.1) dadurch lösen, dass man alle Basen von  $A$  bestimmt – dies sind endlich viele –, den Zielfunktionswert der zugehörigen Basislösung errechnet und die beste Lösung auswählt. Da eine  $(m, n)$ -Matrix bis zu  $\binom{n}{m}$  zulässige Basen haben

kann, ist dieses Verfahren aufgrund des gewaltigen Rechenaufwandes (in jedem Schritt Bestimmung einer inversen Matrix) so gut wie undurchführbar.

Man sollte also versuchen, z. B. ein Verfahren so zu entwerfen, dass man von einer zulässigen Basislösung ausgehend eine neue Basislösung bestimmt, die zulässig ist und einen besseren Zielfunktionswert hat. Ferner muss das Verfahren so angelegt sein, dass der Übergang von einer Basis zur nächsten nicht allzuviel Rechenaufwand erfordert.

Im Prinzip kann man damit nicht gewährleisten, dass nicht alle Basen enumeriert werden, aber aufgrund heuristischer Überlegungen scheint ein solcher Ansatz, wenn er realisiert werden kann, besser als das obige Enumerationsverfahren funktionieren zu können. Im weiteren werden wir zeigen, wie man einen Basisaustausch, der von einer zulässigen Basislösung zu einer besseren zulässigen Basislösung mit relativ wenig Rechenaufwand führt, durchführen kann.

**(9.11) Satz.** Gegeben sei ein Gleichungssystem  $Ay = b$  mit  $\text{rang}(A) = m$ , und  $A_B$  sei eine Basis von  $A$ . Dann gilt

$$x \text{ erfüllt } Ay = b \iff x \text{ erfüllt } x_B = A_B^{-1}b - A_B^{-1}A_Nx_N. \quad \triangle$$

**Beweis.** O. B. d. A. sei  $B = (1, \dots, m)$ , d. h.  $A = (A_B, A_N)$ ,  $x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$ , dann gilt

$$\begin{aligned} Ax = b &\iff (A_B, A_N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = b \iff A_B^{-1}(A_B, A_N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = A_B^{-1}b \\ &\iff (I, A_B^{-1}A_N) \begin{pmatrix} x_B \\ x_N \end{pmatrix} = A_B^{-1}b \\ &\iff x_B + A_B^{-1}A_Nx_N = A_B^{-1}b \\ &\iff x_B = A_B^{-1}b - A_B^{-1}A_Nx_N \quad \square \end{aligned}$$

Die obige Beziehung ist simpel, von rechentechnischer Bedeutung ist jedoch die Tatsache, dass wir die Basisvariablen  $x_j$ ,  $j \in B$ , in Abhängigkeit von den Nichtbasisvariablen darstellen können.

Wir wollen uns nun überlegen, wie wir den Übergang von einer Ecke zu einer anderen Ecke vollziehen können, d. h. wie wir aus einer Basis eine andere Basis ohne viel Aufwand konstruieren können. Um dies rechentechnisch exakt vorzuführen zu können, müssen wir die Indexmengen  $B$ ,  $N$  wie in (9.3)(b) angegeben als Vektoren auffassen.

**(9.12) Satz (Basisaustausch, Pivotoperation).** Gegeben sei ein Gleichungssystem  $Ay = b$  mit  $\text{rang}(A) = m$ .  $B = (p_1, \dots, p_m)$  und  $N = (q_1, \dots, q_{n-m})$  seien Spaltenindexvektoren von  $A$  wie in (9.3)(b) festgelegt, so dass  $A_B$  eine Basis von  $A$  ist. Wir setzen

$$\begin{aligned} \bar{A} &:= A_B^{-1}A_N = (\bar{a}_{rs})_{\substack{1 \leq r \leq m \\ 1 \leq s \leq n-m}}, \\ \bar{b} &:= A_B^{-1}b \in \mathbb{K}^m. \end{aligned}$$

## 9 Die Grundversion des Simplex-Algorithmus

Ist  $\bar{a}_{rs} \neq 0$ , so ist  $A_{B'}$  mit  $B' := (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$  eine Basis von  $A$ . Ferner gilt:

$$A_{B'}^{-1} = EA_B^{-1},$$

wobei  $E$  eine sogenannte  $(m, m)$ -Elementarmatrix ist, die wie folgt definiert ist:

$$E = \begin{pmatrix} 1 & & & \eta_1 & & & & \\ & \ddots & & \vdots & & & & \\ & & 1 & \eta_{r-1} & & & & \\ & & & \eta_r & & & & \\ & & & \eta_{r+1} & 1 & & & \\ & & & \vdots & & \ddots & & \\ & & & \eta_m & & & & 1 \end{pmatrix}.$$

Dabei ist die  $r$ -te Spalte von  $E$  der Vektor  $\eta = (\eta_1, \dots, \eta_m)^T$  (genannt Eta-Spalte) gegeben durch

$$\begin{aligned} \eta_r &:= \frac{1}{\bar{a}_{rs}} \\ \eta_i &:= -\frac{\bar{a}_{is}}{\bar{a}_{rs}}, \quad i \in \{1, \dots, m\} \setminus \{r\}. \end{aligned}$$

Das Element  $\bar{a}_{rs}$  heißt Pivot-Element.

Sind  $x, x'$  die zu  $A_B, A_{B'}$  gehörigen Basislösungen, so gilt:

$$x'_{p_i} = \bar{b}_i - \frac{\bar{a}_{is}}{\bar{a}_{rs}} \bar{b}_r = E_i \cdot \bar{b} = E_i \cdot x_B, \quad 1 \leq i \leq m, i \neq r \quad (\text{alte Basisvariablen})$$

$$x'_{q_s} = \frac{1}{\bar{a}_{rs}} \bar{b}_r (= E_r \cdot \bar{b}_r = E_r \cdot x_B) \quad (\text{neue Basisvariable})$$

$$x'_j = 0 \quad \text{andernfalls} \quad (\text{Nichtbasisvariablen}). \quad \triangle$$

**Beweis.** Es seien

$$d := A_{\cdot q_s}, \quad \bar{d} := A_B^{-1} d = \bar{A}_{\cdot s}, \quad F := A_B^{-1} A_{B'}.$$

$A_{B'}$  erhält man aus  $A_B$  dadurch, dass die  $r$ -te Spalte  $A_{\cdot p_r}$  von  $A_B$  durch  $d$  ersetzt wird.  $F$  ist daher eine  $(m, m)$ -Elementarmatrix, deren  $r$ -te Spalte der Vektor  $\bar{d}$  ist. Offenbar gilt daher

$$F \cdot E = \begin{pmatrix} 1 & & & \bar{d}_1 & & & & \\ & \ddots & & \vdots & & & & \\ & & 1 & \bar{d}_{r-1} & & & & \\ & & & \bar{d}_r & & & & \\ & & & \bar{d}_{r+1} & 1 & & & \\ & & & \vdots & & \ddots & & \\ & & & \bar{d}_m & & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \eta_1 & & & & \\ & \ddots & & \vdots & & & & \\ & & 1 & \eta_{r-1} & & & & \\ & & & \eta_r & & & & \\ & & & \eta_{r+1} & 1 & & & \\ & & & \vdots & & \ddots & & \\ & & & \eta_m & & & & 1 \end{pmatrix} = I,$$

daraus folgt  $E = F^{-1}$  ( $\bar{a}_r \neq 0$  war vorausgesetzt). Da  $F$  und  $A_B$  invertierbar sind, ergibt sich aus  $A_{B'} = A_B F$  sofort  $A_{B'}^{-1} = F^{-1} A_B^{-1} = E A_B^{-1}$ . Die übrigen Formeln ergeben sich durch einfache Rechnung.  $\square$

Bei dem in Satz (9.12) beschriebenen Basisaustausch wird also von einer Basis zu einer anderen Basis übergegangen, indem eine Nichtbasisvariable  $x_{q_s}$  und eine Basisvariable  $x_{p_r}$  gesucht werden, so dass  $\bar{a}_{r_s} \neq 0$  gilt. Wenn diese Bedingung erfüllt ist, kann die Nichtbasisvariable  $x_{q_s}$  zu einer Basisvariablen gemacht werden, wobei  $x_{p_r}$  nichtbasisch wird. Alle übrigen Basisvariablen bleiben erhalten.

**(9.13) Beispiel.** Wir betrachten das Gleichungssystem  $Ay = b$  gegeben durch

$$A = \begin{pmatrix} 1 & 0 & 2 & -\frac{3}{2} & -4 & \frac{11}{2} & \frac{11}{2} \\ 0 & 1 & 0 & 0 & 3 & 0 & 2 \\ 0 & 0 & -1 & \frac{1}{2} & 2 & -\frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 2 \end{pmatrix}.$$

Es seien  $B = (1, 2, 3, 4)$ ,  $N = (5, 6, 7)$ , dann ist  $A_B$  eine Basis von  $A$ , und es gilt

$$A_B^{-1} = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad \bar{A} = A_B^{-1} A_N = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 0 & 2 \\ -1 & 1 & 0 \\ 2 & -1 & -1 \end{pmatrix}.$$

Die zugehörige Basislösung ist  $x_B^T = (1, 2, 3, 4)$ ,  $x_N^T = (0, 0, 0)$ , d. h.  $x^T = (1, 2, 3, 4, 0, 0, 0)$ . Das Element  $\bar{a}_{23} = 2$  ist von Null verschieden. Wir können es also als Pivotelement  $\bar{a}_{rs}$  ( $r = 2, s = 3$ ) wählen. Daraus ergibt sich:

$$B' = (1, 7, 3, 4), \quad N' = (5, 6, 2),$$

$$E = \begin{pmatrix} 1 & -2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & 1 \end{pmatrix}, \quad EA_B^{-1} = A_{B'}^{-1} = \begin{pmatrix} 1 & -2 & 2 & 1 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & \frac{1}{2} & 0 & 2 \end{pmatrix}$$

und

$$A_{B'}^{-1} A = EA_B^{-1} A = \begin{pmatrix} 1 & -2 & 0 & 0 & -5 & 2 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & 1 & \frac{7}{2} & -1 & 0 \end{pmatrix}, \quad \bar{A}' = \begin{pmatrix} -5 & 2 & -2 \\ \frac{3}{2} & 0 & \frac{1}{2} \\ -1 & 1 & 0 \\ \frac{7}{2} & -1 & \frac{1}{2} \end{pmatrix}.$$

Die zugehörige Basislösung ist  $x_{B'}^T = (-3, 1, 3, 5)$ ,  $x_{N'}^T = (0, 0, 0)$ , daraus ergibt sich  $x^T = (-3, 0, 3, 5, 0, 0, 1)$ .  $\triangle$

**(9.14) Bemerkung.** Die neue Basisinverse  $A_{B'}^{-1}$  kann also wegen  $A_{B'}^{-1} = EA_B^{-1}$  leicht aus  $A_B^{-1}$  berechnet werden. Tatsächlich ist jedoch dieses „Pivoting“ der numerisch aufwendigste und schwierigste Schritt im Simplexverfahren. Es gibt hier zahllose „Update“-Varianten, durch die eine schnelle und numerisch stabile Berechnung von  $A_{B'}^{-1}$  erreicht werden soll. Insbesondere für große und dünn besetzte Matrizen ( $n, m \geq 100000$ , large-scale-linear-programming) gibt es spezielle Techniken.  $\triangle$

## 9 Die Grundversion des Simplex-Algorithmus

Die äquivalente Darstellung des Gleichungssystems  $Ax = b$  in Bemerkung (9.11) gibt uns ebenso die Möglichkeit, die Kosten auf einfache Weise über die Nichtbasisvariablen zu berechnen, woraus sich ein sehr einfaches Optimalitätskriterium gewinnen lässt.

**(9.15) Satz.** Gegeben sei ein lineares Programm in Standardform (9.1), und  $A_B$  sei eine Basis von  $A$ .

(a) Für den Zielfunktionswert  $c^T x$  von  $x \in P^=(A, b)$  gilt

$$c^T x = c_B^T A_B^{-1} b + (c_N^T - c_B^T A_B^{-1} A_N) x_N.$$

Der Term  $\bar{c}^T := c_N^T - c_B^T A_B^{-1} A_N$  heißt reduzierte Kosten (von  $x$ ), die Komponenten  $\bar{c}_i$  von  $\bar{c}$  heißen reduzierte Kostenkoeffizienten.

(b) (Simplexkriterium)

Ist  $A_B$  eine zulässige Basis und sind die reduzierten Kosten nicht-positiv, d. h.

$$\bar{c}^T = c_N^T - c_B^T A_B^{-1} A_N \leq 0,$$

dann ist die zugehörige Basislösung  $x$  mit  $x_B = A_B^{-1} b$ ,  $x_N = 0$  optimal für (9.1).

(c) Ist  $A_B$  eine zulässige Basis, und ist die zugehörige Basislösung nichtdegeneriert und optimal, dann gilt  $\bar{c} \leq 0$ . △

**Beweis.** (a) Nach Satz (9.11) gilt  $Ax = b \iff x_B = A_B^{-1} b - A_B^{-1} A_N x_N$  und damit gilt

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

(b) Sei  $y \in P^=(A, b)$  beliebig. Wenn wir zeigen können, dass  $c^T x \geq c^T y$  gilt, sind wir fertig.

$$c^T y = c_B^T A_B^{-1} b + \underbrace{\bar{c}^T}_{\leq 0} \underbrace{y_N}_{\geq 0} \leq c_B^T A_B^{-1} b = c_B^T x_B = c^T x.$$

(c) Sei die Basislösung  $x$  mit  $x_B = A_B^{-1} b$ ,  $x_N = 0$  optimal. Dann gilt für alle  $y \in P^=(A, b)$ :

$$\begin{aligned} c^T x \geq c^T y &\iff c_B^T A_B^{-1} b + \bar{c}^T x_N \geq c_B^T A_B^{-1} b + \bar{c}^T y_N \\ &\iff 0 \geq \bar{c}^T y_N. \end{aligned}$$

Angenommen die  $i$ -te Komponente  $\bar{c}_i$  von  $\bar{c}^T$  wäre größer als Null. Nach Voraussetzung ist  $x$  nichtdegeneriert, also  $A_B^{-1} b > 0$ . Sei  $e_i$  der  $i$ -te Einheitsvektor in  $\mathbb{K}^{n-m}$ , dann gibt es somit ein  $\lambda > 0$  mit

$$A_B^{-1} b \geq A_B^{-1} A_N \lambda e_i.$$

Der Vektor  $x^\lambda$  mit  $x_B^\lambda = A_B^{-1} b - A_B^{-1} A_N \lambda e_i$ ,  $x_N^\lambda = \lambda e_i$  ist somit ein für (9.1) zulässiger Vektor, und es gilt

$$c^T x^\lambda = c_B^T A_B^{-1} b + \bar{c}^T \lambda e_i = c^T x + \lambda \bar{c}_i > c^T x.$$

Widerspruch! □

Aus Satz (9.12) wissen wir, wie ein Basis- bzw. Eckenaustausch vorgenommen werden kann, Satz (9.15) besagt, unter welchen Umständen wir eine zulässige Basis verbessern können, bzw. wann sie optimal ist. Setzen wir diese Informationen zusammen, so gilt:

**(9.16) Satz (Basisverbesserung).** Gegeben sei ein lineares Programm in Standardform (9.1). Sei  $A_B$  eine zulässige Basis mit Basislösung  $x$ . Sei  $\bar{A} = A_B^{-1}A_N$ ,  $\bar{b} = A_B^{-1}b$ , und  $\bar{c}^T = c_N^T - c_B^T A_B^{-1}A_N$  seien die reduzierten Kosten. Sei  $q_s \in N$  ein Index mit  $\bar{c}_s > 0$ , dann gilt

(a) Ist  $\bar{A}_{.s} \leq 0$ , dann ist  $c^T x$  auf  $P^=(A, b)$  unbeschränkt.

(b) Ist  $\bar{A}_{.s} \not\leq 0$ , dann setzen wir

$$\lambda_0 := \min \left\{ \frac{\bar{b}_i}{\bar{a}_{is}} \mid i = 1, \dots, m, \bar{a}_{is} > 0 \right\},$$

und wählen einen Index

$$r \in \left\{ i \in \{1, \dots, m\} \mid \frac{\bar{b}_i}{\bar{a}_{is}} = \lambda_0 \right\}.$$

Dann ist  $A_{B'}$  mit  $B' = (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$  eine zulässige Basis mit Basislösung  $x'$ , so dass  $c^T x' \geq c^T x$  gilt.

(c) Gelten die Voraussetzungen von (b), und ist  $A_B$  nichtdegeneriert, dann gilt  $c^T x' > c^T x$ . △

**Beweis.** (a) Nach (9.11) gilt:  $y \in P^=(A, b) \iff y_B = A_B^{-1}b - A_B^{-1}A_N y_N \geq 0$  und  $y_N \geq 0$ . Für beliebiges  $\lambda \geq 0$  ist daher der Vektor  $y^\lambda \in \mathbb{K}^n$  mit  $y_N^\lambda := \lambda e_s$  und  $y_B^\lambda := A_B^{-1}b - \bar{A} y_N^\lambda = A_B^{-1}b - \lambda \bar{A}_{.s}$  wegen  $e_s \geq 0$  und  $A_B^{-1}b - \lambda \bar{A}_{.s} \geq A_B^{-1}b \geq 0$  in  $P^=(A, b)$ . Da  $c^T y^\lambda = c_B^T A_B^{-1}b + \bar{c}^T y_N^\lambda = c_B^T A_B^{-1}b + \lambda \bar{c}_s$  mit  $\lambda$  über alle Schranken wächst, ist  $c^T x$  auf  $P^=(A, b)$  unbeschränkt.

(b) Wählen wir  $r$  und  $s$  wie im Satz angegeben, so ist  $\bar{a}_{rs} > 0$ . Nach Satz (9.12) ist dann  $B' = (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$  eine Basis mit Basislösung  $x'$ , wobei

$$x'_{p_i} = \bar{b}_i - \frac{\bar{b}_r}{\bar{a}_{rs}} \bar{a}_{is} \begin{cases} \geq \bar{b}_i - \frac{\bar{b}_i}{\bar{a}_{is}} \bar{a}_{is}, & \text{falls } \bar{a}_{is} > 0, i \neq r \quad (\text{Wahl von } \lambda_0!), \\ \geq \bar{b}_i, & \text{falls } \bar{a}_{is} \leq 0, \end{cases}$$

$$x'_{q_s} = \frac{\bar{b}_r}{\bar{a}_{rs}} \geq 0,$$

$$x'_i = 0 \quad \text{andernfalls.}$$

Also ist  $x'$  eine zulässige Basislösung.

Wir zeigen nun  $c^T x' \geq c^T x$ . Zunächst gilt für den in Satz (9.12) definierten Vektor  $\eta$ :

$$c_{B'}^T \eta - c_{p_r} = \frac{\bar{c}_s}{\bar{a}_{rs}} > 0,$$

## 9 Die Grundversion des Simplex-Algorithmus

denn

$$\begin{aligned} 0 < \bar{c}_s &= (c_N^T - c_B^T A_B^{-1} A_N)_s = c_{q_s} - c_B^T \bar{A}_{\cdot s} = c_{q_s} - \sum_{i=1}^m c_{p_i} \bar{a}_{is} \\ &= \left( c_{q_s} - \sum_{i \neq r} c_{p_i} \bar{a}_{is} \right) - c_{p_r} \bar{a}_{rs} = \bar{a}_{rs} (c_{B'}^T \eta - c_{p_r}), \end{aligned}$$

und aus  $\bar{a}_{rs} > 0$  folgt nun die Behauptung. Somit gilt:

$$\begin{aligned} c^T x' &= c_{B'}^T x_{B'} = c_{B'}^T A_{B'}^{-1} b \stackrel{(9.12)}{=} c_{B'}^T E A_B^{-1} b = c_{B'}^T E x_B \\ &= \sum_{i \neq r} c_{p_i} x_{p_i} + c_{B'}^T \eta x_{p_r} \\ &= c_B^T x_B + (c_{B'}^T \eta - c_{p_r}) x_{p_r} \\ &\geq c_B^T x_B = c^T x. \end{aligned}$$

Hieraus folgt (b).

(c) Ist  $x$  nichtdegeneriert, so gilt  $x_{p_r} > 0$ , und somit ist die letzte Ungleichung in der obigen Abschätzung strikt. Daraus ergibt sich (c).  $\square$

Der obige Beweis macht den geometrischen Inhalt des Satzes (9.16) nicht deutlich. Dem Leser wird dringend empfohlen, sich zu überlegen, welche geometrische Bedeutung der Parameter  $\lambda_0$  in (b) hat.

### 9.3 Das Simplexverfahren

Die Grundidee des Simplexverfahrens haben wir schon erläutert, sie kann kurz wie folgt beschrieben werden:

- Finde eine zulässige Basis  $A_B$ .
- Konstruiere aus  $A_B$  eine zulässige Basis  $A_{B'}$ , so dass die zulässige Basislösung von  $A_{B'}$  besser als die von  $A_B$  ist.
- Gibt es keine bessere Basislösung, so ist die letzte optimal.

Die vorhergehenden Sätze geben uns nun die Möglichkeit, die oben informell beschriebene Idee, analytisch darzustellen und algorithmisch zu realisieren. Die Grundversion des Simplexverfahrens lautet dann wie folgt:

#### (9.17) Grundversion des Simplexverfahrens.

**Eingabe:**  $A' \in \mathbb{K}^{(m',n)}$ ,  $b' \in \mathbb{K}^{m'}$ ,  $c \in \mathbb{K}^n$ .



**Ausgabe:** Eine optimale Lösung  $x$  des linearen Programms

$$\begin{aligned} \max \quad & c^T x \\ & A'x = b' \\ & x \geq 0 \end{aligned} \tag{9.18}$$

falls eine solche existiert. Andernfalls stellt das Verfahren fest, dass (9.18) unbeschränkt ist oder keine zulässige Lösung besitzt.

**Phase I:** Test auf Zulässigkeit, Bestimmung einer zulässigen Basislösung

Bestimme ein Subsystem  $Ax = b, x \geq 0$  von  $A'x = b', x \geq 0$ , so dass  $P=(A, b) = P=(A', b')$  gilt und die Zusatzvoraussetzungen

$$\text{rang}(A) = \text{Zeilenzahl von } A < \text{Spaltenzahl von } A$$

erfüllt sind (falls möglich). Sei  $\text{rang}(A) = m < n$ . Bestimme weiterhin eine zulässige Basis, d. h. finde  $B = (p_1, \dots, p_m) \in \{1, \dots, n\}^m$ , so dass  $A_B$  eine zulässige Basis von  $A$  ist, und sei  $N = (q_1, \dots, q_{n-m})$  wie üblich definiert. Berechne

$$\begin{aligned} & A_B^{-1} \\ \bar{A} &= A_B^{-1} A_N, \\ \bar{b} &= A_B^{-1} b, \\ \bar{c}^T &= c_N^T - c_B^T A_B^{-1} A_N, \\ c_0 &= c_B^T \bar{b}. \end{aligned}$$

Wie die gesamte Phase I durchgeführt wird und wie zu verfahren ist, wenn die gewünschten Resultate nicht erzielt werden können, wird in (9.38) angegeben. Im weiteren gehen wir davon aus, dass Phase I erfolgreich war.

**Phase II:** Optimierung

(II.1) Optimalitätsprüfung

Gilt  $\bar{c}_i \leq 0$  für  $i = 1, \dots, n - m$ , so ist die gegenwärtige Basislösung optimal (siehe (9.15)(b)). Gib den Vektor  $x$  mit  $x_B = \bar{b}, x_N = 0$  aus. Der Optimalwert von (9.1) ist  $c^T x = c_B^T \bar{b} = c_0$ . STOP!

(II.2) Bestimmung der Austausch- oder Pivotspalte

Wähle einen Index  $s \in \{1, \dots, n - m\}$ , so dass  $\bar{c}_s > 0$  gilt. (Zur konkreten Realisierung dieser Auswahl, falls mehrere reduzierte Kostenkoeffizienten positiv sind, gibt es sehr viele verschiedene Varianten, die wir in Abschnitt 9.4 besprechen werden.)

(II.3) Prüfung auf Beschränktheit des Optimums

Gilt  $\bar{A}_{\cdot s} \leq 0$ , so ist das lineare Programm unbeschränkt (siehe (9.16)(a)), STOP!

## 9 Die Grundversion des Simplex-Algorithmus

(II.4) Bestimmung der Austausch- oder Pivotzeile

Berechne

$$\lambda_0 := \min \left\{ \frac{\bar{b}_i}{\bar{a}_{is}} \mid \bar{a}_{is} > 0, i = 1, \dots, m \right\}$$

Wähle einen Index  $r \in \{1, \dots, m\}$ , so dass gilt

$$\frac{\bar{b}_r}{\bar{a}_{rs}} = \lambda_0.$$

(Hierzu gibt es verschiedene Möglichkeiten, die wir in Abschnitt 9.4 erläutern werden.)

(II.5) Pivotoperation, Basisaustausch

Setze

$$\begin{aligned} B' &:= (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m), \\ N' &:= (q_1, \dots, q_{s-1}, p_r, q_{s+1}, \dots, q_{n-m}), \\ A_{B'}^{-1} &:= EA_B^{-1} \quad (\text{siehe (9.12)}). \end{aligned}$$

(II.6) Updating

Berechne  $\bar{A}$ ,  $\bar{b}$ ,  $\bar{c}$ , und  $c_0$  neu!

Hierzu gibt es viele numerische Varianten. Wir geben nachfolgend ein didaktisch klares Verfahren an, das aber numerisch umständlich und höchstens für Handrechnung und kleine lineare Programme geeignet ist. Bessere Methoden werden in den Übungen besprochen.

(a) Neuberechnung von  $\bar{A}$ :

$$\text{Setze } \bar{\bar{a}}_{rs} := \frac{1}{\bar{a}_{rs}}.$$

Für  $i = 1, \dots, m, i \neq r$  führe aus

$$\bar{\bar{a}}_{is} := -\bar{\bar{a}}_{rs}\bar{a}_{is}.$$

Für  $j = 1, \dots, n - m, j \neq s$  führe aus

$$\bar{\bar{a}}_{rj} := \bar{\bar{a}}_{rs}\bar{a}_{rj}.$$

Für  $j = 1, \dots, n - m, j \neq s$ , und  $i = 1, \dots, m, i \neq r$  führe aus

$$\bar{\bar{a}}_{ij} := \bar{a}_{ij} - \frac{\bar{a}_{is}}{\bar{a}_{rs}}\bar{a}_{rj} = \bar{a}_{ij} + \bar{\bar{a}}_{is}\bar{\bar{a}}_{rj}.$$

(b) Neuberechnung von  $\bar{b}$ :

$$\bar{\bar{b}}_r := \frac{\bar{b}_r}{\bar{a}_{rs}}.$$

Für  $i = 1, \dots, m, i \neq r$  führe aus

$$\bar{\bar{b}}_i := \bar{b}_i - \frac{\bar{a}_{is}}{\bar{a}_{rs}}\bar{\bar{b}}_r.$$

(c) Neuberechnung von  $\bar{c}$  und  $c_0$ :

Setze für  $j = 1, \dots, n - m, j \neq s$ :

$$\bar{c}_j := \bar{c}_j - \frac{\bar{a}_{rj}}{\bar{a}_{rs}} \bar{c}_s, \quad \bar{c}_s := -\frac{\bar{c}_s}{\bar{a}_{rs}}, \quad c_0 := c_0 + \bar{c}_s \frac{\bar{b}_r}{\bar{a}_{rs}}.$$

(d) Setze  $\bar{A} := \bar{A}, \bar{b} := \bar{b}, \bar{c} := \bar{c}$  und gehe zu (II.1). △

**(9.19) Bemerkung (Die Tableauform der Simplexmethode).** Zur Veranschaulichung der Simplexmethode (nicht zur Verwendung in kommerziellen Codes) benutzt man manchmal ein Tableau  $T$ . Das Tableau entspricht dem Gleichungssystem

$$\begin{pmatrix} c^T \\ A \end{pmatrix} x = \begin{pmatrix} c^T x \\ b \end{pmatrix}$$

bzw. bei gegebener Basis  $A_B$  von  $A$  und geeigneter Permutation der Spalten dem System

$$\begin{array}{|c|c|} \hline c_N^T - c_B^T A_B^{-1} A_N & 0 \\ \hline A_B^{-1} A_N & I \\ \hline \end{array} \begin{pmatrix} x_N \\ x_B \end{pmatrix} = \begin{array}{|c|} \hline c^T x - c_B^T A_B^{-1} b \\ \hline A_B^{-1} b \\ \hline \end{array} \quad \triangle$$

**(9.20) Definition.** Ist  $\max\{c^T x \mid Ax = b, x \geq 0\}$  ein LP in Standardform und ist  $A_B$  eine zulässige Basis von  $A$ , so heißt die  $(m + 1, n + 1)$ -Matrix

$$T_B := \begin{pmatrix} c^T - c_B^T A_B^{-1} A & -c_B^T A_B^{-1} b \\ A_B^{-1} A & A_B^{-1} b \end{pmatrix}$$

ein Simplextableau zur Basis  $B$  mit Zeilen  $i = 0, \dots, m$  und Spalten  $j = 1, \dots, n + 1$ . △

Da ein Simplextableau stets eine volle  $(m, m)$ -Einheitsmatrix enthält, ist eine solche Darstellung für den Rechner redundant. Zum Kennenlernen der Simplexmethode und zur Übung der Rechentechnik ist sie allerdings sehr nützlich.

**(9.21) Satz (Update-Formeln für ein Simplextableau).** Sei  $T_B$  ein Simplextableau zur Basis  $B$ . Ist  $q_s \in N$ , so daß  $B' = (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$  wieder eine zulässige Basis ist, so gelten für das Simplextableau  $T_{B'}$  folgende Transformationsregeln

$$\begin{aligned} (T_{B'})_{i.} &= (T_B)_{i.} - \frac{\bar{a}_{is}}{\bar{a}_{rs}} (T_B)_{r.} \quad \text{für } i = 0, \dots, m, i \neq r, \\ (T_{B'})_{r.} &= \frac{1}{\bar{a}_{rs}} (T_B)_{r.} \end{aligned} \tag{a}$$

oder

$$T_{B'} = E' \cdot T_B \tag{b}$$

## 9 Die Grundversion des Simplex-Algorithmus

mit

$$E' = \begin{pmatrix} 1 & & & \eta'_0 & & & & & \\ & \ddots & & \vdots & & & & & \\ & & 1 & \eta'_{r-1} & & & & & \\ & & & \eta'_r & & & & & \\ & & & \eta'_{r+1} & 1 & & & & \\ & & & \vdots & & \ddots & & & \\ & & & \eta'_m & & & & & 1 \end{pmatrix}, \quad \begin{aligned} \eta'_i &= -\frac{(T_B)_{is}}{\bar{a}_{rs}}, \quad i \neq r \\ \eta'_r &= \frac{1}{\bar{a}_{rs}} \end{aligned}$$

(Die Transformationsformeln (9.21) sind im wesentlichen die gleichen wie in Satz (9.12), nur dass diesmal die Zielfunktion gleich mittransformiert wird.)  $\triangle$

**Beweis.** Formel (a):

$$T_{B'} := \begin{pmatrix} c^T - c_{B'}^T A_{B'}^{-1} A & -c_{B'}^T A_{B'}^{-1} b \\ A_{B'}^{-1} A & A_{B'}^{-1} b \end{pmatrix},$$

d. h. mit Satz (9.12) gilt  $(A_{B'}^{-1} = EA_B^{-1})$ :

$$T_{B'} := \begin{pmatrix} (c^T, 0) - c_{B'}^T E \cdot (A_B^{-1} A, A_B^{-1} b) \\ E \cdot (T_B)_{\{1, \dots, m\}} \end{pmatrix}.$$

Sei  $K := \{1, \dots, m\}$ . Aus  $(T_{B'})_K = E \cdot (T_B)_K$  ergeben sich die beiden unteren Transformationsformeln in (a). Sei  $z \in \mathbb{K}^m$  und  $y = A_B^{-1} z$ , dann gilt

$$c_{B'}^T A_{B'}^{-1} z = c_{B'}^T E A_B^{-1} z = c_{B'}^T E y = c_{B'}^T y + (c_{B'}^T \eta - c_{p_r}) y_r$$

(mit  $\eta$  aus Satz (9.12)). Mit (9.15) folgt nun  $c_{B'}^T \eta - c_{p_r} = \frac{\bar{c}_s}{\bar{a}_{rs}} = \frac{(T_B)_{0s}}{\bar{a}_{rs}}$ . Also gilt:

$$\begin{aligned} (c^T, 0) - c_{B'}^T A_{B'}^{-1} (A, b) &= (c^T, 0) - c_{B'}^T A_B^{-1} (A, b) - (T_B)_{0s} \bar{a}_{rs} (A_B^{-1} (A, b))_r \\ &= (T_B)_0 - \frac{(T_B)_{0s}}{\bar{a}_{rs}} (T_B)_r. \end{aligned}$$

Die Formeln (b) ergeben sich direkt aus (a).  $\square$

**(9.22) Beispiel.** Wir betrachten das lineare Programm

$$\begin{aligned} \max \quad & x_1 + 2x_2 \\ & x_1 \leq 4 \\ & 2x_1 + x_2 \leq 10 \\ & -x_1 + x_2 \leq 5 \\ & 3x_2 \leq 20 \\ & x_1, x_2 \geq 0, \end{aligned}$$

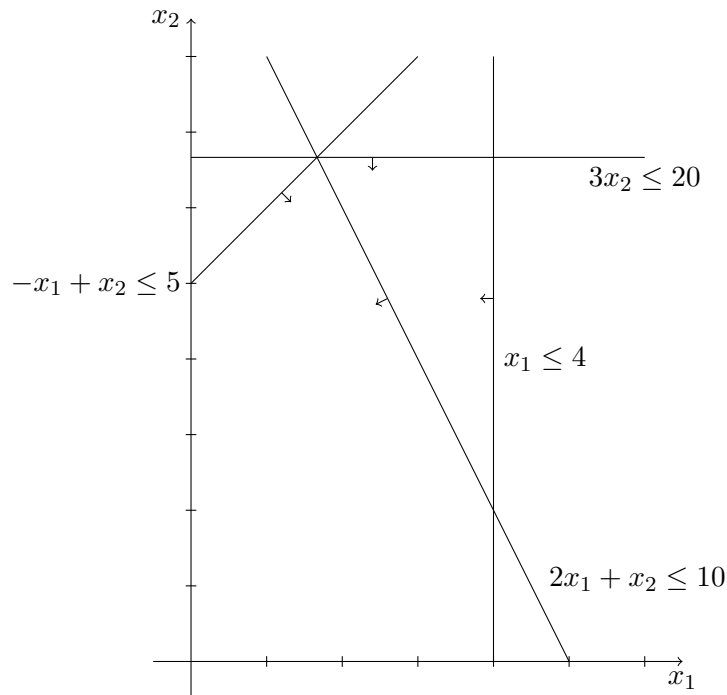


Abbildung 9.3: Zulässige Menge des LPs aus Beispiel (9.22).

dessen Lösungsmenge in Abbildung 9.3 graphisch dargestellt ist.

Wir führen Schlupfvariablen  $x_3, x_4, x_5, x_6$  ein und transformieren unser LP in folgende Form:

$$\max \quad (1, 2, 0, 0, 0) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 3 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \\ 5 \\ 20 \end{pmatrix}$$

$$x_i \geq 0 \quad i = 1, \dots, 6.$$

Die Matrix  $A$  des Gleichungssystems des transformierten Systems hat 4 Zeilen und 6 Spalten und enthält eine Einheitsmatrix. Folglich hat sie vollen Zeilenrang.

## 9 Die Grundversion des Simplex-Algorithmus

Daraus folgt, dass die Spalten 3, 4, 5, 6 eine Basis von  $A$  definieren, und diese ist zulässig, da die zugehörige Basislösung  $x_3 = 4$ ,  $x_4 = 10$ ,  $x_5 = 5$ ,  $x_6 = 20$ ,  $x_1 = 0$ ,  $x_2 = 0$  nichtnegativ ist. Diese Basislösung entspricht übrigens der Ecke  $x_1 = 0$ ,  $x_2 = 0$  des ursprünglichen Problems. Wir stellen nun das zur Anfangsbasis gehörige Tableau auf und führen den Simplexalgorithmus durch. Die Anfangsbasis ist gegeben durch  $B = (3, 4, 5, 6)$ , und somit ist  $N = (1, 2)$ .

$$T_0 : \begin{array}{cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\ & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ x_3 & 1 & 0 & 1 & 0 & 0 & 0 & 4 \\ x_4 & 2 & 1 & 0 & 1 & 0 & 0 & 10 \\ x_5 & -1 & \boxed{1} & 0 & 0 & 1 & 0 & 5 \\ x_6 & 0 & 3 & 0 & 0 & 0 & 1 & 20 \end{array} \quad a_{rs} = \bar{a}_{32} = 1$$

Spaltenauswahl: Steilster Anstieg, d. h. der größte reduzierte Kostenkoeffizient wird gewählt. Das Pivotelement ist:  $\bar{a}_{rs} = \bar{a}_{32} = 1$ .

$$T_1 : \begin{array}{cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\ & 3 & 0 & 0 & 0 & -2 & 0 & -10 \\ x_3 & 1 & 0 & 1 & 0 & 0 & 0 & 4 \\ x_4 & \boxed{3} & 0 & 0 & 1 & -1 & 0 & 5 \\ x_2 & -1 & 1 & 0 & 0 & 1 & 0 & 5 \\ x_6 & \boxed{3} & 0 & 0 & 0 & -3 & 1 & 5 \end{array} \quad \begin{array}{l} a_{rs} = \bar{a}_{21} = 3 \quad \text{oder} \\ a_{rs} = \bar{a}_{41} = 3 \end{array}$$

Im Tableau  $T_1$  ist die Wahl des Pivotelements nicht eindeutig, da das minimale  $\lambda_0$  in zwei Zeilen angenommen wird. Entscheiden wir uns für  $a_{rs} = \bar{a}_{41} = 3$ , so ergibt sich als nächstes Tableau

$$T_2 : \begin{array}{cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\ & 0 & 0 & 0 & 0 & 1 & -1 & -15 \\ x_3 & 0 & 0 & 1 & 0 & 1 & -1/3 & 7/3 \\ x_4 & 0 & 0 & 0 & 1 & \boxed{2} & 0 & 0 \\ x_2 & 0 & 1 & 0 & 0 & 0 & 1/3 & 20/3 \\ x_1 & 1 & 0 & 0 & 0 & -1 & 1/3 & 5/3 \end{array} \quad a_{rs} = \bar{a}_{25} = 3$$

Weiteres Pivotisieren führt zu

$$T_3 : \begin{array}{cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\ & 0 & 0 & 0 & -1/2 & 0 & -1/3 & -15 \\ x_3 & 0 & 0 & 1 & -1/2 & 0 & -1/3 & 7/3 \\ x_5 & 0 & 0 & 0 & 1/2 & 1 & 0 & 0 \\ x_2 & 0 & 1 & 0 & 0 & 0 & 1/3 & 20/3 \\ x_1 & 1 & 0 & 0 & 1/2 & 0 & 1/3 & 5/3 \end{array}$$

Das Tableau  $T_3$  ist optimal. Die optimale Lösung ist  $(\frac{5}{3}, \frac{20}{3}, \frac{7}{3}, 0, 0, 0)$  bzw.  $x_1 = \frac{5}{3}$ ,  $x_2 = \frac{20}{3}$ . Wenn wir für Tableau  $T_1$  das Pivotelement  $a_{rs} = \bar{a}_{21} = 3$  wählen, ergibt sich Tableau

$T'_2$ , das optimal ist:

$$T'_2 : \begin{array}{cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & \\ & 0 & 0 & 0 & -1 & -1 & 0 & -15 \\ \hline x_3 & 0 & 0 & 1 & -1/3 & 1/3 & 0 & 7/3 \\ x_1 & 1 & 0 & 0 & 1/3 & -1/3 & 0 & 5/3 \\ x_2 & 0 & 1 & 0 & 1/3 & 2/3 & 0 & 20/3 \\ x_6 & 0 & 0 & 0 & -1 & -2 & 1 & 0 \end{array}$$

Wir sehen an diesem Beispiel, dass es lästig ist, immer die Einheitsmatrix im Tableau mitzuschleppen. Wenn wir uns die gegenwärtigen Basis- und Nichtbasisvariablen geeignet (z. B. am Rand des Tableaus) merken, können wir die Einheitsmatrix weglassen. Ein derartiges Tableau nennt man *verkürztes Tableau*. Wir schreiben die zweite Variante der vorhergehenden Rechnung noch einmal in dieser verkürzten Form auf.

$$T_0 : \begin{array}{cc|c} & x_1 & x_2 & \\ & 1 & 2 & 0 \\ \hline x_3 & 1 & 0 & 4 \\ x_4 & 2 & 1 & 10 \\ x_5 & -1 & \boxed{1} & 5 \\ x_6 & 0 & 3 & 20 \end{array}$$

$$T_1 : \begin{array}{cc|c} & x_1 & x_5 & \\ & 3 & -2 & -10 \\ \hline x_3 & 1 & 0 & 4 \\ x_4 & \boxed{3} & -1 & 5 \\ x_2 & -1 & 1 & 5 \\ x_6 & \boxed{3} & -3 & 5 \end{array}$$

$$T'_2 : \begin{array}{ccc|c} & x_4 & x_5 & \\ & -1 & -1 & -15 \\ \hline x_3 & -1/3 & 1/3 & 7/3 \\ x_1 & 1/3 & -1/3 & 5/3 \\ x_2 & 1/3 & 2/3 & 20/3 \\ x_6 & -1 & -2 & 0 \end{array}$$

△

Ein weiteres Beispiel soll zur Einübung der Rechentechnik dienen.

**(9.23) Beispiel.** Wir betrachten das folgende lineare Programm.

$$\begin{aligned} \max \quad & x_1 + x_2 - x_3 \\ & 3x_1 - 2x_2 \leq 3 \\ & 2x_1 + x_3 \leq 4 \\ & x_1 + 3x_2 - 2x_3 \leq 6 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

## 9 Die Grundversion des Simplex-Algorithmus

Wir führen Schlupfvariablen ein und erhalten das folgende zulässige Anfangstableau.

$$\begin{array}{l}
 T_0 : \\
 \begin{array}{ccccccc|c}
 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\
 4 & \boxed{3} & -2 & 0 & 1 & 0 & 0 & 3 \\
 5 & 2 & 0 & 1 & 0 & 1 & 0 & 4 \\
 6 & 1 & 3 & -2 & 0 & 0 & 1 & 6
 \end{array}
 \end{array}
 \quad
 \begin{array}{l}
 s = 1 \\
 r = 1
 \end{array}$$
  

$$\begin{array}{l}
 T_1 : \\
 \begin{array}{ccccccc|c}
 & 0 & \frac{5}{3} & -1 & -\frac{1}{3} & 0 & 0 & -1 \\
 1 & 1 & -\frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 1 \\
 5 & 0 & \frac{4}{3} & 1 & -\frac{2}{3} & 1 & 0 & 2 \\
 6 & 0 & \boxed{5} & -2 & -\frac{1}{3} & 0 & 1 & 5
 \end{array}
 \end{array}
 \quad
 \begin{array}{l}
 s = 2 \\
 r = 3
 \end{array}$$
  

$$\begin{array}{l}
 T_2 : \\
 \begin{array}{ccccccc|c}
 & 0 & 0 & -\frac{1}{11} & -\frac{6}{33} & 0 & -\frac{5}{11} & -\frac{36}{11} \\
 1 & 1 & 0 & -\frac{4}{11} & \frac{9}{33} & 0 & \frac{2}{11} & \frac{21}{11} \\
 5 & 0 & 0 & \frac{19}{11} & -\frac{18}{33} & 1 & -\frac{4}{11} & \frac{2}{11} \\
 2 & 0 & 1 & -\frac{6}{11} & -\frac{1}{11} & 0 & \frac{3}{11} & \frac{15}{11}
 \end{array}
 \end{array}$$

Als Optimallösung ergibt sich somit:  $x_1 = \frac{21}{11}$ ,  $x_2 = \frac{15}{11}$ ,  $x_3 = 0$ , mit  $c^T x = \frac{36}{11}$ .  $\triangle$

Über die Grundversion der Simplexmethode können wir die folgende Aussage machen.

**(9.24) Satz.** *Sei (9.1) ein lineares Programm, so dass  $P=(A, b) \neq \emptyset$  und alle zulässigen Basislösungen nicht entartet sind. Dann findet die Grundversion des Simplexalgorithmus (d. h. alle Schritte, bei denen eine nicht spezifizierte Auswahl besteht, werden in irgendeiner beliebigen Form ausgeführt) nach endlich vielen Schritten eine optimale Lösung oder stellt fest, dass das Problem unbeschränkt ist.*  $\triangle$

**Beweis.** Bei jedem Durchlauf des Simplexalgorithmus wird eine neue Basis erzeugt. Da jede Basislösung nicht entartet ist, hat die neue Basislösung nach (9.16)(c) einen strikt besseren Zielfunktionswert. Ist  $A_{B_1}, \dots, A_{B_i}, \dots$  die Folge der mit dem Simplexalgorithmus generierten Basen, dann folgt daraus, dass keine Basis mehrfach in dieser Folge auftreten kann. Da die Anzahl der verschiedenen Basen von  $A$  endlich ist, ist die Folge der Basen endlich. Ist  $A_{B_k}$  die letzte erzeugte Basis, so ist sie entweder optimal oder in Schritt (II.3) wurde Unbeschränktheit festgestellt.  $\square$

Besitzt (9.1) degenerierte Ecken, so kann es sein, dass die Grundversion des Simplexverfahrens irgendwann einmal nur noch zulässige Basen erzeugt, die alle zu derselben Ecke gehören und bei der die gleichen Matrizen  $\bar{A}$  immer wiederkehren. Man sagt, das Verfahren kreist oder *kreiselt*. In der Tat sind Beispiele konstruiert worden, bei denen dieses Phänomen auftreten kann, siehe (9.25).



(9.25) **Beispiel (Kreiseln).** Gegeben sei das folgende LP:

$$\begin{aligned} \max \quad & \frac{4}{5}x_1 - 18x_2 - x_3 - x_4 \\ & \frac{16}{5}x_1 - 84x_2 - 12x_3 + 8x_4 \leq 0 \\ & \frac{1}{5}x_1 - 5x_2 - \frac{2}{3}x_3 + \frac{1}{3}x_4 \leq 0 \\ & x_1 \leq 1 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

Das verkürzte Tableau zur Anfangsbasis  $B = (5, 6, 7)$  der Schlupfvariablen lautet:

	1	2	3	4		
	$\frac{4}{5}$	-18	-1	-1		0
$\frac{16}{5}$	-84	-12	8	0		5
$\frac{1}{5}$	-5	$-\frac{2}{3}$	$\frac{1}{3}$	0		6
1	0	0	0	1		7

Wir führen nun den Simplexalgorithmus aus. Die Pivotelemente sind in den Tableaus gekennzeichnet.

5	2	3	4		
$-\frac{1}{4}$	3	2	-3		0
$\frac{5}{16}$	$-\frac{105}{4}$	$-\frac{15}{4}$	$\frac{5}{2}$	0	1
$-\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{12}$	$-\frac{1}{6}$	0	6
$-\frac{5}{16}$	$\frac{105}{4}$	$\frac{15}{4}$	$-\frac{5}{2}$	1	7

5	6	3	4		
$\frac{1}{2}$	-12	1	-1		0
$-\frac{25}{4}$	105	$\frac{5}{3}$	-15	0	1
$-\frac{1}{4}$	4	$\frac{1}{3}$	$-\frac{2}{3}$	0	2
$\frac{25}{4}$	-105	-5	15	1	7

5	6	1	4		
$\frac{7}{4}$	-33	$-\frac{1}{5}$	2	0	
$-\frac{5}{4}$	21	$\frac{1}{5}$	-3	0	3
$\frac{1}{6}$	-3	$-\frac{1}{15}$	$\frac{1}{3}$	0	2
0	0	1	0	1	7

5	6	1	2		
$\frac{3}{4}$	-15	$\frac{1}{5}$	-6	0	
$\frac{1}{4}$	-6	$-\frac{2}{5}$	9	0	3
$\frac{1}{2}$	-9	$-\frac{1}{5}$	3	0	4
0	0	1	0	1	7

3	6	1	2		
-3	3	$\frac{7}{5}$	-33	0	
4	-24	$-\frac{8}{5}$	36	0	5
-2	$\frac{3}{5}$	$\frac{3}{5}$	-15	0	4
0	0	1	0	1	7

3	4	1	2		
-1	-1	$\frac{4}{5}$	-18	0	
-12	8	$\frac{16}{5}$	-84	0	5
$-\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{5}$	-5	0	6
0	0	1	0	1	7

## 9 Die Grundversion des Simplex-Algorithmus

Das letzte Tableau ist bis auf Spaltenvertauschung das gleiche wie das erste. Alle Basen der oben generierten Folge gehören zur Ecke  $x^T = (0, 0, 0, 0)$  des Ausgangsproblems. Die bei der Berechnung des Beispiels benutzte Variante des Simplexverfahrens würde also nicht nach endlich vielen Schritten abbrechen.  $\triangle$

Wie Satz (9.24) zeigt, funktioniert die Grundversion der Simplexmethode, gleichgültig welche Arten von Auswahlregeln man in den Schritten (II.2) und (II.4) benutzt, wenn alle Basislösungen nicht entartet sind. Es ist daher sinnvoll zu fragen, ob man nicht das Ausgangsproblem so modifizieren (stören) kann, dass das neue Problem nicht entartet ist. Dass dies geht, zeigt der folgende Satz. Wie üblich bezeichnen wir das LP  $\max\{c^T x \mid Ax = b, x \geq 0\}$  mit (9.1). Sei für ein  $\varepsilon^T = (\varepsilon_1, \dots, \varepsilon_n)$ ,  $\varepsilon > 0$

$$\begin{aligned} \max c^T x \\ Ax = b + A\varepsilon \\ x \geq 0. \end{aligned} \tag{9.1\varepsilon}$$

### (9.26) Satz (Perturbationsmethode).

- (a) *Das lineare Programm (9.1) ist genau dann lösbar, wenn das LP (9.1\varepsilon) lösbar ist für jedes  $\varepsilon > 0$ .*
- (b) *Es gibt ein  $\varepsilon_0 > 0$  derart, dass für alle  $0 < \varepsilon < \varepsilon_0$  jede zulässige Basis von (9.1\varepsilon) zum einen eine nichtdegenerierte Basislösung von (9.1\varepsilon) bestimmt und zum anderen eine zulässige Basis von (9.1) ist. Ferner bestimmt die zu einer optimalen Basislösung von (9.1\varepsilon) gehörige Basis eine optimale Basislösung von (9.1).*  $\triangle$

**Beweis.** Siehe Kall (1976, S. 52–53).  $\square$

Aufgrund von Satz (9.26) könnte also jedes LP durch Störung in ein nichtentartetes LP umgewandelt werden, und damit wäre die Konvergenz der Grundversion des Simplexalgorithmus gewährleistet. In der Praxis wird diese Methode jedoch kaum angewendet, da das  $\varepsilon_0$  nicht einfach zu bestimmen ist und der Ansatz mit irgendeiner Zahl  $\varepsilon > 0$  nicht zum Ziele führen muss. Ferner könnten durch eine geringe Störung des Problems durch Rundungsfehler numerische Schwierigkeiten auftreten, deren Auswirkungen nicht immer abschätzbar sind.

Wir werden im nächsten Abschnitt andere Methoden zur Vermeidung des Kreiseln kennenlernen. In der Praxis werden diese häufig jedoch gar nicht implementiert, da bei praktischen Problemen trotz gelegentlicher Degeneration ein Kreiseln sehr selten beobachtet wurde. Einer der Gründe dafür dürfte darin zu sehen sein, dass der Computer durch seine begrenzte Rechengenauigkeit sozusagen „von allein“ eine Störungsmethode durchführt. So wird bei jedem Auf- oder Abrunden eine kleine Änderung des vorliegenden Programms vorgenommen, die sich positiv auf die Konvergenzeigenschaften des Simplexverfahrens auswirkt.

## 9.4 Spalten- und Zeilenauswahlregeln

Wir wollen uns nun damit beschäftigen, wie die noch unspezifizierten Schritte des Simplexverfahrens „vernünftig“ ausgeführt werden können.

**(9.27) Bemerkung (Varianten der Spaltenauswahl in Schritt II.2).** Sei  $S := \{j \in \{1, \dots, n\} \mid \bar{c}_j > 0\} \neq \emptyset$ . Folgende Regeln sind in der Literatur eingehend diskutiert worden und werden bei schlichten Implementationen des Simplexverfahrens angewendet.

- (1) *Kleinster-Index-Regel*: Wähle  $s = \min\{j \in S\}$ .
- (2) *Kleinster-Variablenindex-Regel*: Wähle  $s \in S$ , so dass  $q_s = \min\{q_j \in N \mid j \in S\}$ .
- (3) *Steilster-Anstieg-Regel*: Wähle  $s \in S$ , so dass  $\bar{c}_s = \max\{\bar{c}_j \mid j \in S\}$ .
- (4) *Größter-Fortschritt-Regel*: Berechne für jedes  $j \in S$  zunächst

$$\lambda_0^j = \min\left\{\frac{\bar{b}_i}{\bar{a}_{ij}} \mid \bar{a}_{ij} > 0, i = 1, \dots, m\right\} \quad \text{und} \quad g_j = \bar{c}_j \lambda_0^j.$$

Wähle  $s \in S$ , so dass  $g_s = \max\{g_j \mid j \in S\}$ .

- (5) *Varianten von (4)*: Es gibt unzählige Varianten von (4), einige davon sind beschrieben in:

- Goldfarb and Reid (1977): “A practicable steepest-edge simplex algorithm”
- Harris (1973): “Pivot selection methods for the Devex LP code”
- Crowder and Hattingh (1975): “Partially normalized pivot selection in linear programming” △

Die Regel (9.27)(1) ist die rechentechnisch einfachste. Man durchläuft den Vektor  $\bar{c}$ , sobald ein Index  $s$  mit  $\bar{c}_s > 0$  gefunden ist, hört man auf und geht zum nächsten Schritt. Die reduzierten Kosten müssen also nicht alle berechnet werden. Dadurch wird viel Aufwand gespart, aber aufgrund der simplen Wahl von  $s$  ist die Gesamtzahl der Pivotoperationen im Vergleich zu anderen Regeln recht hoch.

Ähnliches gilt für Regel (2). Hier müssen jedoch alle reduzierten Kostenkoeffizienten berechnet werden. Diese Regel ist jedoch aus einem theoretischen Grund, der noch diskutiert werden soll, interessant.

Die Regel (3) ist die Regel, die bei einfachen Implementationen (keine ausgeklügelten Tricks) am häufigsten verwendet wird und bei Problemen bis zu mittleren Größenordnungen (jeweils bis zu 500 Variablen und Zeilen) recht gute Ergebnisse zeigt. Ihr liegt die Idee zu Grunde, dass diejenige Variable in die Basis genommen werden sollte, die pro Einheit den größten Zuwachs in der Zielfunktion bringt.

Es kann natürlich sein, dass die Nichtbasisvariable mit dem größten Zuwachs pro Einheit nur um sehr wenige Einheiten erhöht werden kann und dass ein Wechsel zu einer anderen Basis einen wesentlich größeren Gesamtfortschritt bringt. Hierzu ist Regel (4) geschaffen. Bei ihr wird für jeden möglichen Basiswechsel der tatsächliche Zuwachs  $g_j$  der

Zielfunktion berechnet, und es wird der Basiswechsel vorgenommen, der den insgesamt größten Fortschritt bringt. Diese Verbesserung wird natürlich durch einen erheblichen rechnerischen Mehraufwand erkauft, bei dem es fraglich ist, ob er sich lohnt.

Aufgrund von Erfahrungen in der Praxis kann gesagt werden, dass sich die trivialen Regeln (1), (2) und (3) für kleinere bis mittlere Problemgrößen bewährt haben, jedoch für große LPs, Modifizierungen von (9.27)(4), wie sie in den Literaturangaben beschrieben sind, benutzt werden. Die trivialen Regeln führen im allgemeinen zu insgesamt mehr Pivotoperationen. Die komplizierten Regeln versuchen die Anzahl der Pivotoperationen minimal zu gestalten. Hierbei ist ein wesentlich höherer Rechenaufwand erforderlich (viele Spalten von  $\bar{A}$  sind zu generieren und für jede Spalte ist  $\lambda_0$  zu berechnen), der bei kleineren Problemen durchaus einem mehrfachen Basisaustausch entspricht und sich daher nicht auszahlt. Bei großen Problemen ist i. A. der Basiswechsel rechenaufwendiger als die komplizierte Auswahl der Pivotspalte. Hier zahlt sich dann die sorgfältige Auswahl der Pivotspalte bei der Gesamtrechenzeit aus.

Bei der Auswahl von Pivotzeilen gibt es nicht so viele interessante Regeln, aber hier kann durch geschickte Wahl die Endlichkeit des Simplexverfahrens erzwungen werden.

Im folgenden gehen wir davon aus, dass der Spaltenindex  $s$  durch eine Spaltenauswahlregel (9.27) bestimmt wurde und setzen  $R := \{i \in \{1, \dots, m\} \mid \frac{\bar{b}_i}{\bar{a}_{is}} = \lambda_0\}$ . Wir treffen zunächst eine Definition.

**(9.28) Definition.** Ein Vektor  $x^T = (x_1, \dots, x_n) \in \mathbb{K}^n$  heißt lexikographisch positiv (wir schreiben:  $x \succ 0$ ), wenn die erste von Null verschiedene Komponente positiv ist (d. h. ist  $i := \min \text{supp}(x)$ , dann ist  $x_i > 0$ ). Wir schreiben  $x \succ y$ , falls  $x - y \succ 0$  gilt.  $\triangle$

**(9.29) Bemerkung.** „ $\succ$ “ definiert eine totale Ordnung im  $\mathbb{K}^n$ . Wir schreiben  $x \succeq y \iff x \succ y \vee x = y$  und bezeichnen mit  $\text{lex-min } S$  das lexikographische Minimum einer endlichen Menge  $S \subseteq \mathbb{K}^n$ .  $\triangle$

**(9.30) Bemerkung (1. Lexikographische Zeilenauswahlregel).** Wähle  $r \in R$  so dass

$$\frac{1}{\bar{a}_{rs}}(A_B^{-1}A)_{r.} = \text{lex-min} \left\{ \frac{1}{\bar{a}_{is}}(A_B^{-1}A)_{i.} \mid \bar{a}_{is} > 0, i \in R \right\} \quad \triangle$$

Mit Regel (9.30) wird jedes Simplexverfahren unabhängig von der Spaltenauswahlregel endlich.

**(9.31) Bemerkung (Endlichkeit des Simplexalgorithmus).** Wird im Simplexverfahren die 1. Lexikographische Regel (9.30) verwendet, so endet der Algorithmus nach endlich vielen Schritten, gleichgültig, welche Spaltenauswahlregel verwendet wird.  $\triangle$

**Beweis.** Für den Beweis benutzen wir eine „umgekehrte“ Version des Simplextableaus:

$$\tilde{T}_B := \begin{pmatrix} -c_B^T A_B^{-1} b & c^T - c_B^T A_B^{-1} A \\ A_B^{-1} b & A_B^{-1} A \end{pmatrix}$$

mit Zeilen  $i = 0, \dots, m$  und Spalten  $j = 0, \dots, n$  (vgl. (9.20)). Dann lautet die Auswahlregel (9.30): Wähle  $r \in \{1, \dots, m\}$  so dass  $\bar{a}_{rs} > 0$  und

$$\frac{1}{\bar{a}_{rs}}(\tilde{T}_B)_r. = \text{lex-min} \left\{ \frac{1}{\bar{a}_{is}}(\tilde{T}_B)_i. \mid \bar{a}_{is} > 0 \right\}.$$

Wir zeigen zuerst:

$$\frac{1}{\bar{a}_{rs}}(\tilde{T}_B)_r. < \frac{1}{\bar{a}_{is}}(\tilde{T}_B)_i. \quad \forall i \in \{1, \dots, m\}, i \neq r \text{ mit } \bar{a}_{is} > 0, \quad (9.32)$$

das heißt, das lexikographische Minimum wird eindeutig angenommen. Sind  $r$  und  $i$  Indizes mit

$$\frac{1}{\bar{a}_{rs}}(\tilde{T}_B)_r. = \frac{1}{\bar{a}_{is}}(\tilde{T}_B)_i.,$$

dann folgt

$$\frac{1}{\bar{a}_{rs}}(A_B^{-1}A)_r. = \frac{1}{\bar{a}_{is}}(A_B^{-1}A)_i.. \quad (9.33)$$

O. B. d. A. sei  $A = (A_B, A_N)$ . Dann gilt  $(A_B^{-1}A)_r. = (A_B^{-1})_r. A = (e_r^T, \bar{A}_r.)$  und  $(A_B^{-1}A)_i. = (e_i^T, \bar{A}_i.)$  und (9.33) impliziert  $(e_r^T, \bar{A}_r.) = \frac{\bar{a}_{rs}}{\bar{a}_{is}}(e_i^T, \bar{A}_i.)$ . Damit folgt  $e_r = \frac{\bar{a}_{rs}}{\bar{a}_{is}}e_i$  und somit  $r = i$ . Also gilt (9.32).

Sei  $A_B$  nun eine zulässige Basis von  $A$  (z. B. die Startbasis (evtl. auch die der Phase I)) und sei  $\tilde{T}_B$  das zugehörige (umgekehrte) Simplextableau. O. B. d. A. sei

$$\tilde{T}_B = \begin{pmatrix} -c_B^T A_B^{-1} b & 0 & \bar{c} \\ A_B^{-1} b & I & \bar{A} \end{pmatrix}$$

(evtl. ist eine Permutation der Spalten der Ausgangsmatrix notwendig).

Sei  $B' = (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$ . Dann gilt

$$(\tilde{T}_{B'})_r. \succ 0$$

denn  $(\tilde{T}_B)_r. \succ 0$  und  $\bar{a}_{rs} > 0$  (vgl. Formel (9.21)(a)). Da nach Voraussetzung  $(\tilde{T}_B)_i. \succ 0$  für alle  $i = 1, \dots, m$ , folgt

$$(\tilde{T}_{B'})_i. \succ 0 \quad \forall i \text{ mit } \bar{a}_{is} \leq 0.$$

Aus (9.32) folgt nun  $(\tilde{T}_B)_i. \succ \frac{\bar{a}_{is}}{\bar{a}_{rs}}(\tilde{T}_B)_r.$  für alle  $i \neq r$  mit  $\bar{a}_{is} > 0$ , d. h.

$$(\tilde{T}_{B'})_i. \succ 0 \quad \forall i \text{ mit } \bar{a}_{is} > 0.$$

Somit sind bei Anwendung der 1. Lexikographischen Regel stets alle Zeilen  $(\tilde{T}_B)_i.$  ( $i \neq 0$ ) des (umgekehrten) Simplextableaus lexikographisch positiv, d. h. es gilt für zwei aufeinanderfolgende Basen  $B$  und  $B'$  (vgl. (9.21)(a))

$$(\tilde{T}_{B'})_0. - (\tilde{T}_B)_0. = -\frac{(\tilde{T}_B)_{0s}}{\bar{a}_{rs}}(\tilde{T}_B)_r. < 0,$$

denn  $(\tilde{T}_B)_{0s} > 0$ . Also sind alle Simplextableaus verschieden und da nur endlich viele zulässige Basislösungen und somit Simplextableaus existieren, folgt nun die Behauptung.  $\square$

**(9.34) Bemerkung (2. Lexikographische Zeilenauswahlregel).** Wähle  $r \in R$ , so dass

$$\frac{1}{a_{rs}}(A_B^{-1})_r = \text{lex-min} \left\{ \frac{1}{a_{is}}(A_B^{-1})_i \mid \bar{a}_{is} > 0, i \in \{1, \dots, m\} \right\}. \quad \triangle$$

Die Anwendung der 2. Lexikographischen Regel garantiert auch Endlichkeit des Simplexverfahrens, analog zu Satz (9.31) (siehe Kall oder Dantzig S. 269, diese folgt aus der Störungsmethode).

**(9.35) Bemerkung (Weitere Zeilenauswahlregeln).**

(1) *Kleinster-Index-Regel:*  $r := \min R$ .

(2) *Kleinster-Variablenindex-Regel:* Wähle  $r \in R$ , so dass  $p_r = \min\{p_i \in B \mid i \in R\}$ .  $\triangle$

Keine der beiden Regeln in (9.35) kann für sich allein Endlichkeit des Simplexverfahrens bewirken. Aber eine Kombination von (9.35)(2) und (9.27)(2) schafft es.

**(9.36) Bemerkung (Bland-Regel).** Bland (1977) hat gezeigt, dass das Simplexverfahren auch endlich ist, wenn sowohl bei der Spaltenauswahl als auch bei der Zeilenauswahl die Kleinster-Variablen-Index Regel angewendet wird.  $\triangle$

Von Avis and Chvátal (1978) sind Untersuchungen über die Praktikabilität der Bland-Regel gemacht worden. Dabei hat sich gezeigt, dass i. A. die Anzahl der Pivotoperationen bei Anwendung der Bland-Regel wesentlich höher liegt als z. B. bei der Steilster-Anstieg Regel (9.27)(3). Für Computerimplementierungen ist die Bland-Regel selbst bei hochdegenerierten Problemen nicht gut geeignet.

**(9.37) Bemerkung (Worst-Case Verhalten des Simplexverfahrens).** Zu fast allen bekannten Pivotauswahlregeln (speziell zu allen, die hier genannt wurden) kennt man heute eine Klasse von Polyedern und Zielfunktionen, so dass der Simplexalgorithmus bei Anwendung einer zugehörigen Auswahlregel durch alle Ecken des Polyeders läuft. Da die Anzahl der Ecken dieser Polyeder exponentiell mit der Anzahl der Variablen wächst, sagt man, dass das Simplexverfahren ein exponentielles worst-case Verhalten besitzt.  $\triangle$

## 9.5 Die Phase I

Zur vollständigen Beschreibung der Grundversion der Simplexmethode fehlt noch eine Darstellung der Phase I von (9.17). Diese wollen wir nun nachholen. Ist das Ungleichungssystem  $Ax = b$ ,  $x \geq 0$  gegeben, so können wir stets o. B. d. A.  $b \geq 0$  annehmen, denn Multiplikation einer Gleichung mit  $-1$  verändert das Polyeder nicht.

**(9.38) Phase I des Simplexverfahrens.**

**Eingabe:**  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$  mit  $b \geq 0$ .

**Ausgabe:**

(a)  $P^=(A, b) = \emptyset$  oder

- (b)  $P^=(A, b) = \{x\}$  oder
- (c) Ausgabe von  $I \subseteq \{1, \dots, m\}$  und  $B = (p_1, \dots, p_k)$  mit folgenden Eigenschaften:
- (1) Mit  $A' := A_I, b' := b_I$  gilt  $P^=(A', b') \neq \emptyset, \text{rang}(A') = |I| = k, k < n$  und  $P^=(A', b') = P^=(A, b)$ .
  - (2)  $A'_B$  ist eine zulässige Basis von  $A'$ .

Wir formulieren zunächst ein lineares „Hilfsprogramm“. Sei  $D := (A, I)$  und betrachte das LP:

$$\begin{aligned} \max \quad & \mathbf{1}^T Ax \\ D \begin{pmatrix} x \\ y \end{pmatrix} &= b \\ x, y &\geq 0 \end{aligned} \tag{9.39}$$

wobei  $x^T = (x_1, \dots, x_n), y^T = (y_{n+1}, \dots, y_{n+m})$  gesetzt wird. Wir nennen die Variablen  $y_{n+1}, \dots, y_{n+m}$  *künstliche Variablen* und die Variablen  $x_1, \dots, x_n$  *Originalvariablen*. Das LP (9.39) erfüllt die Voraussetzungen (9.2)(a), (b), (c), und mit  $B = (n+1, \dots, n+m)$  ist  $D_B$  eine zulässige Basis mit zugehöriger Basislösung  $x = 0, y = b$ . Es gilt offenbar

$$\begin{aligned} D \begin{pmatrix} x \\ y \end{pmatrix} = b \\ x, y \geq 0 \end{aligned} \iff \begin{aligned} Ax + y = b \\ x, y \geq 0 \end{aligned} ,$$

daraus folgt  $\mathbf{1}^T Ax = \mathbf{1}^T b - \mathbf{1}^T y$ , d. h. (9.39) ist äquivalent zu  $\max\{\mathbf{1}^T b - \mathbf{1}^T y \mid Ax + y = b, x, y \geq 0\}$  bzw. zu

$$\begin{aligned} \mathbf{1}^T b - \min \quad & \mathbf{1}^T y \\ Ax + y &= b \\ x, y &\geq 0. \end{aligned} \tag{9.40}$$

(9.40) bzw. (9.39) besagen also, dass die künstlich eingeführten Variablen möglichst klein gewählt werden sollen.

(I.1) Wende die Grundversion (9.17) des Simplexverfahrens auf (9.39) an. Die Matrix  $D$  hat vollen Zeilenrang und weniger Zeilen als Spalten,  $B = (n+1, \dots, n+m)$  definiert eine zulässige Basis, zu der die Matrix  $\bar{A}$  und die Vektoren  $\bar{b}, \bar{c}$  und  $c_0$  trivial zu berechnen sind. Wir können also direkt mit diesen Daten mit Phase II beginnen.

Das LP (9.40) ist offenbar durch  $\mathbf{1}^T b - 0 = \mathbf{1}^T b$  nach oben beschränkt, also ist (9.39) durch  $\mathbf{1}^T b$  nach oben beschränkt, und somit besitzt (9.39) eine optimale Lösung. Daher endet das Simplexverfahren mit einer optimalen Basis  $D_B$  und zugehöriger Basislösung  $z = \begin{pmatrix} x \\ y \end{pmatrix}$ .

Wir werden nun die optimale Lösung bzw. die optimale Basis analysieren, um aus Eigenschaften der Lösung bzw. Basis die Schlussfolgerungen (a), (b) oder (c) ziehen zu können.

## 9 Die Grundversion des Simplex-Algorithmus

- (I.2) Falls  $\mathbf{1}^T Ax < \mathbf{1}^T b$ , so wird das Minimum in (9.40) nicht in  $y = 0$  angenommen. Hieraus folgt offenbar, dass  $P^=(A, b) = \emptyset$  gilt, und wir können das Verfahren mit Antwort (a) beenden, STOP!

Die folgenden Schritte behandeln den Fall  $\mathbf{1}^T Ax = \mathbf{1}^T b$ . D. h. es gilt  $\mathbf{1}^T Ax = \mathbf{1}^T b - \mathbf{1}^T y = \mathbf{1}^T b$  bzw.  $\mathbf{1}^T y = 0$ . Somit gilt  $y = 0$ , und  $x$  ist zulässig bzgl.  $Ax = b$ ,  $x \geq 0$ . Wir müssen nun noch evtl. Zeilen von  $A$  streichen, um die Rangbedingung zu erfüllen.

- (I.3) Falls  $B \cap \{n+1, \dots, n+m\} = \emptyset$ , so befinden sich in der optimalen Basis keine künstlichen Variablen, d. h. es gilt  $D_B = A_B$ . Da  $D_B = A_B$  regulär ist, gilt  $\text{rang}(A) = m$ .

Falls  $N \cap \{1, \dots, n\} = \emptyset$ , so sind alle Originalvariablen Basisvariablen, und wir haben  $m = n$ . Dann gilt  $x = A^{-1}b$  und  $P^=(A, b) = \{x\}$ . In diesem Falle können wir das Verfahren beenden mit Antwort (b), STOP! Der Vektor  $x$  ist die Optimallösung eines jeden LP über  $P^=(A, b)$ .

Andernfalls ist  $m < n$  und  $A_B$  eine zulässige Basis von  $A$ . Wir setzen  $I = \{1, \dots, m\}$  und schließen das Verfahren mit Antwort (c) ab, STOP!

Die beiden abschließenden Schritte sind für den Fall vorgesehen, dass sich noch künstliche Variablen in der Basis befinden, d. h. falls  $B \cap \{n+1, \dots, n+m\} \neq \emptyset$ . Wir versuchen zunächst, diese künstlichen Variablen aus der Basis zu entfernen. Da alle künstlichen Variablen  $y_{n+1}, \dots, y_{n+m}$  Null sind, ist die Basislösung  $z = (z_B, z_N)$ ,  $z_B = D_B^{-1}b$  degeneriert. Sei o. B. d. A.

$$z_B^T = (y_{p_1}, \dots, y_{p_t}, x_{p_{t+1}}, \dots, x_{p_m}),$$

d. h.  $B \cap \{n+1, \dots, n+m\} = \{p_1, \dots, p_t\}$ . Wenn wir also künstliche Variablen aus der Basis entfernen wollen, müssen wir prüfen, ob wir sie aus der Basis „hinauspivotisieren“ können, d. h. ob in den Zeilen  $\bar{D}_i$ . ( $i = 1, \dots, t$ ) ( $\bar{D} = (\bar{d}_{rs}) = D_B^{-1}D_N$ ) von Null verschiedene Pivotelemente vorhanden sind. Da wir ja nur Nichtbasisvariable gegen degenerierte Basisvariable austauschen, also Nullen gegen Nullen tauschen, können wir auch Pivotoperationen auf negativen Elementen zulassen.

- (I.4) Falls es  $r \in \{1, \dots, t\}$  und  $s \in \{1, \dots, n\}$  gibt, so dass  $z_{q_s}$  keine künstliche Variable ist und  $\bar{d}_{rs} \neq 0$  gilt, dann führe eine Pivotoperation mit dem Pivotelement  $\bar{d}_{rs}$  entsprechend Satz (9.12) durch. Wir erhalten dadurch eine neue Basis  $D_{B'}$ , mit einer künstlichen Variablen weniger, setzen  $B := B'$  und gehen zu (I.3).

- (I.5) Falls  $\bar{d}_{ij} = 0 \forall i \in \{1, \dots, t\} \forall s \in \{1, \dots, n\}$ , für die  $z_{q_s}$  keine künstliche Variable ist, dann hat das Gleichungssystem  $z_B = D_B^{-1}b - D_B^{-1}D_N z_N$  bis auf Permutationen



der Zeilen und Spalten folgende Form:

$$\begin{pmatrix} z_{p_1} \\ \vdots \\ z_{p_t} \end{pmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} - \left[ \begin{array}{c|c} 0 & * \end{array} \right] \begin{pmatrix} x_N \\ y_N \end{pmatrix}$$

und

$$\begin{pmatrix} z_{p_{t+1}} \\ \vdots \\ z_{p_m} \end{pmatrix} = \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix} - \left[ \begin{array}{c|c} * & * \end{array} \right] \begin{pmatrix} x_N \\ y_N \end{pmatrix}$$

(mit  $y_B = (z_{p_1}, \dots, z_{p_t})$ ,  $x_B = (z_{p_{t+1}}, \dots, z_{p_m})$ ,  $z_N = (x_N, y_N)$ ).

Setzen wir also die künstlichen Variablen Null, so ist der obere Teil des Gleichungssystems

$$y_B = 0 - 0x_N$$

unabhängig von der Belegung von  $x_N$  stets erfüllt, d. h. wir können alle Zeilen  $D_1, \dots, D_t$  streichen. Ist  $I = \{t+1, \dots, m\}$  und  $J = \{j \in \{1, \dots, n\} \mid z_{q_j} \text{ ist keine künstliche Variable}\}$ , so gilt

$$x_B = (D_B^{-1}b)_I - \bar{D}_{IJ}x_N$$

und  $B = (p_{t+1}, \dots, p_m)$  ist eine zulässige Basis von  $A' := A_I$  mit  $\text{rang}(A') = k := m - t$ .

Ist  $k = n$ , so gilt wieder  $P^-(A, b) = \{x\}$ , und wir enden mit Antwort (b). Andernfalls gilt  $k < n$ , und wir beenden das Verfahren mit Antwort (c), STOP!  $\triangle$

Die Korrektheit des Phase-I-Verfahrens (9.38) ist aufgrund der oben gemachten Bemerkungen offensichtlich. Algorithmus (9.38) zeigt auch, dass man bei der Lösung von linearen Programmen im wesentlichen mit der Grundversion des Simplexalgorithmus zur Lösung von Phase II auskommt. Es ist lediglich zusätzlich etwas Buchhaltung (Einführung künstlicher Variablen, Streichen von Spalten und Zeilen) zusätzlich durchzuführen. Außerdem sind u. U. im Schritt (I.4) Pivotoperationen auf negativen Pivotelementen erforderlich. Die Pivotformeln ändern sich dadurch aber nicht.

**(9.41) Bemerkung.** Liegen Probleme mit nicht vorzeichenbeschränkten Variablen vor, so wendet man die Transformationsregeln (2.4) aus Kapitel 2 an, um das Problem in ein LP mit vorzeichenbeschränkten Variablen zu transformieren.  $\triangle$

Zur Auffindung einer Anfangslösung gibt es noch andere Varianten (z. B. die  $M$ -Methode), jedoch wird in kommerziellen Codes meist die Zweiphasenmethode verwendet.

**(9.42) Beispiel (für Phase I).** Gegeben sei das folgende lineare Programm in Standardform

$$\begin{aligned} \max \quad & x_1 - x_2 + 2x_3 \\ & 2x_1 - 3x_2 + x_3 = 3 \\ & -x_1 + 2x_2 + x_3 = -1 \\ & 3x_1 - 5x_2 = 4 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Durch Multiplikation der zweiten Zeile mit  $-1$  machen wir die rechten Seiten nichtnegativ. Es gilt dann  $\mathbf{1}^T A = (6, -10, 0)$ ,  $\mathbf{1}^T b = 8$ . Wir stellen nun unser Simplextableau auf, wobei wir künstliche Variablen  $s_1, s_2, s_3$  einführen. Wir fügen dem Simplextableau eine neue oberste Zeile für die künstliche Zielfunktion hinzu.

$$T_0 : \begin{array}{cccccc|cc} & x_1 & x_2 & x_3 & s_1 & s_2 & s_3 & & \\ & 6 & -10 & 0 & 0 & 0 & 0 & 0 & \leftarrow \text{künstliche Zielfunktion} \\ & 1 & -1 & 2 & 0 & 0 & 0 & 0 & \leftarrow \text{akt. Zielfunktion} \\ s_1 & \hline 2 & -3 & 1 & 1 & 0 & 0 & 3 & \\ s_2 & \boxed{1} & -2 & -1 & 0 & 1 & 0 & 1 & \\ s_3 & 3 & -5 & 0 & 0 & 0 & 1 & 4 & \end{array}$$

$$T_1 : \begin{array}{cccccc|cc} & x_1 & x_2 & x_3 & s_1 & s_2 & s_3 & & \\ & 0 & 2 & 6 & 0 & -6 & 0 & -6 & \\ & 0 & 1 & 3 & 0 & -1 & 0 & -1 & \\ s_1 & \hline 0 & 1 & \boxed{3} & 1 & -2 & 0 & 1 & \\ x_1 & 1 & -2 & -1 & 0 & 1 & 0 & 1 & \\ s_3 & 0 & 1 & 3 & 0 & -3 & 1 & 1 & \end{array}$$

$$T_2 : \begin{array}{cccccc|cc} & x_1 & x_2 & x_3 & s_1 & s_2 & s_3 & & \\ & 0 & 0 & 0 & -2 & -2 & 0 & -8 & \\ & 0 & 0 & 0 & -1 & 1 & 0 & -2 & \\ x_3 & \hline 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & \frac{1}{3} & \\ x_1 & 1 & -\frac{5}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{4}{3} & \\ s_3 & 0 & 0 & 0 & -1 & -1 & 1 & 0 & \end{array}$$

Das Tableau  $T_2$  ist optimal bezüglich der künstlichen Zielfunktion. Es gilt  $B \cap \{n + 1, \dots, n + m\} = \{n + 3\} = \{p_3\} \neq \emptyset$  (d. h.  $s_3$  ist in der Basis). Schritt (I.4) wird nicht ausgeführt, da  $\bar{d}_{32} = 0$ , also dürfen wir entsprechend (I.5) die letzte Zeile und die Spalten 4, 5 und 6 streichen. Dies ergibt das folgende zulässige Tableau für das ursprüngliche Problem:

$$\begin{array}{ccc|c}
 & x_1 & x_2 & x_3 & \\
 & 0 & 0 & 0 & -2 \\
 x_3 & 0 & \frac{1}{3} & 1 & \frac{1}{3} \\
 x_1 & 1 & -\frac{5}{3} & 0 & \frac{4}{3}
 \end{array}$$

Dieses Tableau ist optimal bezüglich der ursprünglichen Zielfunktion.  $x_1 = \frac{4}{3}$ ,  $x_2 = 0$ ,  $x_3 = \frac{1}{3}$ ,  $c^T x = 2$ .  $\triangle$

## Literaturverzeichnis

- D. Avis and V. Chvátal. Notes on Bland's pivoting rule. *Mathematical Programming Study*, 8:24–34, 1978.
- R. Bland. New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2:103–107, 1977.
- V. Chvátal. *Linear Programming*. Freeman, New York, 1983.
- H. Crowder and J. M. Hattingh. Partially normalized pivot selection in linear programming. *Mathematical Programming Study*, 4:12–25, 1975.
- G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, 1998.
- D. Goldfarb and J. K. Reid. A practicable steepest-edge simplex algorithm. *Mathematical Programming*, 12:361–371, 1977.
- P. M. S. Harris. Pivot selection methods for the devex LP code. *Mathematical Programming*, 5:1–28, 1973.
- P. Kall. *Mathematische Methoden des Operations Research*. Teubner Studienbücher, Stuttgart, 1976.



# 10 Fourier-Motzkin-Elimination und Projektion

Wir haben nun mit dem Simplex-Algorithmus ein Werkzeug zur Hand, um lineare Programme zu lösen. Insbesondere haben wir auch gesehen, wie wir mit dem Simplex-Algorithmus die Frage, ob ein lineares Ungleichungssystem  $Ax \leq b$  eine Lösung hat oder nicht, entscheiden können. Wir werden in diesem Kapitel eine weitere Methode dafür kennenlernen, die gänzlich anders vorgeht und diese für eine theoretische Analyse der (Un)Zulässigkeit linearer Programme nutzen.

Diese Methode geht auf Fourier (1827) zurück, der sie vor rund 180 Jahren formuliert und analysiert hat. Sie wurde vergessen und mehrmals wiederentdeckt, u. a. von Motzkin (1936). Zu weiteren historischen Details zu diesem Thema siehe Schrijver (1998, sect. 12.2).

Wir beginnen mit der Beschreibung eines Algorithmus, der aus einer  $(m, n)$ -Matrix  $A$  und einem Vektor  $b \in \mathbb{K}^m$  eine  $(r, n)$ -Matrix  $D$  und einen Vektor  $d \in \mathbb{K}^r$  macht, so dass eine Spalte von  $D$  aus lauter Nullen besteht und dass gilt:

$$Ax \leq b \text{ hat eine Lösung genau dann, wenn } Dx \leq d \text{ eine Lösung hat.}$$

## 10.1 Fourier-Motzkin-Elimination

**(10.1) Fourier-Motzkin-Elimination (der  $j$ -ten Variablen).**

**Eingabe:** Eine  $(m, n)$ -Matrix  $A = (a_{ij})$ , ein Vektor  $b \in \mathbb{K}^m$  und ein Spaltenindex  $j \in \{1, \dots, n\}$  der Matrix  $A$ .

**Ausgabe:** Eine  $(r, n)$ -Matrix  $D = (d_{ij})$  ( $r$  wird im Algorithmus berechnet) und ein Vektor  $d \in \mathbb{K}^r$ , so dass  $D_{\cdot j}$  (die  $j$ -te Spalte von  $D$ ) der Nullvektor ist.

1. Partitioniere die Menge der Zeilenindizes  $M = \{1, \dots, m\}$  von  $A$  wie folgt:

$$N := \{i \in M \mid a_{ij} < 0\},$$

$$Z := \{i \in M \mid a_{ij} = 0\},$$

$$P := \{i \in M \mid a_{ij} > 0\}.$$

(Die Menge  $Z \cup (N \times P)$  wird die Zeilenindexmenge von  $D$ .)

2. Setze  $r := |Z \cup (N \times P)|$ ,  $R := \{1, \dots, r\}$ , sei  $p : R \rightarrow Z \cup (N \times P)$  eine Bijektion (d. h. eine kanonische Indizierung der Elemente von  $Z \cup (N \times P)$ ).
3. Führe für  $i = 1, 2, \dots, r$  aus:

## 10 Fourier-Motzkin-Elimination und Projektion

- (a) Falls  $p(i) \in Z$ , dann setze  $D_i := A_{p(i)\cdot}$ ,  $d_i := b_{p(i)}$  (d. h., die  $i$ -te Zeile von  $D$  ist gleich der  $p(i)$ -ten Zeile von  $A$ ).
- (b) Falls  $p(i) = (s, t) \in N \times P$ , dann setze

$$\begin{aligned} D_i &:= a_{tj}A_s - a_{sj}A_t, \\ d_i &:= a_{tj}b_s - a_{sj}b_t \end{aligned} \tag{10.2}$$

(d. h., das  $a_{sj}$ -fache der  $t$ -ten Zeile von  $A$  wird vom  $a_{tj}$ -fachen der  $s$ -ten Zeile von  $A$  abgezogen und als  $i$ -te Zeile von  $D$  betrachtet).  $\triangle$

*Hinweis:* Die in (10.1) konstruierte Menge  $Z \cup (N \times P)$  kann leer und somit  $r = 0$  sein. Die Matrix  $D$  ist dann eine Matrix mit 0 Zeilen und  $n$  Spalten und somit  $Dx \leq d$  ein „leeres Ungleichungssystem“. Die Menge der zulässigen Lösungen  $P(D, d) = \{x \in \mathbb{K}^n \mid Dx \leq d\}$  unterliegt folglich keiner Einschränkung und ist daher der ganze Raum  $\mathbb{K}^n$ .

Bevor wir den Algorithmus analysieren, betrachten wir ein Beispiel:

$$-7x_1 + 6x_2 \leq 25 \tag{1}$$

$$+x_1 - 5x_2 \leq 1 \tag{2}$$

$$+x_1 \leq 7 \tag{3}$$

$$-x_1 + 2x_2 \leq 12 \tag{4}$$

$$-x_1 - 3x_2 \leq 1 \tag{5}$$

$$+2x_1 - x_2 \leq 10 \tag{6}$$

Wir wollen die zweite Variable  $x_2$  eliminieren und erhalten in Schritt 1 von (10.1):

$$N = \{2, 5, 6\}, \quad Z = \{3\}, \quad P = \{1, 4\}.$$

Daraus ergibt sich  $|Z \cup (N \times P)| = 7$ . Die Nummerierung der Zeilen von  $D$  ist dann  $R = \{1, \dots, 7\}$ , und wir setzen:

$$\begin{aligned} p(1) &= 3, \quad p(2) = (2, 1), \quad p(3) = (2, 4), \quad p(4) = (5, 1), \quad p(5) = (5, 4), \\ p(6) &= (6, 1), \quad p(7) = (6, 7). \end{aligned}$$

Die zweite Zeile von  $D$  ergibt sich dann als die Summe vom 6-fachen der zweiten Zeile von  $A$  und dem 5-fachen der ersten Zeile von  $A$ . Die zweite Ungleichung des Systems  $Dx \leq d$  hat somit die Form  $-29x_1 + 0x_2 \leq 131$ .

Insgesamt ergibt sich (wenn wir die aus lauter Nullen bestehende zweite Spalte weglassen), das folgende System  $Dx \leq d$ :

$$+x_1 \leq 7 \tag{1}$$

$$-29x_1 \leq 131 \tag{2}$$

$$-3x_1 \leq 62 \tag{3}$$

$$-27x_1 \leq 81 \tag{4}$$

$$-5x_1 \leq 38 \tag{5}$$

$$+5x_1 \leq 85 \tag{6}$$

$$+3x_1 \leq 32 \tag{7}$$

Ganz offensichtlich können einige der Koeffizienten gekürzt werden. Wir sehen, dass die Zeilen (1), (6), (7) obere und die Ungleichungen (2), (3), (4), (5) untere Schranken auf den für  $x_1$  zulässigen Wert liefern. Wir sehen aber auch, dass die Eliminationsmethode viele überflüssige (wir werden das später *redundante* nennen) Ungleichungen liefert.

Rechentechisch ist es offensichtlich geschickter, die Zeilen der Matrix  $A$  vor Beginn des Algorithmus zu skalieren. In einem Vorbereitungsschritt multiplizieren wir jede Zeile  $A_i$  der Matrix  $A$  und die Zahl  $b_i$ ,  $i \in P \cup N$  mit dem Wert  $|\frac{1}{a_{ij}}|$ . Nennen wir die so skalierte Matrix ebenfalls  $A$ , so haben die Elemente der  $j$ -ten Spalte den Wert 0, +1 oder  $-1$ . Die Berechnung der Zeile  $D_i$  der Matrix  $D$  und der rechten Seite  $d_i$  des neuen Ungleichungssystems  $Dx \leq d$  in Schritt 3(b) geschieht dann durch

$$\begin{aligned} D_i &:= A_s + A_t, \\ d_i &:= b_s + b_t. \end{aligned} \tag{10.3}$$

Wir wollen nun das Ergebnis der Elimination der  $j$ -ten Variablen analysieren. Wir schauen zunächst einige Trivialfälle an.

Wenn alle Elemente der Spalte  $A_{\cdot j}$  Null sind, so ist  $A = D$ . Gilt  $M = P$  oder  $M = N$ , sind also alle Elemente der Spalte  $A_{\cdot j}$  von Null verschieden und haben dasselbe Vorzeichen, dann ist  $r = 0$  und somit  $D$  die leere Matrix mit 0 Zeilen und  $n$  Spalten.  $P(D, d)$  ist dann der gesamte Raum  $\mathbb{K}^n$ .

**(10.4) Satz.** Seien  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$ ,  $j$  ein Spaltenindex von  $A$ , und seien  $D \in \mathbb{K}^{(r,n)}$ ,  $d \in \mathbb{K}^r$  die in (10.1) konstruierte Matrix bzw. Vektor. Es gilt:

- (a) Die  $j$ -te Spalte von  $D$  ist der Nullvektor.
- (b) Jede Ungleichung  $D_i x \leq d_i$ ,  $i = 1, \dots, r$ , ist eine konische Kombination der Ungleichungen  $A_k x \leq b_k$ ,  $k = 1, \dots, m$ , d. h. es existiert ein Vektor  $u \in \mathbb{K}^m$ ,  $u \geq 0$  mit  $u^T A = D_i$  und  $u^T b = d_i$ .
- (c) Aus (b) folgt: Es existiert eine  $(r, m)$ -Matrix  $U$ ,  $U \geq 0$ , mit  $UA = D$  und  $Ub = d$ .
- (d) Sei  $\bar{x} \in \mathbb{K}^n$  mit  $\bar{x}_j = 0$ , sei  $e_j$  der  $j$ -te Einheitsvektor und seien  $N, Z, P$  die in Schritt 1 von (10.1) definierten Mengen. Setzen wir:

$$\begin{aligned} \lambda_i &:= \frac{1}{a_{ij}}(b_i - A_i \bar{x}) \quad \text{für alle } i \in P \cup N, \\ L &:= \begin{cases} -\infty & \text{falls } N = \emptyset, \\ \max\{\lambda_i \mid i \in N\} & \text{falls } N \neq \emptyset, \end{cases} \\ U &:= \begin{cases} +\infty & \text{falls } P = \emptyset, \\ \min\{\lambda_i \mid i \in P\} & \text{falls } P \neq \emptyset. \end{cases} \end{aligned}$$

Dann gilt

- (d<sub>1</sub>)  $\bar{x} \in P(D, d) \implies L \leq U$  und  $\bar{x} + \lambda e_j \in P(A, b) \forall \lambda \in [L, U]$
- (d<sub>2</sub>)  $\bar{x} + \lambda e_j \in P(A, b) \implies \lambda \in [L, U]$  und  $\bar{x} \in P(D, d)$ . △

**Beweis.** (a) Falls  $D$  die leere Matrix ist, so ist die Behauptung natürlich richtig. Wurde eine Zeile  $i$  von  $D$  in Schritt 3(a) definiert, so ist  $d_{ij} = 0$ , da  $a_{p(i)j} = 0$ . Andernfalls gilt  $d_{ij} = a_{tj}a_{sj} - a_{sj}a_{tj} = 0$  nach Konstruktion. Das heißt, die  $j$ -te Spalte  $D_{.j}$  von  $D$  ist der Nullvektor. Man beachte auch, dass im Falle  $N = \emptyset$  oder  $P = \emptyset$  die Matrix  $D$  nur aus den Zeilen  $A_i$  von  $A$  besteht mit  $i \in Z$ .

(b) ist offensichtlich nach Definition und impliziert (c).

(d) Sei  $\bar{x} \in \mathbb{K}^n$  mit  $\bar{x}_j = 0$  ein beliebiger Vektor. Wir zeigen zunächst (d<sub>1</sub>). Angenommen  $\bar{x} \in P(D, d)$ . Wir zeigen zunächst, dass  $L \leq U$  gilt. Ist  $P = \emptyset$  oder  $N = \emptyset$ , so gilt offensichtlich  $L \leq U$  nach Definition. Wir können also annehmen, dass  $P \neq \emptyset$ ,  $N \neq \emptyset$  gilt. Es seien  $s \in N$ ,  $t \in P$  und  $q \in R$  so gewählt, dass  $p(q) = (s, t)$  und  $\lambda_s = L$ ,  $\lambda_t = U$  gelten.

Dann gilt

$$a_{tj}A_s.\bar{x} - a_{sj}A_t.\bar{x} = D_q.\bar{x} \leq d_q = a_{tj}b_s - a_{sj}b_t,$$

woraus folgt

$$a_{sj}(b_t - A_t.\bar{x}) \leq a_{tj}(b_s - A_s.\bar{x})$$

und somit (wegen  $a_{tj} > 0$  und  $a_{sj} < 0$ )

$$U = \lambda_t = \frac{1}{a_{tj}}(b_t - A_t.\bar{x}) \geq \frac{1}{a_{sj}}(b_s - A_s.\bar{x}) = \lambda_s = L. \quad (10.5)$$

Wir zeigen nun  $A_i.(\bar{x} + \lambda e_j) \leq b_i$  für alle  $\lambda \in [L, U]$  und alle  $i \in M = P \cup N \cup Z$ , wobei  $e_j$  den  $j$ -ten Einheitsvektor bezeichnet.

Ist  $i \in Z$ , so gibt es ein  $j \in R$  mit  $p(j) = i$  und  $D_j. = A_i.$ ,  $d_j = b_i$ . Aus  $D_j.\bar{x} \leq d_j$  folgt daher  $A_i.(\bar{x} + \lambda e_j) = A_i.\bar{x} + a_{ij}\lambda = A_i.\bar{x} \leq b_i$ .

Ist  $i \in P$ , dann gilt  $U < +\infty$  und somit

$$A_i.(\bar{x} + \lambda e_j) = A_i.\bar{x} + a_{ij}\lambda \leq A_i.\bar{x} + a_{ij}U \leq A_i.\bar{x} + a_{ij}\lambda_i = b_i.$$

Analog folgt die Behauptung für  $i \in N$ .

Der Beweis von (d<sub>2</sub>) sei dem Leser überlassen. □

**(10.6) Folgerung.**  $P(A, b) \neq \emptyset \iff P(D, d) \neq \emptyset.$  △

Sprechweise für (10.6):

$$Ax \leq b \text{ ist konsistent (lösbar)} \iff Dx \leq d \text{ ist konsistent (lösbar)}.$$

Die Fourier-Motzkin-Elimination der  $j$ -ten Variablen kann man natürlich auch auf ein System von normalen und strikten Ungleichungen anwenden. Eine Nachvollziehung des obigen Beweises liefert:



**(10.7) Satz.** Seien  $A$  eine  $(m, n)$ -Matrix,  $b \in \mathbb{K}^m$ ,  $j \in \{1, \dots, n\}$  und  $I, J$  eine Partition der Zeilenindexmenge  $M = \{1, \dots, m\}$ . Seien  $D \in \mathbb{K}^{r \times n}$ ,  $d \in \mathbb{K}^r$  die durch Fourier-Motzkin-Elimination der Variablen  $j$  gewonnene Matrix bzw. Vektor.

Setze  $E := p^{-1}((Z \cap I) \cup ((N \times P) \cap (I \times I)))$ ,  $F := R \setminus E$ , (wobei  $p$  die in (10.1) Schritt 2 definierte Abbildung ist), dann gilt: Das System

$$A_I \cdot x \leq b_I, \quad A_J \cdot x < b_J$$

hat eine Lösung genau dann, wenn das System

$$D_E \cdot x \leq d_E, \quad D_F \cdot x < d_F$$

eine Lösung hat. △

**Beweis.** Hausaufgabe. □

Wir können nun die Fourier-Motzkin-Elimination sukzessive auf alle Spaltenindizes anwenden. Ist in einem Schritt die neu konstruierte Matrix leer, so ist das System konsistent. Wir wollen herausfinden, was passiert, wenn  $Ax \leq b$  nicht konsistent ist. Wir nehmen daher an, dass keine der sukzessiv konstruierten Matrizen leer ist.

Eliminieren wir die 1. Spalte von  $A$ , so erhalten wir nach Satz (10.4) eine  $(r_1, n)$ -Matrix  $D_1$  und einen Vektor  $d_1 \in \mathbb{K}^{r_1}$  mit

$$Ax \leq b \text{ ist konsistent} \iff D_1x \leq d_1 \text{ ist konsistent.}$$

Die erste Spalte von  $D_1$  ist eine Nullspalte. Nun eliminieren wir die zweite Spalte von  $D_1$  und erhalten mit Satz (10.4) eine  $(r_2, n)$ -Matrix  $D_2$  und einen Vektor  $d_2 \in \mathbb{K}^{r_2}$  mit

$$D_1x \leq d_1 \text{ ist konsistent} \iff D_2x \leq d_2 \text{ ist konsistent.}$$

Die erste und die zweite Spalte von  $D_2$  bestehen aus lauter Nullen. Wir eliminieren nun die dritte Spalte von  $D_2$  und fahren auf diese Weise fort, bis wir die  $n$ -te Spalte eliminiert haben. Insgesamt erhalten wir:

$$\begin{aligned} Ax \leq b \text{ ist konsistent} &\iff D_1x \leq d_1 \text{ ist konsistent} \iff \dots \\ &\iff D_nx \leq d_n \text{ ist konsistent.} \end{aligned}$$

Was haben wir durch diese äquivalenten Umformungen gewonnen? Nach Konstruktion hat die  $n$ -te Matrix  $D_n$  dieser Folge von Matrizen nur noch Nullspalten, ist also eine Nullmatrix. Das System  $D_nx \leq d_n$  ist also nichts anderes als  $0x \leq d_n$ , und die Konsistenz dieses letzten Ungleichungssystems ist äquivalent zur Konsistenz von  $Ax \leq b$ . Die Konsistenz von  $0x \leq d_n$  ist trivial überprüfbar, denn  $0x \leq d_n$  hat genau dann keine Lösung, wenn der Vektor  $d_n \in \mathbb{K}^{r_n}$  eine negative Komponente hat; das heißt,  $d_n \geq 0$  ist äquivalent zur Konsistenz von  $Ax \leq b$ .

Aus Satz (10.4) folgt außerdem: Es existiert eine Matrix  $U_1 \in \mathbb{K}^{(r_1, m)}$  mit  $U_1 \geq 0$  und

$$D_1 = U_1A, \quad d_1 = U_1b.$$

## 10 Fourier-Motzkin-Elimination und Projektion

Erneute Anwendung von (10.4) liefert die Existenz einer Matrix  $U_2 \in \mathbb{K}^{(r_2, r_1)}$  mit  $U_2 \geq 0$  und

$$D_2 = U_2 D_1, \quad d_2 = U_2 d_1.$$

Setzen wir die Schlussweise fort, so erhalten wir eine Matrix  $U_n \in \mathbb{K}^{(r_n, r_{n-1})}$  mit  $U_n \geq 0$  und

$$0 = D_n = U_n D_{n-1}, \quad d_n = U_n d_{n-1}.$$

Für die Matrix

$$U := U_n \cdots U_1 \in \mathbb{K}^{(r_n, m)}$$

gilt dann

$$U \geq 0, \quad 0 = UA, \quad d_n = Ub.$$

Daraus folgt, dass jede Zeile der Matrix  $D_n$ , welche die Nullmatrix ist, eine konische Kombination von Zeilen von  $A$  ist.

Wir haben uns oben überlegt, dass  $Ax \leq b$  genau dann nicht konsistent ist, wenn  $D_n x \leq d_n$  nicht konsistent ist. Dies ist genau dann der Fall, wenn es eine Komponente von  $d_n$  gibt, die negativ ist, anders ausgedrückt, wenn es einen Vektor  $u$  (eine Zeile der Matrix  $U$ ) gibt mit  $0^T = u^T A$  und  $u^T b < 0$ . Fassen wir diese Beobachtung zusammen.

**(10.8) Folgerung.** *Es seien  $A \in \mathbb{K}^{(m, n)}$  und  $b \in \mathbb{K}^m$ , dann gilt: Das Ungleichungssystem  $Ax \leq b$  hat genau dann keine Lösung, wenn es einen Vektor  $u \in \mathbb{K}^m$ ,  $u \geq 0$  gibt mit  $u^T A = 0^T$  und  $u^T b < 0$ .*  $\triangle$

Diese (einfache) Konsequenz aus unserem Eliminationsverfahren (10.1) ist eines der nützlichsten Resultate der Polyedertheorie.

### Ergänzende Abschweifung

Das Fourier-Motzkin-Eliminationsverfahren ist ein Spezialfall eines allgemeinen Projektionsverfahrens auf lineare Teilräume des  $\mathbb{K}^n$ .

**(10.9) Definition.** *Sei  $L$  ein linearer Unterraum von  $\mathbb{K}^n$ . Ein Vektor  $\bar{x} \in \mathbb{K}^n$  heißt orthogonale Projektion eines Vektors  $x \in \mathbb{K}^n$  auf  $L$ , falls  $\bar{x} \in L$  und  $(x - \bar{x})^T \bar{x} = 0$  gilt.*  $\triangle$

**(10.10) Bemerkung.** Sei  $B \in \mathbb{K}^{(m, n)}$  eine Matrix mit vollem Zeilenrang,  $L = \{x \in \mathbb{K}^n \mid Bx = 0\}$ , dann ist für jeden Vektor  $x \in \mathbb{K}^n$ , der Vektor

$$\bar{x} := (I - B^T(BB^T)^{-1}B)x$$

die orthogonale Projektion auf  $L$ .  $\triangle$

Die Fourier-Motzkin-Elimination der  $j$ -ten Variablen kann man, wie wir jetzt zeigen, als orthogonale Projektion von  $P(A, b)$  auf den Vektorraum  $L = \{x \mid x_j = 0\}$  deuten. Wir zeigen nun, wie man ein Polyeder  $P(A, b)$  auf  $L = \{y \mid c^T y = 0\}$ ,  $c \neq 0$ , projiziert.

**(10.11) Algorithmus.** Orthogonale Projektion eines Polyeders auf

$$L = \{y \in \mathbb{K}^n \mid c^T y = 0\}, \quad c \neq 0$$

(kurz: Projektion entlang  $c$ ).

**Eingabe:** Ein Vektor  $c \in \mathbb{K}^n$ ,  $c \neq 0$  (die Projektionsrichtung), eine Matrix  $A \in \mathbb{K}^{(m,n)}$ , ein Vektor  $b \in \mathbb{K}^m$ .

**Ausgabe:** Eine Matrix  $D \in \mathbb{K}^{(r,n)}$  ( $r$  wird im Algorithmus berechnet), ein Vektor  $d \in \mathbb{K}^r$ .

1. Partitioniere die Menge  $M = \{1, \dots, m\}$  der Zeilenindizes von  $A$  wie folgt

$$N := \{i \in M \mid A_i \cdot c < 0\},$$

$$Z := \{i \in M \mid A_i \cdot c = 0\},$$

$$P := \{i \in M \mid A_i \cdot c > 0\}.$$

2. Setze

$$r := |Z \cup N \times P|,$$

$$R := \{1, 2, \dots, r\},$$

und  $p: R \rightarrow Z \cup N \times P$  sei eine Bijektion.

3. Für  $i = 1, 2, \dots, r$  führe aus:

(a) Ist  $p(i) \in Z$ , dann setze

$$D_{i \cdot} := A_{p(i) \cdot}, \quad d_i = b_{p(i)}.$$

(D. h. die  $i$ -te Zeile von  $D$  ist gleich der  $p(i)$ -ten Zeile von  $A$ .)

(b) Ist  $p(i) = (s, t) \in N \times P$ , dann setze

$$D_{i \cdot} := (A_t \cdot c)A_s - (A_s \cdot c)A_t,$$

$$d_i := (A_t \cdot c)b_s - (A_s \cdot c)b_t.$$

4. Gib  $D$  und  $d$  aus. △

**(10.12) Satz.** Seien  $D \in \mathbb{K}^{(r,n)}$  und  $d \in \mathbb{K}^r$  die in (10.11) konstruierte Matrix bzw. Vektor. Dann gilt

(a) Die Zeilen von  $D$  sind orthogonal zu  $c$ , d. h., die Zeilenvektoren von  $D$  sind in  $L = \{y \in \mathbb{K}^n \mid c^T y = 0\}$  enthalten und konische Kombinationen der Zeilen von  $A$ .

(b)  $P(D, d) \cap L$  ist die orthogonale Projektion von  $P(A, b)$  auf  $L$ . △

**(10.13) Satz.** Die Projektion eines Polyeders ist ein Polyeder! △

Durch sukzessive Projektion kann man ein Polyeder auf einen beliebigen Unterraum  $\{y \in \mathbb{K}^n \mid By = 0\}$  des  $\mathbb{K}^n$  projizieren.

## 10.2 Lineare Optimierung und Fourier-Motzkin-Elimination

Die Fourier-Motzkin-Elimination ist nicht nur eine Projektionsmethode, man kann sie auch zur Lösung linearer Programme verwenden. Dies geht wie folgt.

Wir beginnen mit einem linearen Programm der Form

$$\begin{aligned} \max \quad & c^T x \\ & Ax \leq a \end{aligned}$$

mit  $A \in \mathbb{K}^{(m,n)}$ ,  $a \in \mathbb{K}^m$  und setzen  $P = P(A, a)$ . Wir führen nun eine zusätzliche Variable  $x_{n+1}$  ein und fügen zur Matrix  $A$  eine zusätzliche Spalte und Zeile hinzu. Diese neue Matrix nennen wir  $B$ . Die  $(m+1)$ -te Zeile von  $B$  enthält in den ersten  $n$  Spalten den Vektor  $c$ , d. h.  $b_{m+1,j} = c_j$ ,  $j = 1, \dots, n$ ; wir setzen  $b_{i,n+1} = 0$  für  $i = 1, \dots, m$  und  $b_{m+1,n+1} = -1$ . Wir verlängern den Vektor  $a \in \mathbb{K}^m$  um eine Komponente, die den Wert 0 enthält, und nennen diesen  $(m+1)$ -Vektor  $b$ :

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

Wenden wir die Fourier-Motzkin-Elimination auf  $Ax \leq a$  an, so können wir feststellen, ob die Lösungsmenge  $P$  des linearen Programms leer ist oder nicht. Führen wir die Fourier-Motzkin-Elimination mit  $Bx \leq b$  durch und eliminieren wir nur die ersten  $n$  Variablen, so erhalten wir am Ende ein System  $D_n x \leq d_n$ , welches nichts anderes ist als ein Ungleichungssystem der Form  $\alpha_i \leq x_{n+1} \leq \beta_j$ , wobei die  $\beta_j$  die positiven und die  $\alpha_i$  die negativen Einträge des Vektors  $d_n$  sind. Das Minimum über die Werte  $\beta_j$  liefert den maximalen Wert der Zielfunktion  $c^T x$  des gegebenen linearen Programms.

Setzen wir  $\bar{x}_i := 0$ ,  $i = 1, \dots, n$  und  $\bar{x}_{n+1} := \min\{\beta_j\}$ , so können wir mit diesem Vektor  $\bar{x} = (\bar{x}_i)$  beginnend durch Rücksubstitution – wie in Satz (10.4)(d) beschrieben – eine Optimallösung von  $\max c^T x$ ,  $Ax \leq a$  berechnen.

Wenn wir statt des Maximums von  $c^T x$  über  $P$  das Minimum finden wollen, so fügen wir bei der Konstruktion der Matrix  $B$  statt der Zeile  $(c^T, -1)$  die Zeile  $(-c^T, 1)$  ein. Wollen wir das Maximum und Minimum gleichzeitig bestimmen, so führen wir die Fourier-Motzkin-Elimination der ersten  $n$  Variablen auf der Matrix durch, die um die beiden Zeilen  $(c^T, -1)$  und  $(-c^T, 1)$  erweitert wurde.

Dieses Verfahren zur Lösung linearer Programme ist konzeptionell extrem einfach und klar. Versuchen Sie einmal, damit lineare Programme „anständiger Größenordnung“ zu lösen. Sie werden sich wundern und wissen dann, warum Fourier-Motzkin-Elimination nicht zur linearen Optimierung benutzt wird.

Wir beschließen die Diskussion der Fourier-Motzkin-Elimination als Methode zur Lösung linearer Programme durch numerische Berechnung eines Beispiels.

**(10.14) Beispiel.** Wir beginnen mit folgendem System von 5 Ungleichungen mit 2 Variablen:

$$\begin{aligned} -x_2 &\leq 0 & (1) \\ -x_1 - x_2 &\leq -1 & (2) \\ -x_1 + x_2 &\leq 3 & (3) \\ +x_1 &\leq 3 & (4) \\ +x_1 + 2x_2 &\leq 9 & (5) \end{aligned}$$

Wir wollen die Zielfunktion

$$x_1 + 3x_2$$

über der Menge der zulässigen Lösungen sowohl maximieren als auch minimieren. Dazu schreiben wir (wie oben erläutert) das folgende Ungleichungssystem

$$\begin{aligned} -x_2 &\leq 0 & (1) \\ -x_1 - x_2 &\leq -1 & (2) \\ -x_1 + x_2 &\leq 3 & (3) \\ +x_1 &\leq 3 & (4) \\ +x_1 + 2x_2 &\leq 9 & (5) \\ +x_1 + 3x_2 - x_3 &\leq 0 & (6) \\ -x_1 - 3x_2 + x_3 &\leq 0 & (7) \end{aligned}$$

auf und eliminieren die Variable  $x_1$ . Das Ergebnis ist das folgende Ungleichungssystem:

$$\begin{aligned} (1) \quad -x_2 &\leq 0 & (1) \\ (2, 4) \quad -x_2 &\leq 2 & (2) \\ (2, 5) \quad +x_2 &\leq 8 & (3) \\ (2, 6) \quad +2x_2 - x_3 &\leq -1 & (4) \\ (3, 4) \quad +x_2 &\leq 6 & (5) \\ (3, 5) \quad +x_2 &\leq 4 & (6) \\ (3, 6) \quad +4x_2 - x_3 &\leq 3 & (7) \\ (7, 4) \quad -3x_2 + x_3 &\leq 3 & (8) \\ (7, 5) \quad -x_2 + x_3 &\leq 9 & (9) \end{aligned}$$

Wir haben oben die Variable  $x_1$  weggelassen. Die erste Spalte zeigt an, woher die neue Ungleichung kommt. So ist z. B. die 5. Ungleichung aus der Kombination der 3. und 4. Ungleichung des vorhergehenden Systems entstanden. Ungleichungen der Form  $0x_2 + 0x_3 \leq \alpha$  haben wir weggelassen. Eine solche ist z. B. die Ungleichung, die aus der Kombination der 7. und 6. Ungleichung entsteht.

## Literaturverzeichnis

Die Elimination der Variablen  $x_2$  liefert nun analog

$$(1, 4) \quad -x_3 \leq -1 \quad (1)$$

$$(1, 7) \quad -x_3 \leq 3 \quad (2)$$

$$(2, 4) \quad -x_3 \leq 3 \quad (3)$$

$$(2, 7) \quad -x_3 \leq 11 \quad (4)$$

$$(8, 3) \quad +x_3 \leq 27 \quad (5)$$

$$(8, 4) \quad -x_3 \leq 3 \quad (6)$$

$$(8, 5) \quad +x_3 \leq 21 \quad (7)$$

$$(8, 6) \quad +x_3 \leq 15 \quad (8)$$

$$(8, 7) \quad +x_3 \leq 21 \quad (9)$$

$$(9, 3) \quad +x_3 \leq 17 \quad (10)$$

$$(9, 4) \quad +x_3 \leq 17 \quad (11)$$

$$(9, 5) \quad +x_3 \leq 15 \quad (12)$$

$$(9, 6) \quad +x_3 \leq 13 \quad (13)$$

$$(9, 7) \quad +3x_3 \leq 39 \quad (14)$$

Die größte untere Schranke für  $x_3$  ist 1. Sie wird von der ersten Ungleichung geliefert. Die kleinste obere Schranke hat den Wert 13, dieser folgt aus den Ungleichungen (13) und (14). Damit ist 13 der Maximalwert der Zielfunktion, 1 ist der Minimalwert. Setzen wir  $x_3 = 13$  in das vorhergehende Ungleichungssystem ein, so ergibt sich der Wert  $x_2 = 4$ . Durch Substitution dieser beiden Werte in das Ausgangssystem erhalten wir  $x_1 = 1$ . Der maximale Zielfunktionswert wird also im zulässigen Punkt  $x_1 = 1, x_2 = 4$  angenommen.  $\triangle$

## Literaturverzeichnis

A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, New York, 1998.

J. Stoer and C. Witzgall. *Convexity and optimization in finite dimensions*. Springer, Berlin, 1970.

Detaillierte Hinweise auf „alte“ Literatur zum Thema und ihre Auffindbarkeit wie Nachdrucke und Übersetzungen finden sich insbesondere in dem Buch von Schrijver (1998), so auch die Fundstellen für Fourier (1827) und Motzkin (1936).

# 11 Das Farkas-Lemma und Dualitätstheorie

## 11.1 Verschiedene Versionen des Farkas-Lemmas

(11.1) Satz ((Allgemeines) Farkas-Lemma). Für dimensionsverträgliche Matrizen  $A, B, C, D$  und Vektoren  $a, b$  gilt:

$$\left. \begin{array}{l} \text{Es existieren } x, y \text{ mit} \\ Ax + By \leq a \\ Cx + Dy = b \\ x \geq 0 \end{array} \right\} \dot{\vee} \left\{ \begin{array}{l} \text{Es existieren } u, v \text{ mit} \\ u^T A + v^T C \geq 0^T \\ u^T B + v^T D = 0^T \\ u \geq 0 \\ u^T a + v^T b < 0. \end{array} \right.$$

(Hierbei bezeichnet „ $\dot{\vee}$ “ das „entweder oder“, d. h. eine der beiden Aussagen gilt, aber niemals beide gleichzeitig, also eine Alternative.) △

**Beweis.** Durch die Transformationsregeln (2.4) führen wir die obige Aussage auf Folgerung (10.8) zurück. Die linke Aussage der Alternative lässt sich schreiben als „ $\exists z$  mit  $\bar{A}z \leq \bar{a}$ “ wobei

$$\bar{A} := \begin{pmatrix} A & B \\ C & D \\ -C & -D \\ -I & 0 \end{pmatrix} \quad \text{und} \quad \bar{a} = \begin{pmatrix} a \\ b \\ -b \\ 0 \end{pmatrix}.$$

Nach (10.8) hat dieses System genau dann keine Lösung, wenn gilt:

$$\exists y^T = (u^T, \bar{v}^T, \bar{\bar{v}}^T, w^T) \geq 0$$

mit  $y^T \bar{A} = 0^T$  und  $y^T \bar{a} < 0$ . Ausführlich geschrieben heißt dies:

$$\exists \begin{pmatrix} u \\ \bar{v} \\ \bar{\bar{v}} \\ w \end{pmatrix} \geq 0 \quad \text{mit} \quad \left\{ \begin{array}{l} u^T A + \bar{v}^T C - \bar{\bar{v}}^T C - w^T = 0^T \\ u^T B + \bar{v}^T D - \bar{\bar{v}}^T D = 0^T \\ u^T a + \bar{v}^T b - \bar{\bar{v}}^T b < 0. \end{array} \right.$$

Setzen wir  $v := \bar{v} - \bar{\bar{v}}$  und betrachten wir  $w$  als einen Vektor von Schlupfvariablen, so

lässt sich dieses System äquivalent schreiben als:

$$\exists u, v \quad \text{mit} \quad \begin{cases} u^T A + v^T C \geq 0^T \\ u^T B + v^T D = 0^T \\ u \geq 0 \\ u^T a + v^T b < 0, \end{cases}$$

und dies ist das gewünschte Resultat. □

Durch Spezialisierung von (11.1) erhalten wir:

**(11.2) Satz ((Spezielles) Farkas-Lemma).** *Es seien  $A \in \mathbb{K}^{(m,n)}$  und  $b \in \mathbb{K}^m$ , dann gilt:*

- (a)  $\exists x \quad \text{mit } Ax \leq b \quad \dot{\vee} \quad \exists u \geq 0 \quad \text{mit } u^T A = 0^T \quad \text{und } u^T b < 0.$
- (b)  $\exists x \geq 0 \quad \text{mit } Ax \leq b \quad \dot{\vee} \quad \exists u \geq 0 \quad \text{mit } u^T A \geq 0^T \quad \text{und } u^T b < 0.$
- (c)  $\exists x \geq 0 \quad \text{mit } Ax = b \quad \dot{\vee} \quad \exists u \quad \text{mit } u^T A \geq 0^T \quad \text{und } u^T b < 0.$
- (d)  $\exists x \quad \text{mit } Ax = b \quad \dot{\vee} \quad \exists u \quad \text{mit } u^T A = 0^T \quad \text{und } u^T b < 0. \quad \triangle$

Das Farkas-Lemma (11.2) in seiner Version (a) charakterisiert also

(a) *Die Lösbarkeit eines linearen Ungleichungssystems,*

in seiner Version (11.2)(b)

(b) *die nichtnegative Lösbarkeit eines linearen Ungleichungssystems,*

in seiner Version (11.2)(c)

(c) *die nichtnegative Lösbarkeit eines linearen Gleichungssystems,*

in seiner Version (11.2)(d)

(d) *die Lösbarkeit eines linearen Gleichungssystems,*

wobei (d) natürlich bereits aus der linearen Algebra bekannt ist. Die zweite Alternative in (d) ist nichts anderes als eine Umformulierung von  $\text{rang}(A) \neq \text{rang}(A, b)$ .

Die linken Alternativen in (11.1) und (11.2)(a)–(d) sagen aus, dass gewisse Polyeder nicht leer sind. Die Lösungsmengen der rechten Alternativen sind dagegen keine Polyeder, weil jeweils eine strikte Ungleichung in den Gleichungs- und Ungleichungssystemen vorkommt. Da in allen Fällen die rechten Seiten der rechten Alternativen Nullvektoren sind, sind die Lösungsmengen (ohne die strikte Ungleichung) nach (2.8) Kegel. Folglich können die Lösungsvektoren skaliert werden. Aus dieser Beobachtung folgt:

**(11.3) Satz (Farkas-Lemma (polyedrische Version)).** *Es seien  $A \in \mathbb{K}^{(m,n)}$  und  $b \in \mathbb{K}^m$ , dann gilt:*

- (a)  $P(A, b) \neq \emptyset \quad \dot{\vee} \quad P^= \left( \begin{pmatrix} A^T \\ b^T \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right) \neq \emptyset.$
- (b)  $P^=(A, b) \neq \emptyset \quad \dot{\vee} \quad P \left( \begin{pmatrix} -A^T \\ b^T \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right) \neq \emptyset. \quad \triangle$

Polyedrische Formulierungen von (11.2)(b), (d) und (11.1) seien dem Leser überlassen.



## 11.2 Alternativ- und Transpositionssätze

Wir haben das Farkas-Lemma nun in verschiedenen Versionen formuliert. Folgerung (10.8) beschreibt ein zur Unlösbarkeit eines Ungleichungssystems äquivalentes Kriterium. Die Sätze (11.1) und (11.2) haben die Form von Alternativen, während in (11.3) die Lösbarkeit eines Systems mit der Unlösbarkeit eines durch Transposition gewonnenes System festgestellt wird. Man nennt deshalb Sätze des Typs (11.1), (11.2) bzw. (11.3) Alternativ- bzw. Transpositionssätze. Die Literatur ist außerordentlich reich an derartigen Sätzen, da sie – wie wir später noch sehen werden – auf vielerlei Weise nützlich sind. Wir wollen hier noch einige weitere Sätze dieser Art angeben.

**(11.4) Satz.** *Es seien  $A \in \mathbb{K}^{(p,n)}$ ,  $B \in \mathbb{K}^{(q,n)}$ ,  $a \in \mathbb{K}^p$ ,  $b \in \mathbb{K}^q$ , dann gilt genau eine der beiden folgenden Alternativen:*

- (a)  $\exists x \in \mathbb{K}^n$  mit  $Ax \leq a$ ,  $Bx < b$
- (b) (b<sub>1</sub>)  $\exists u \in \mathbb{K}_+^p, v \in \mathbb{K}_+^q \setminus \{0\}$  mit  $u^T A + v^T B = 0^T$ ,  $u^T a + v^T b \leq 0$ , oder  
 (b<sub>2</sub>)  $\exists u \in \mathbb{K}_+^p$  mit  $u^T A = 0^T$ ,  $u^T a < 0$ . △

**Beweis.** Angenommen, (a) und (b<sub>2</sub>) gelten gleichzeitig, dann gibt es also einen Vektor  $x$  mit  $Ax \leq a$  und einen Vektor  $u \in \mathbb{K}_+^p$  mit  $u^T A = 0^T$ ,  $u^T a < 0$ , was (11.2)(a) widerspricht. Angenommen, (a) und (b<sub>1</sub>) gelten gleichzeitig, dann gilt:  $0^T x = (u^T A + v^T B)x = u^T(Ax) + v^T(Bx) < u^T a + v^T b \leq 0$ , ein Widerspruch.

Wir nehmen nun an, dass (a) nicht gilt und wollen zeigen, dass dann (b) gilt. Wir wenden die Fourier-Motzkin-Elimination  $n$ -mal iterativ auf  $\begin{pmatrix} A \\ B \end{pmatrix}$  und  $\begin{pmatrix} a \\ b \end{pmatrix}$  an. Nach (10.7) und erhalten wir nach  $n$  Schritten Matrizen  $C$ ,  $D$  und Vektoren  $c$ ,  $d$ , so dass  $Ax \leq a$ ,  $Bx < b$  genau dann lösbar ist, wenn  $Cx \leq c$ ,  $Dx < d$  lösbar ist, wobei  $C = 0$ ,  $D = 0$  gilt. Nach Annahme gilt (a) nicht, also muss es einen Zeilenindex  $i$  von  $C$  geben mit  $c_i < 0$  oder einen Zeilenindex  $j$  von  $D$  mit  $d_j \leq 0$ .

Wendet man die Fourier-Motzkin-Elimination nur auf  $A$ ,  $a$  und  $n$ -mal sukzessiv an, so erhält man nach (10.7) das System  $C$ ,  $c$ . Das heißt, die Lösbarkeit von  $Ax \leq a$  ist äquivalent zur Lösbarkeit von  $Cx \leq c$ . Ist also  $Ax \leq a$  nicht lösbar, so gibt es nach (10.8) einen Vektor  $u \geq 0$  mit  $u^T A = 0^T$ ,  $u^T a < 0$ , das heißt (b<sub>2</sub>) gilt. Ist  $Ax \leq a$  lösbar, so gibt es keinen Vektor  $u \geq 0$  mit  $u^T A = 0^T$ ,  $u^T a < 0$ , d. h.  $c_i \geq 0$  für alle  $i$ . Folglich muss  $d_j \leq 0$  für ein  $j$  gelten. Die Zeile  $D_j \cdot = 0^T$  ist eine konische Kombination von Zeilen von  $A$  und Zeilen von  $B$ . Nach Definition, siehe (10.7), muss dabei mindestens eine Zeile von  $B$  mit einem positiven Multiplikator beteiligt sein, also gibt es  $u \in \mathbb{K}_+^p$ ,  $v \in \mathbb{K}_+^q \setminus \{0\}$  mit  $u^T A + v^T B = D_j \cdot = 0^T$  und  $u^T a + v^T b = d_j \leq 0$ , das heißt (b<sub>1</sub>) gilt. □

**(11.5) Folgerung.** *Ist  $P(A, a) \neq \emptyset$ , dann gilt genau eine der beiden folgenden Alternativen:*

- (a)  $\exists x \in \mathbb{K}^n$  mit  $Ax \leq a$ ,  $Bx < b$ .
- (b)  $\exists \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{K}_+^{p+q}$ ,  $v \neq 0$  mit  $u^T A + v^T B = 0^T$  und  $u^T a + v^T b \leq 0$ . △

**(11.6) Satz (Gordan (1873)).** *Es gilt genau eine der beiden folgenden Alternativen:*

- (a)  $\exists x$  mit  $Ax < 0$ .
- (b)  $\exists u \geq 0, u \neq 0$  mit  $u^T A = 0^T$ . △

**Beweis.** Setze in (11.5)  $B := A$  (wobei  $A$  die obige Matrix ist),  $b := 0, A := 0, a := 0$ . □

**(11.7) Satz (Stiemke (1915)).** *Es gilt genau eine der folgenden Alternativen:*

- (a)  $\exists x > 0$  mit  $Ax = 0$ .
- (b)  $\exists u$  mit  $u^T A \geq 0^T, u^T A \neq 0$ . △

**Beweis.** Setze in (11.5)  $B := -I, b = 0, A := \begin{pmatrix} A \\ -A \end{pmatrix}, a := 0$ . □

Der Satz von Stiemke charakterisiert also die strikt positive Lösbarkeit eines homogenen Gleichungssystems, während der Satz von Gordan die semipositive Lösbarkeit des homogenen Gleichungssystems  $u^T A = 0$  kennzeichnet. Eine Sammlung weiterer Alternativsätze kann man in Mangasarian (1969) finden.

### 11.3 Trennsätze

So genannte Trennsätze spielen in der Konvexitätstheorie und – in noch allgemeinerer Form – in der Funktionalanalysis eine wichtige Rolle. Diese Sätze sind i. A. vom folgenden Typ: Seien  $S$  und  $T$  zwei Mengen, dann gibt es unter gewissen Voraussetzungen eine Hyperebene  $H$ , die  $S$  und  $T$  trennt. Geometrisch heißt dies, dass  $S$  auf der einen,  $T$  auf der anderen Seite von  $H$  liegt. Genauer: Sei  $H = \{x \mid c^T x = \gamma\}, c \neq 0$ , eine Hyperebene, dann sagen wir, dass  $H$  zwei Mengen  $P$  und  $Q$  trennt, falls  $P \cup Q \not\subseteq H$  und  $P \subseteq \{x \mid c^T x \leq \gamma\}, Q \subseteq \{x \mid c^T x \geq \gamma\}$  gilt. Wir sagen, dass  $H$  die Mengen  $P$  und  $Q$  strikt trennt, falls  $P \subseteq \{x \mid c^T x < \gamma\}$  und  $Q \subseteq \{x \mid c^T x > \gamma\}$ .

Das Farkas-Lemma impliziert einige interessante Trennsätze für Polyeder  $P \subseteq \mathbb{K}^n$ . Wir wollen hier einen Satz dieser Art angeben, der als Prototyp für die allgemeineren Trennsätze angesehen werden kann.

**(11.8) Satz (Trennsatz für Polyeder).** *Es seien  $P = P(A, a)$  und  $Q = P(B, b)$  zwei Polyeder im  $\mathbb{K}^n$ . Es gibt eine  $P$  und  $Q$  strikt trennende Hyperebene genau dann, wenn  $P \cap Q = \emptyset$  gilt, es sei denn, eines der Polyeder ist leer und das andere der  $\mathbb{K}^n$ . △*

**Beweis.** Gibt es eine  $P$  und  $Q$  strikt trennende Hyperebene, dann ist  $P \cap Q$  offensichtlich leer.

Sei  $P \cap Q = \emptyset$ . Zu zeigen ist die Existenz eines Vektors  $c \in \mathbb{K}^n, c \neq 0$ , und einer Zahl  $\gamma \in \mathbb{K}$ , so dass  $P \subseteq \{x \mid c^T x < \gamma\}$  und  $Q \subseteq \{x \mid c^T x > \gamma\}$  gilt. Sind  $P$  und  $Q$  leer, so leistet jede beliebige Hyperebene das Gewünschte. Ist eines der Polyeder leer, sagen wir

$Q = \emptyset$ ,  $P \neq \emptyset$ , dann ist wegen  $P \neq \mathbb{K}^n$  eine der Zeilen von  $A$  von Null verschieden, sagen wir  $A_i \neq 0$ . Dann haben  $c^T := A_i$ ,  $\gamma := a_i + 1$  die geforderten Eigenschaften. Sind  $P$  und  $Q$  nicht leer, dann gilt nach Voraussetzung

$$P \left( \begin{pmatrix} A \\ B \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix} \right) = \emptyset.$$

Folglich existiert nach (11.2)(a) ein Vektor  $\begin{pmatrix} u \\ v \end{pmatrix} \geq 0$  mit  $u^T A + v^T B = 0^T$  und  $u^T a + v^T b < 0$ . Wir setzen

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

Daraus folgt:

$$x \in P \implies c^T x = u^T A x \leq u^T a < u^T a - \frac{1}{2}(u^T a + v^T b) = \frac{1}{2}(u^T a - v^T b) = \gamma.$$

$$x \in Q \implies c^T x = -v^T B x \geq -v^T b > -v^T b + \frac{1}{2}(u^T a + v^T b) = \frac{1}{2}(u^T a - v^T b) = \gamma.$$

Der Vektor  $c$  ist offenbar von Null verschieden.  $\square$

Der folgende Satz, den wir nicht mit den bisher entwickelten Methoden beweisen können, soll als Beispiel für die oben angesprochene allgemeine Art von Trennsätzen dienen.

**(11.9) Satz (Trennsatz für konvexe Mengen).** *Es seien  $P$  und  $Q$  zwei nichtleere konvexe Teilmengen des  $\mathbb{R}^n$ . Es gibt eine  $P$  und  $Q$  trennende Hyperebene  $H = \{x \in \mathbb{R}^n \mid c^T x = \gamma\}$  (d. h.  $P \cup Q \not\subseteq H$ ,  $P \subseteq \{x \in \mathbb{R}^n \mid c^T x \leq \gamma\}$ ,  $Q \subseteq \{x \in \mathbb{R}^n \mid c^T x \geq \gamma\}$ ) genau dann, wenn der Durchschnitt des relativ Inneren von  $P$  mit dem relativ Inneren von  $Q$  leer ist.  $\triangle$*

**Beweis.** Siehe Stoer and Witzgall (1970, Seite 98, Satz (3.3.9)) oder Leichtweiß (1980, Seite 28, Satz 3.1).  $\square$

Der aufmerksame Leser wird bemerkt haben, dass in (11.9) der Vektorraum  $\mathbb{R}^n$  (und nicht wie üblich  $\mathbb{K}^n$ ) benutzt wurde. In der Tat ist Satz (11.9) für konvexe Teilmengen von  $\mathbb{Q}^n$  im folgenden Sinne falsch. Es ist nicht immer möglich, zwei konvexe Mengen in  $\mathbb{Q}^n$ , deren relativ innere Mengen kein gemeinsames Element besitzen, durch eine Hyperebene  $c^T x = \gamma$  zu trennen, so dass der Vektor  $(c^T, \gamma)^T$  rational ist. Man betrachte nur  $P = \{x \in \mathbb{Q} \mid x < \sqrt{2}\}$ ,  $Q = \{x \in \mathbb{Q} \mid x > \sqrt{2}\}$ . Dieses Beispiel zeigt auch, dass die bisher entwickelten konstruktiven Beweistechniken, die ja für  $\mathbb{Q}$  und  $\mathbb{R}$  gleichermaßen arbeiten, nicht mächtig genug sind, um (11.9) abzuleiten. Offenbar muss das Vollständigkeitsaxiom der reellen Zahlen auf irgendeine Weise benutzt werden.

Satz (11.9) können wir allerdings für Polyeder beweisen. Zu seiner Formulierung müssen wir jedoch schärfere Anforderungen an die Darstellung der involvierten Polyeder stellen.

**(11.10) Satz (Schwacher Trennsatz).** *Seien  $P = \{x \in \mathbb{K}^n \mid Ax = a, Bx \leq b\}$ ,  $Q = \{x \in \mathbb{K}^n \mid Cx = c, Dx \leq d\}$  zwei nichtleere Polyeder, so dass keine der Ungleichungen  $Bx \leq b$  von allen Punkten in  $P$  und keine der Ungleichungen  $Dx \leq d$  von allen Punkten in  $Q$  mit Gleichheit erfüllt wird. Es gibt eine  $P$  und  $Q$  trennende Hyperebene genau dann, wenn  $R := \{x \in \mathbb{K}^n \mid Ax = a, Cx = c, Bx < b, Dx < d\}$  leer ist.  $\triangle$*

**Beweis.** Ist  $P \cap Q = \emptyset$ , so ist  $R = \emptyset$ , und die Behauptung folgt (in schärferer Form) aus Satz (11.8). Sei also  $P \cap Q \neq \emptyset$ .

Die Menge  $R$  ist offenbar die Lösungsmenge des Systems

$$Ax \leq a, -Ax \leq -a, Cx \leq c, -Cx \leq -c, Bx < b, Dx < d.$$

Dieses hat nach Folgerung (11.5) genau dann keine Lösung, wenn es einen Vektor

$$(\bar{u}^T, \bar{\bar{u}}^T, \bar{v}^T, \bar{\bar{v}}^T, w^T, y^T) \geq 0$$

mit  $(w^T, y^T) \neq 0^T$  gibt mit der Eigenschaft

$$\begin{aligned} \bar{u}^T A - \bar{\bar{u}}^T A + \bar{v}^T C - \bar{\bar{v}}^T C + w^T B + y^T D &= 0^T, \\ \bar{u}^T a - \bar{\bar{u}}^T a + \bar{v}^T c - \bar{\bar{v}}^T c + w^T b + y^T d &\leq 0. \end{aligned}$$

Wir setzen  $u := \bar{u} - \bar{\bar{u}}$ ,  $v := \bar{v} - \bar{\bar{v}}$  und

$$\begin{aligned} r^T &:= u^T A + w^T B (= -v^T C - y^T D), \\ \rho &:= \frac{1}{2}(u^T a + w^T b - v^T c - y^T d). \end{aligned}$$

Sei  $H := \{x \mid r^T x = \rho\}$ , dann ist  $H$  eine  $P$  und  $Q$  trennende Hyperebene, denn

$$\begin{aligned} (1) \quad x \in P &\implies r^T x = u^T Ax + w^T Bx \leq u^T a + w^T b \\ &\leq u^T a + w^T b - \frac{1}{2}(u^T a + w^T b + v^T c + y^T d) \\ &= \rho \end{aligned}$$

$$\begin{aligned} (2) \quad x \in Q &\implies r^T x = -v^T Cx - y^T Dx \geq -v^T c - y^T d \\ &\geq -v^T c - y^T d + \frac{1}{2}(u^T a + w^T b + v^T c + y^T d) \\ &= \rho. \end{aligned}$$

$R = \emptyset$  ist also äquivalent zur Existenz der Hyperebene  $H = \{x \mid r^T x = \rho\}$ . Aus der Voraussetzung folgt, dass Punkte  $p \in P$  und  $q \in Q$  existieren mit  $Bp < b$  und  $Dq < d$ . Die Abschätzungen (1), (2) zeigen, dass aus  $(w^T, y^T) \neq 0^T$  folgt, dass  $r^T p < \rho$  oder  $r^T q > \rho$  gilt. Somit gilt  $r \neq 0$  und  $P \cup Q \not\subseteq H$ ; das aber heißt,  $H$  ist eine trennende Hyperebene.

Gibt es umgekehrt eine  $P$  und  $Q$  trennende Hyperebene  $H = \{x \mid r^T x = p\}$ , dann gibt es einen Punkt in  $P \cup Q$ , der nicht in  $H$  ist; sagen wir  $P \not\subseteq H$ . Daraus folgt  $P = \{x \mid Ax = a, Bx \leq b, r^T x \leq \rho\}$  und  $\{x \mid Ax = a, Bx < b\} = \{x \mid Ax = q, Bx < b, r^T x < \rho\}$  und somit ergibt sich, dass  $R = \{x \mid Ax = a, Cx = c, Bx < b, r^T x < \rho, Dx < d\}$  leer ist, denn  $\{x \mid Cx = c, Dx < d, r^T x < \rho\} = \emptyset$ , weil  $Q = \{x \mid Cx = c, Dx \leq d\} \subseteq \{x \mid r^T x \geq \rho\}$ .  $\square$

Satz (11.10) folgt aus (11.9), wenn man sich überlegt, dass für ein Polytop  $P = \{x \mid Ax = a, Bx \leq b\}$  mit der Eigenschaft, dass keine der Ungleichungen in  $Bx \leq b$  durch alle Punkte aus  $P$  mit Gleichheit erfüllt ist, das relativ Innere von  $P$  genau die Menge  $\{x \mid Ax = a, Bx < b\}$  ist.

## 11.4 Der Dualitätssatz der linearen Programmierung

In Satz (2.11) haben wir bereits den sogenannten schwachen Dualitätssatz angegeben und bewiesen. Die Gleichheit der optimalen Zielfunktionswerte kann man (unter anderem) mit Hilfe des Farkas-Lemmas beweisen. Man kann den (nachfolgend formulierten) Dualitätssatz als eine Optimierungsversion des Farkas-Lemmas ansehen. Beide Sätze sind in dem Sinne äquivalent, dass der eine ohne Mühe aus dem anderen gefolgert werden kann und umgekehrt. So wie das Farkas-Lemma für die Polyedertheorie von zentraler Bedeutung ist, ist der Dualitätssatz von außerordentlicher Tragweite in der linearen Optimierung und zwar sowohl in algorithmisch-praktischer als auch in theoretischer Hinsicht. Wenn man heutzutage in mathematischen (Forschungs-)Aufsätzen einen Satz wie „... it follows by an LP-argument...“ liest, dann heißt das so gut wie immer, dass der Dualitätssatz in einer seiner Versionen geschickt anzuwenden ist.

Wir wollen zunächst das folgende lineare Programm

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned} \tag{11.11}$$

wobei  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$  und  $c \in \mathbb{K}^n$  gegeben sind, untersuchen. Unsere Aufgabe besteht darin, unter den Vektoren  $x \in \mathbb{K}^n$ , die zulässig für (11.11) sind, d. h. die  $Ax \leq b$  erfüllen, einen Vektor  $x^* \in \mathbb{K}^n$  zu finden, der optimal ist, d. h.  $c^T x^* \geq c^T x$  für alle zulässigen  $x$ . In Kapitel 2 haben wir bereits eine Motivation für die Dualitätstheorie gegeben, hier wollen wir noch einen weiteren Zugang beschreiben. Zunächst wollen wir (11.11) polyedrisch darstellen. Wir führen dazu eine neue Variable  $z$  ein und schreiben (11.11) in der Form

$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & \begin{pmatrix} 1 & -c^T \\ 0 & A \end{pmatrix} \begin{pmatrix} z \\ x \end{pmatrix} \leq \begin{pmatrix} 0 \\ b \end{pmatrix} \end{aligned} \tag{11.12}$$

Daraus folgt, dass jedes lineare Programm in ein LP umgeformt werden kann, bei dem lediglich Lösungsvektoren gesucht werden, deren erste Komponente so groß wie möglich ist. Schreiben wir (11.12) noch einmal um und bringen wir  $z$  auf die rechte Seite, so kann (11.12) folgendermaßen geschrieben werden.

$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & \begin{pmatrix} -c^T \\ A \end{pmatrix} x \leq \begin{pmatrix} -z \\ b \end{pmatrix} \end{aligned} \tag{11.13}$$

Für festes  $z \in \mathbb{K}$  ist die Lösungsmenge von (11.13) ein Polyeder im  $\mathbb{K}^n$ . Wir setzen für alle  $z \in \mathbb{K}$

$$P_z := \left\{ x \in \mathbb{K}^n \mid \begin{pmatrix} -c^T \\ A \end{pmatrix} x \leq \begin{pmatrix} -z \\ b \end{pmatrix} \right\}. \tag{11.14}$$

Nunmehr können wir (11.13) wie folgt darstellen:

$$\max \{ z \mid P_z \neq \emptyset \}. \tag{11.15}$$

Wir suchen also ein  $z \in \mathbb{K}$ , das so groß wie möglich ist unter der Nebenbedingung, dass das Polyeder  $P_z \neq \emptyset$  ist.

Üblicherweise nennt man die Aufgabe zu zeigen, dass eine Menge nicht leer ist, ein *primales Problem*. Zu zeigen, dass eine Menge leer ist, ist ein *duales Problem*. Diese beiden Aufgaben sind i. A. nicht von gleicher Schwierigkeit. Betrachten wir z. B. ein Polyeder  $P \subseteq \mathbb{Q}^n$ . Legen wir einfach eine Abzählung  $x_1, x_2, x_3, \dots$  der Elemente von  $\mathbb{Q}^n$  fest und überprüfen wir, ob  $x_i \in P$  ist für  $i = 1, 2, \dots$ , so sind wir sicher, dass dieses Verfahren nach endlich vielen Schritten abbricht, falls  $P \neq \emptyset$  ist. Jedoch kann dieses Verfahren niemals nach endlich vielen Schritten folgern, dass  $P$  leer ist. Ein Ziel von Dualitätstheorien ist es, „gute Kriterien“ für die Lösung der dualen Probleme zu finden. Das Farkas-Lemma liefert ein solches. Gilt  $P = \{x \mid Ax \leq b\}$ , so ist nach (11.3)(a)  $P$  genau dann leer, wenn  $Q = P^{\left(\begin{pmatrix} A^T \\ b^T \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)}$  nicht leer ist. Durch Anwendung des obigen Abzählverfahrens auf  $P$  und  $Q$  gleichzeitig können wir also nach endlich vielen Schritten entscheiden, ob  $P$  leer ist oder nicht.

Das angegebene Abzählverfahren sollte nicht als ein ernsthafter Vorschlag zur algorithmischen Lösung des primalen und dualen Problems angesehen werden. Diese Bemerkungen sollen lediglich das Problembewusstsein des Lesers schärfen!

Zurück zu (11.15). Es kann sein, dass  $P_z \neq \emptyset$  gilt für alle  $z \in \mathbb{K}$ . Dann ist (11.11) unbeschränkt und hat keine Optimallösung. Andernfalls können wir versuchen den Maximalwert in (11.15) nach oben zu beschränken. Die bestmögliche Schranke ist natürlich gegeben durch

$$\inf\{z \mid P_z = \emptyset\}. \tag{11.16}$$

Nach (11.2)(a) ist  $P_z = \emptyset$  äquivalent dazu, dass das System  $y^T A = \lambda c^T$ ,  $y^T b < \lambda z$ ,  $y \geq 0$ ,  $\lambda \geq 0$  eine Lösung hat. Man überlegt sich leicht, dass dieses System – im Falle der Lösbarkeit von (11.11) – genau dann lösbar ist, wenn  $y^T A = c^T$ ,  $y^T b < z$ ,  $y \geq 0$  lösbar ist. Wenn wir das kleinste  $z$  in (11.16) finden wollen, müssen wir also einen möglichst kleinen Wert  $y^T b$  bestimmen. (11.16) ist somit äquivalent zu

$$\begin{aligned} \min \quad & y^T b \\ & y^T A = c^T \\ & y \geq 0 \end{aligned} \tag{11.17}$$

Problem (11.17) ist offenbar wiederum ein lineares Programm. Wir nennen es das *zum primalen Problem (11.11) duale lineare Programm*. Wir wollen nun zeigen, dass (11.11) und (11.17) den gleichen Optimalwert haben.

**(11.18) Satz (Dualitätssatz).** *Es seien  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$  und  $c \in \mathbb{K}^n$ , dann haben die beiden linearen Programme*

$$\begin{aligned} \max \quad & c^T x \\ & Ax \leq b \end{aligned} \tag{P} \quad \text{und} \quad \begin{aligned} \min \quad & y^T b \\ & y^T A = c^T \\ & y \geq 0 \end{aligned} \tag{D}$$

optimale Lösungen, deren Zielfunktionswerte gleich sind, genau dann, wenn sie zulässige Lösungen besitzen.  $\triangle$

**Beweis.** Wenn (P) und (D) optimale Lösungen haben, dann haben sie auch zulässige. Haben (P) und (D) zulässige Lösungen, so gilt nach (2.11), dass der Wert von (P) nicht größer als der von (D) ist. (P) und (D) haben zulässige Lösungen mit gleichem Zielfunktionswert genau dann, wenn das System

$$\begin{aligned} \begin{pmatrix} 0 \\ b^T \end{pmatrix} y + \begin{pmatrix} A \\ -c^T \end{pmatrix} x &\leq \begin{pmatrix} b \\ 0 \end{pmatrix} \\ A^T y &= c \\ y &\geq 0 \end{aligned} \tag{1}$$

eine Lösung hat. Dies ist nach (11.1) äquivalent dazu, dass das System

$$\begin{aligned} zb^T + v^T A^T &\geq 0^T \\ u^T A - zc^T &= 0^T \\ u &\geq 0 \\ z &\geq 0 \\ u^T b + v^T c &< 0 \end{aligned} \tag{2}$$

keine Lösung hat. Nehmen wir an, dass (2) eine Lösung  $(u^T, v^T, z)$  hat. Gibt es eine Lösung von (2) mit  $z = 0$ , so hat nach (11.1) das System

$$\begin{aligned} Ax &\leq b \\ A^T y &= c \\ y &\geq 0 \end{aligned}$$

keine Lösung. Das aber heißt, dass (P) oder (D) keine zulässige Lösung hat. Widerspruch! Gibt es eine Lösung von (2) mit  $z > 0$ , so können wir durch Skalieren eine Lösung finden mit  $z = 1$ . Daraus folgt

$$0 > u^T b + v^T c \geq -u^T A v + v^T A^T u = 0.$$

Widerspruch! Dies impliziert, dass (2) inkonsistent ist. Also hat nach (11.1) das System (1) eine Lösung, und Satz (11.18) ist bewiesen.  $\square$

**(11.19) Folgerung.**  $P$  bzw.  $D$  seien die Lösungsmengen von (P) bzw. (D). Wir setzen

$$z^* := \begin{cases} +\infty, & \text{falls (P) unbeschränkt,} \\ -\infty, & \text{falls } P = \emptyset, \\ \max\{c^T x \mid x \in P\}, & \text{andernfalls,} \end{cases}$$

$$u^* := \begin{cases} -\infty, & \text{falls (D) unbeschränkt,} \\ +\infty, & \text{falls } D = \emptyset, \\ \min\{y^T b \mid y \in D\}, & \text{andernfalls.} \end{cases}$$

Dann gilt:

- (a)  $-\infty < z^* = u^* < +\infty \iff z^*$  endlich  $\iff u^*$  endlich.
- (b)  $z^* = +\infty \implies D = \emptyset$ .
- (c)  $u^* = -\infty \implies P = \emptyset$ .
- (d)  $P = \emptyset \implies D = \emptyset$  oder  $u^* = -\infty$ .
- (e)  $D = \emptyset \implies P = \emptyset$  oder  $z^* = +\infty$ . △

**Beweis.** (a) Aus  $z^* = u^*$  endlich folgt natürlich  $u^*$  und  $z^*$  endlich.

Sei  $z^*$  endlich, dann ist  $P \neq \emptyset$  und

$$\begin{aligned} -c^T x &\leq -z \\ Ax &\leq b \end{aligned} \tag{1}$$

ist für  $z = z^*$  lösbar, jedoch unlösbar für jedes  $z > z^*$ . Sei also  $z > z^*$ , dann ist nach (11.2)(a) das duale Problem

$$\begin{aligned} -uc^T + y^T A &= 0^T \\ -uz + y^T b &< 0 \\ u, y &\geq 0 \end{aligned} \tag{2}$$

lösbar. Hat dieses System eine Lösung mit  $u = 0$ , so hat  $y^T A = 0^T$ ,  $y^T b < 0$ ,  $y \geq 0$  eine Lösung. Also folgt aus (11.2)(a), dass  $P = \emptyset$ , ein Widerspruch. Folglich muss es eine Lösung von (2) geben mit  $u > 0$ . Durch Skalieren können wir eine derartige Lösung mit  $u = 1$  finden. Daraus folgt, dass das System

$$\begin{aligned} y^T b &< z \\ y^T A &= c^T \\ y &\geq 0 \end{aligned}$$

eine Lösung besitzt. Folglich ist  $D \neq \emptyset$ , und aus Satz (11.18) folgt die Behauptung.

Ist  $u^*$  endlich, so verläuft der Beweis analog.

- (b) Ist  $D \neq \emptyset$ , so impliziert der schwache Dualitätssatz (2.11)  $z^* \leq \min\{y^T b \mid y \in D\} = u^* < +\infty$ .
- (c) Analog zu (b).
- (d) Angenommen  $P = \emptyset$  und  $u^* > -\infty$ . Dann gilt entweder  $u^* = +\infty$  und somit  $D = \emptyset$  oder nach (a) folgt  $z^* = u^*$  endlich, also  $P \neq \emptyset$ . Widerspruch!
- (e) Analog zu (d). □



Man könnte vermuten, dass in (11.19)(b) und (c) auch die umgekehrte Richtung gilt. Die Aussagen (d) und (e) zeigen aber, dass beim Versuch eines Beweises der Rückrichtung Komplikationen auftreten können. Das nachfolgende Beispiel zeigt, dass es in der Tat zueinander duale lineare Programme gibt, die beide leere Lösungsmengen haben.

**(11.20) Beispiel.** Es seien

$$P = \left\{ x \in \mathbb{K}^2 \mid \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} x \leq \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\},$$

$$D = \left\{ y \in \mathbb{K}^2 \mid \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} y = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, y \geq 0 \right\}.$$

Dann sind die linearen Programme

$$\max_{x \in P} x_1 + x_2 \quad \text{und} \quad \min_{y \in D} -y_2$$

zueinander dual, und es gilt:

(a)  $P = \emptyset$ , da nach (11.2)(a) das System

$$\begin{aligned} (u_1, u_2) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} &= (0, 0) \\ u_1, u_2 &\geq 0 \\ -u_2 &< 0 \end{aligned}$$

eine Lösung hat (z. B.  $u_1 = u_2 = 1$ ).

(b)  $D = \emptyset$ , da nach (11.2)(c) das System

$$\begin{aligned} (v_1, v_2) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} &\geq (0, 0) \\ v_1 + v_2 &< 0, \end{aligned}$$

eine Lösung hat (z. B.  $v_1 = v_2 = -1$ ).

△

Wir haben bereits die Sprechweise „zueinander duale lineare Programme“ benutzt, ohne dafür eine Begründung zu geben. Sie erfolgt hiermit:

**(11.21) Bemerkung.** Gegeben seien dimensionsverträgliche Matrizen  $A, B, C, D$  und Vektoren  $a, b, c, d$ , dann gilt: Das zum *primale lineare Programm*

$$\begin{aligned} \max \quad & c^T x + d^T y \\ & Ax + By \leq a \\ & Cx + Dy = b \\ & x \geq 0 \end{aligned} \tag{11.22}$$

primales LP		duales LP	
(P <sub>1</sub> )	$\max c^T x, Ax \leq b, x \geq 0$	(D <sub>1</sub> )	$\min y^T b, y^T A \geq c^T, y \geq 0$
(P <sub>2</sub> )	$\min c^T x, Ax \geq b, x \geq 0$	(D <sub>2</sub> )	$\max y^T b, y^T A \leq c^T, y \geq 0$
(P <sub>3</sub> )	$\max c^T x, Ax = b, x \geq 0$	(D <sub>3</sub> )	$\min y^T b, y^T A \geq c^T$
(P <sub>4</sub> )	$\min c^T x, Ax = b, x \geq 0$	(D <sub>4</sub> )	$\max y^T b, y^T A \leq c^T$
(P <sub>5</sub> )	$\max c^T x, Ax \leq b$	(D <sub>5</sub> )	$\min y^T b, y^T A = c^T, y \geq 0$
(P <sub>6</sub> )	$\min c^T x, Ax \geq b$	(D <sub>6</sub> )	$\max y^T b, y^T A = c^T, y \geq 0$

Tabelle 11.1: Beziehung zwischen primalem und dualem LP für verschiedene Typen von LPs.

duale lineare Programm ist

$$\begin{aligned}
 \min \quad & u^T a + v^T b \\
 & u^T A + v^T C \geq c^T \\
 & u^T B + v^T D = d^T \\
 & u \geq 0.
 \end{aligned} \tag{11.23}$$

Das zum linearen Programm (11.23) duale Programm ist (11.22). △

**Beweis.** Wir benutzen die Transformationsregeln (2.4) und schreiben (11.22) in der Form

$$\begin{aligned}
 \max \quad & c^T x + d^T y \\
 & Ax + By \leq a \\
 & Cx + Dy \leq b \\
 & -Cx - Dy \leq -b \\
 & -Ix + 0y \leq 0
 \end{aligned}$$

Das hierzu duale Programm ist nach Definition

$$\begin{aligned}
 \min \quad & u^T a + v_1^T b - v_2^T b \\
 & u^T A + v_1^T C - v_2^T C - w^T = c^T \\
 & u^T B + v_1^T D - v_2^T D = d^T \\
 & u, \quad v_1, \quad v_2, \quad w \geq 0.
 \end{aligned}$$

Setzen wir  $v := v_1 - v_2$  und lassen wir  $w$  weg, so erhalten wir (11.23). Analog folgt, dass (11.22) dual zu (11.23) ist. □

Von nun an können wir also sagen, dass das duale Programm zum dualen Programm das primale ist, bzw. von einem Paar dualer Programme sprechen. Zur Übersicht und als Nachschlagewerk listet Tabelle 11.1 eine Reihe von Paaren dualer Programme auf. Die Korrektheit der Aussagen ergibt sich direkt aus (11.21). Aus den obigen primal-dualen Relationen kann man die Transformationsregeln in Tabelle 11.2 ableiten. Speziell bedeutet dies z. B. bezüglich des LP  $\max c^T x, Ax \leq b, x \geq 0$ :

primal	dual
Gleichung oder Ungleichung	Variable
Ungleichung	nichtnegative Variable
Gleichung	nicht vorzeichenbeschränkte Variable
nichtnegative Variable	Ungleichung
nicht vorzeichenbeschränkte Variable	Gleichung

Tabelle 11.2: Transformationsregeln für primal-duale LP-Paare.

- Jeder primalen Ungleichung  $A_i \cdot x \leq b_i$  ist eine nichtnegative duale Variable  $y_i$  zugeordnet.
- Jeder primalen nichtnegativen Variablen  $x_j$  ist die duale Ungleichung  $y^T A_{\cdot j} \geq c_j$  zugeordnet.

Aus dem Vorhergehenden dürfte klar sein, dass der Dualitätssatz nicht nur für das spezielle Paar linearer Programme (P), (D) aus Satz (11.18) gilt, sondern für alle Paare dualer linearer Programme. Um später darauf Bezug nehmen zu können, wollen wir den Dualitätssatz in voller Allgemeinheit formulieren.

**(11.24) Satz (Allgemeiner Dualitätssatz).** *Es seien (P) ein lineares Programm und (D) das zu (P) duale Programm.*

- Haben (P) und (D) zulässige Lösungen, so haben (P) und (D) Optimallösungen, und die Zielfunktionswerte aller Optimallösungen stimmen überein.*
- Hat eines der beiden Programme (P) oder (D) eine Optimallösung, so auch das andere.*
- Hat eines der Programme (P) oder (D) keine zulässige Lösung, so ist das andere unbeschränkt oder hat keine zulässige Lösung.*
- Ist eines der Programme (P) oder (D) unbeschränkt, so hat das andere keine zulässige Lösung.* △

Wir wollen nun zwei weitere wichtige Sätze zur Charakterisierung von Optimallösungen linearer Programme formulieren und beweisen.

**(11.25) Satz (Satz vom schwachen komplementären Schlupf).** *Es seien  $A, B, C, D$  und  $a, b, c, d$  dimensionsverträglich und (11.22), (11.23) das zugehörige Paar dualer linearer Programme. Die Vektoren  $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$  bzw.  $\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix}$  seien zulässig für (11.22) bzw. (11.23), dann sind die folgenden Aussagen äquivalent:*

- $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$  ist optimal für (11.22) und  $\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix}$  ist optimal für (11.23).*
- $(c^T - (\bar{u}^T A + \bar{v}^T C))\bar{x} - \bar{u}^T (a - (A\bar{x} + B\bar{y})) = 0.$*

- (c) Für alle Komponenten  $\bar{u}_i$  der Duallösung gilt:  $\bar{u}_i > 0 \implies A_i \bar{x} + B_i \bar{y} = a_i$ .  
 Für alle Komponenten  $\bar{x}_j$  der Primallösung gilt:  $\bar{x}_j > 0 \implies \bar{u}^T A_{.j} + \bar{v}^T C_{.j} = c_j$ .
- (d) Für alle Zeilenindizes  $i$  von  $A$  und  $B$  gilt:  $A_i \bar{x} + B_i \bar{y} < a_i \implies \bar{u}_i = 0$ .  
 Für alle Spaltenindizes  $j$  von  $A$  und  $C$  gilt:  $\bar{u}^T A_{.j} + \bar{v}^T C_{.j} > c_j \implies \bar{x}_j = 0$ .  $\triangle$

**Beweis.** (a)  $\iff$  (b):

$$\begin{aligned} \text{(a)} &\iff c^T \bar{x} + d^T \bar{y} = \bar{u}^T a + \bar{v}^T b \quad (\text{nach (11.24)}) \\ &\iff c^T \bar{x} + (\bar{u}^T B + \bar{v}^T D) \bar{y} - \bar{u}^T a - \bar{v}^T (C \bar{x} + D \bar{y}) = 0 \\ &\iff c^T \bar{x} + \bar{u}^T B \bar{y} + \bar{v}^T D \bar{y} - \bar{u}^T a - \bar{v}^T C \bar{x} - \bar{v}^T D \bar{y} - \bar{u}^T A \bar{x} + \bar{u}^T A \bar{x} = 0 \\ &\iff c^T \bar{x} - (\bar{u}^T A + \bar{v}^T C) \bar{x} - \bar{u}^T a + \bar{u}^T (A \bar{x} + B \bar{y}) = 0 \\ &\iff \text{(b)}. \end{aligned}$$

(b)  $\implies$  (c): Es seien  $t^T := c^T - (\bar{u}^T A + \bar{v}^T C)$  und  $s := a - (A \bar{x} + B \bar{y})$ . Nach Voraussetzung gilt also  $t \leq 0$  und  $s \geq 0$ . Aus  $\bar{x} \geq 0$  und  $\bar{u} \geq 0$  folgt daher  $t^T \bar{x} \leq 0$  und  $\bar{u}^T s \geq 0$ , mithin  $t^T \bar{x} - \bar{u}^T s \leq 0$ . Es kann also  $t^T \bar{x} - \bar{u}^T s = 0$  nur dann gelten, wenn  $t^T \bar{x} = 0$  und  $\bar{u}^T s = 0$ . Daraus ergibt sich (c).

(c)  $\implies$  (b): Trivial.

(c)  $\iff$  (d): Durch Negation.  $\square$

Aus (11.25) folgt für jedes Paar dualer linearer Programme durch Spezialisierung die Gültigkeit eines Satzes vom schwachen komplementären Schlupf.

**(11.26) Satz (Satz vom starken komplementären Schlupf).** *Besitzen beide dualen linearen Programme*

$$\begin{array}{ccc} \max c^T x & & \min y^T b \\ Ax \leq b & \text{(P)} & \text{und} & y^T A = c^T & \text{(D)} \\ & & & y \geq 0 & \end{array}$$

zulässige Lösungen, so existieren optimale Lösungen  $\bar{x}, \bar{u}$ , so dass für alle Zeilenindizes  $i$  von  $A$  gilt:

$$\begin{array}{ccc} \bar{u}_i > 0 & \iff & A_i \bar{x} = b_i \\ \text{bzw.} & & A_i \bar{x} < b_i \iff \bar{u}_i = 0 \end{array} \quad \triangle$$

**Beweis.** Aufgrund von Satz (11.25) genügt es zu zeigen, dass optimale Lösungen  $\bar{x}, \bar{u}$  existieren mit

(i)  $A_i \bar{x} = b_i \implies \bar{u}_i > 0$ ,

$$(ii) \bar{u}_i = 0 \implies A_i \bar{x} < b_i.$$

Setzen wir

$$\bar{A} := \begin{pmatrix} -c^T & b^T \\ A & 0 \\ 0 & A^T \\ 0 & -A^T \\ 0 & -I \end{pmatrix}, \quad \bar{a} := \begin{pmatrix} 0 \\ b \\ c \\ -c \\ 0 \end{pmatrix}, \quad \bar{B} := (A, -I), \quad \bar{b} := b,$$

dann gilt:

$$\begin{aligned} \bar{x}, \bar{u} \text{ optimal} &\iff \bar{A} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} \leq \bar{a}, \\ \bar{x}, \bar{u} \text{ erfüllen (i) und (ii)} &\iff \bar{B} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} < \bar{b}. \end{aligned}$$

Zum Beweis der Behauptung genügt es also zu zeigen, dass

$$\bar{A} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} \leq \bar{a}, \quad \bar{B} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} < \bar{b}$$

konsistent ist. Mit Satz (11.24) hat nach Voraussetzung  $\bar{A} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} \leq \bar{a}$  eine Lösung, also ist nach Folgerung (11.5) die Konsistenz dieses Systems äquivalent zur Inkonsistenz von

$$p \geq 0, \quad 0 \neq q \geq 0, \quad p^T \bar{A} + q^T \bar{B} = 0^T, \quad p^T \bar{a} + q^T \bar{b} \leq 0.$$

Angenommen, dieses System besitzt eine Lösung, sagen wir  $p^T = (\lambda, p_1^T, p_2^T, p_3^T, p_4^T) \geq 0, 0 \neq q \geq 0$ , dann gilt:

$$\begin{aligned} (p_1 + q)^T A &= \lambda c^T \\ (p_2 - p_3)^T A^T - (p_4 + q)^T &= -\lambda b^T \\ (p_1 + q)^T b + (p_2 - p_3)^T c &\leq 0. \end{aligned}$$

Daraus folgt

$$\begin{aligned} 0 &\geq \lambda((p_1 + q)^T b + (p_2 - p_3)^T c) = (p_1 + q)^T \lambda b + (p_2 - p_3)^T \lambda c \\ &= (p_1 + q)^T A(p_3 - p_2) + (p_1 + q)^T (p_4 + q) + (p_2 - p_3)^T A^T (p_1 + q) \\ &= p_1^T p_4 + p_1^T q + q^T p_4 + q^T q \\ &\geq q^T q \\ &> 0. \end{aligned}$$

Daraus folgt  $(p_1 + q)^T b + (p_2 - p_3)^T c > 0$ , ein Widerspruch. □

Es gibt eine natürliche Merkregel für die verschiedenen Sätze vom komplementären Schlupf. Wir wissen bereits, dass zu jeder primalen Variablen eine duale Restriktion (auch komplementäre Restriktion genannt) und zu jeder primalen Restriktion eine duale (komplementäre) Variable gehören. Die Sätze (11.25) und (11.26) zeigen nun, dass zur Charakterisierung von Optimallösungen vorzeichenbeschränkte Variable und Ungleichungen besonders wichtig sind. Zu jeder vorzeichenbeschränkten primalen (dualen) Variablen gehört eine duale (primale) Ungleichung und umgekehrt. Ist  $a^T x \leq \alpha$  eine Ungleichung und  $\bar{x}$  ein Vektor, so nennen wir die Ungleichung *straff* (bezüglich  $\bar{x}$ ), falls  $a^T \bar{x} = \alpha$  gilt, andernfalls nennen wir sie *locker*. Der Satz vom schwachen komplementären Schlupf (11.25) sagt dann aus: Gewisse Vektoren sind optimal für (11.22) bzw. (11.23) genau dann, wenn für jede lockere Nichtnegativitätsbedingung die komplementäre Ungleichung straff ist, d. h. wenn in der komplementären Ungleichung kein Schlupf auftritt. Satz (11.26) kann wie folgt formuliert werden. Es gibt Optimallösungen, bei denen Schlüpfen komplementär auftreten, bzw. bei denen die duale Nichtnegativitätsbedingung genau dann locker ist, wenn die komplementäre primale Ungleichung straff ist.

## Literaturverzeichnis

- K. Leichtweiß. *Konvexe Mengen*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1980.
- O. L. Mangasarian. *Nonlinear Programming*. McGraw-Hill, New York, 1969.
- A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, New York, 1998.
- J. Stoer and C. Witzgall. *Convexity and optimization in finite dimensions*. Springer, Berlin, 1970.

## 12 Varianten der Simplex-Methode

Der Simplex-Algorithmus ist das „Arbeitspferd“ der Optimierung. Effiziente Implementationen von Varianten dieses Verfahrens sind die Basis vieler akademischer und kommerzieller Codes zur Lösung von Optimierungsproblemen. Insbesondere die kombinatorische und gemischt-ganzzahlige Optimierung kommen bei der Lösung von schwierigen Problemen nicht ohne diese Methode aus.

In Kapitel 9 haben wir das Grundprinzip des Simplex-Algorithmus erläutert. Wir setzen hier die Kenntnis dieses Kapitels (Definitionen, Sätze, Notation, etc.) voraus. Im Nachfolgenden wollen wir das Verständnis des Verfahrens vertiefen und insbesondere die Ingredienzien guter Implementierungen erläutern. Die in Kapitel 9 eingeführte Tableau-Methode ist in der Praxis unbrauchbar. Bitte nicht implementieren! Wie es besser gemacht werden kann, wird in diesem Kapitel (zumindest teilweise) skizziert, wobei auf numerische Aspekte speziell in den Übungen eingegangen wird.

### 12.1 Der revidierte Simplexalgorithmus

Betrachtet man die Grundversion des Simplexalgorithmus (9.17), so stellt man fest, dass während einer Iteration i. A. nicht sämtliche Spalten der Matrix  $\bar{A}$  benötigt werden. Der revidierte Simplexalgorithmus nutzt diese Tatsache aus, indem er stets nur solche Spalten von  $\bar{A}$  berechnet, die im jeweiligen Schritt benötigt werden.

Alle numerischen Implementationen des Simplexverfahrens benutzen als Grundgerüst den revidierten Simplexalgorithmus. Wir wollen deshalb im folgenden den revidierten Simplexalgorithmus in Unterprogramme (Subroutines) zerlegen und später zeigen, wie diese Subroutines bei den jeweiligen numerischen Implementationsvarianten realisiert werden. Wir gehen immer davon aus, dass wir ein LP in Standardform

$$\begin{aligned} \max \quad & c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned} \tag{*}$$

vorliegen haben (siehe Definition (9.1)). Alle Bezeichnungen werden, wie im Kapitel 9 festgelegt, verwendet:  $\{1, \dots, m\}$  ist die Zeilenindexmenge,  $\{1, \dots, n\}$  die Spaltenindexmenge von  $A$ ,  $B = (p_1, \dots, p_m)$  und  $N = (q_1, \dots, q_{n-m})$  bezeichnen Spaltenindexvektoren von  $A$  und werden gleichzeitig als Teilmengen von  $\{1, \dots, n\}$  aufgefasst, und statt  $A_B$  bzw.  $A_N$  schreiben wir vereinfacht  $A_B$  bzw.  $A_N$ ; vgl. die Konventionen in (9.3).

Der nachfolgende Algorithmus beschreibt nur die Phase II des Simplexverfahrens. Wir gehen also davon aus, dass zu Beginn eine zulässige Basis  $A_B$  vorliegt.

**(12.1) Algorithmus (Revidierter Simplexalgorithmus).**

**Eingabe:**  $A \in \mathbb{K}^{(m,n)}$ ,  $b \in \mathbb{K}^m$ ,  $c \in \mathbb{K}^n$ , Spaltenindexvektor  $B$ , zugehörige Matrix  $A_B^{-1}$  und Vektor  $\bar{b} = A_B^{-1}b$ .

**Ausgabe:** Eine optimale Lösung  $x$  des linearen Programms (\*).

1. **BTRAN** (Backward Transformation)  
Berechne  $\pi^T := c_B^T A_B^{-1}$ . (Wir werden diese Werte später *Schattenpreise* nennen.)
2. **PRICE** (Pivotspaltenauswahl)  
Berechne die reduzierten Kostenkoeffizienten  $\bar{c}_j := (c_N^T)_j - \pi^T A_N e_j$ , für  $j = 1, \dots, n-m$  und wähle einen Index  $s$  mit  $\bar{c}_s > 0$ .  
**Abbruchkriterium** (vgl. Schritt (II.1) von (9.17)):  
Gilt  $\bar{c}_j \leq 0$  für  $j = 1, \dots, n-m$ , so ist die gegenwärtige Basislösung optimal. Gib den Vektor  $x$  mit  $x_B = \bar{b}$ ,  $x_N = 0$  aus. Der Optimalwert von (\*) ist  $c^T x = c_B^T \bar{b} = c_0$ .  
STOP!
3. **FTRAN** (Forward Transformation)  
Aktualisiere die Pivotspalte  $\bar{d} := A_B^{-1}d = A_B^{-1}A_{\cdot q_s}$ .
4. **CHUZR** (Pivotzeilenauswahl)  
Berechne  $\lambda_0 = \min\{\frac{\bar{b}_i}{\bar{d}_i} \mid \bar{d}_i > 0, i = 1, \dots, m\}$  und wähle einen Index  $r \in \{i \in \{1, \dots, m\} \mid \frac{\bar{b}_i}{\bar{d}_i} = \lambda_0\}$ .
5. **WRETA** (Updating der Basis)  
Entferne das  $r$ -te Element von  $B$  und ersetze es durch  $q_s$ . Der neue Spaltenindexvektor heie  $B'$ .  
Berechne  $A_{B'}^{-1}$ ,  $\bar{b} = A_{B'}^{-1}b$ . △

Wir werden später und in den Übungen auf numerische „Tricks“ zur Implementation der Updates eingehen. Es seien hier jedoch bereits einige Vorteile der revidierten Simplexmethode festgehalten:

- Wesentlich weniger Rechenaufwand, speziell bei Programmen mit erheblich mehr Variablen als Gleichungen.
- Die Ausgangsmatrix  $A$  kann auf externem Speicher bzw. bei dünn besetzten Matrizen durch spezielle Speichertechniken gehalten werden. Spalten von  $A$  werden je nach Bedarf generiert. (Sparse-Matrix-Techniques)
- Die Kontrolle über die Rechengenauigkeit ist besser. Bei der Tableaumethode akkumulieren die Rundefehler. Beim revidierten Simplexalgorithmus kann bei Bedarf mit Hilfe eines Inversionsunterprogramms  $A_B^{-1}$  neu berechnet werden.
- Es gibt besondere Speichertechniken, die eine explizite Speicherung von  $A_B^{-1}$  vermeiden. Man merkt sich lediglich die Etavektoren und berechnet dann über die Formel aus Satz (9.12) aus der ersten Basisinversen durch Linksmultiplikation mit den Elementarmatrizen die gegenwärtige Basisinverse.



## 12.2 Die Behandlung oberer Schranken

In linearen Programmen kommen häufig Beschränkungen der folgenden Form vor:

$$0 \leq x \leq u.$$

Da diese strukturell sehr einfach sind, kann man sie algorithmisch besser behandeln als allgemeine Ungleichungen. Normalerweise führen wir Ungleichungen durch Einfügung von Schlupfvariablen wie folgt in Gleichungen und Nichtnegativitätsbedingungen über:

$$x + \bar{x} = u, \quad x \geq 0, \quad \bar{x} \geq 0.$$

Ist jedoch der Wert von  $x$  (bzw.  $\bar{x}$ ) festgelegt, so ist der Wert der Komplementärvariablen  $\bar{x}$  (bzw.  $x$ ) bereits eindeutig bestimmt. Diesen Vorteil kann man nun so ausnutzen, dass man im Simplexverfahren nur eine der beiden Variablen mitschleppt und die zusätzliche Gleichung völlig weglässt. Für jede Zeile oder Spalte muss jedoch festgehalten werden, ob sie  $x$  oder der Komplementärvariablen  $\bar{x} = u - x$  entspricht. Die Zeilenauswahlregeln sind ebenfalls etwas komplizierter, da eine neue in die „Basis“ hineinzunehmende Variable nur maximal bis zu ihrer Beschränkung erhöht werden darf und die übrigen Basisvariablen ebenfalls ihre Schranken nicht überschreiten dürfen.

Wir wollen im folgenden die sogenannte *Obere-Schranken-Technik* zur Behandlung von linearen Programmen der Form

$$\begin{aligned} \max \quad & c^T x \\ & Ax = b \\ & 0 \leq x \leq u \end{aligned} \tag{12.2}$$

besprechen. Diese „upper-bound-technique“ behandelt nur explizit nach oben beschränkte Variablen. Sind einige der Variablen nicht explizit nach oben beschränkt, so legen wir, um die weitere Diskussion und die Formeln zu vereinfachen, fest, dass ihre obere Schranke  $+\infty$  ist. Das heißt, die Komponenten von  $u$  haben Werte aus der Menge  $\mathbb{R}_+ \cup \{+\infty\}$ .

Weiter benötigen wir eine erweiterte Definition von Basis und Nichtbasis. Wir halten uns im Prinzip an Konvention (9.3), lassen jedoch zu, dass der Vektor der Indizes von Nichtbasisvariablen positive und negative Indizes enthalten kann. Durch das Vorzeichen wollen wir uns merken, ob eine Nichtbasisvariable die obere oder untere Schranke annimmt, und zwar legen wir fest, dass für eine Nichtbasisvariable  $q_s$  der Wert  $x_{q_s}$  Null ist, falls das Vorzeichen von  $q_s$  positiv ist, andernfalls ist der Wert von  $x_{q_s}$  die obere Schranke  $u_{q_s}$ . Um dies formeltechnisch einfach aufschreiben zu können, treffen wir die folgenden Vereinbarungen. Ist  $B = (p_1, \dots, p_m)$  ein Spaltenindexvektor, so dass  $A_B$  eine Basis ist und  $N = (q_1, \dots, q_{n-m})$  wie in (9.3) definiert, dann sei

$$\bar{N} := (\bar{q}_1, \dots, \bar{q}_{n-m}) \quad \text{mit } \bar{q}_i = q_i \text{ oder } \bar{q}_i = -q_i, \tag{12.3}$$

$$\begin{aligned} N^+ &:= (q_i \mid \bar{q}_i > 0), \\ N^- &:= (q_i \mid \bar{q}_i < 0). \end{aligned} \tag{12.4}$$

Die in  $B$ ,  $N^+$  und  $N^-$  enthaltenen Indizes bilden also eine Partition der Spaltenindexmenge  $\{1, \dots, n\}$  von  $A$ . Bei einer Rechnerimplementation genügt es natürlich, sich  $\overline{N}$  zu merken, denn  $N = (|\overline{q}_1|, \dots, |\overline{q}_{n-m}|)$ . Ist  $A_B$  regulär, so nennen wir  $A_B$  eine *E-Basis* (erweiterte Basis) von  $A$  und  $\overline{N}$  eine *E-Nichtbasis* von  $A$ . Der Vektor  $x \in \mathbb{K}^n$  mit

$$\begin{aligned} x_{N^+} &= 0 \\ x_{N^-} &= u_{N^-} \\ x_B &= A_B^{-1}b - A_B^{-1}A_{N^-}u_{N^-} \end{aligned} \tag{12.5}$$

heißt *E-Basislösung* zur E-Basis  $A_B$ . Eine  $\left\{ \begin{array}{l} \text{E-Basis } A_B \\ \text{E-Basislösung } x \end{array} \right\}$  heißt *zulässig*, wenn  $0 \leq x_B \leq u_B$  und heißt *nicht degeneriert* (*nicht entartet*), falls  $0 < x_B < u_B$ , andernfalls *degeneriert* (*entartet*). Gibt es keine oberen Schranken, so gilt offenbar  $N = \overline{N} = N^+$  und alle Formeln reduzieren sich auf den bereits behandelten Fall.

**(12.6) Satz.** Gegeben sei ein Polyeder  $P := \{x \in \mathbb{K}^n \mid Ax = b, 0 \leq x \leq u\}$  mit  $\text{rang}(A) = m$ . Ein Vektor  $x \in \mathbb{K}^n$  ist genau dann eine Ecke von  $P$ , wenn  $x$  zulässige E-Basislösung ist.  $\triangle$

**Beweis.** Es gilt  $P = P(D, d)$  mit

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

Also folgt mit Satz (8.8):  $x$  Ecke von  $P \iff \text{rang}(D_{\text{eq}(\{x\})}) = n$ . Seien  $J = \text{eq}(\{x\})$  und  $J_1, J_2, J_3, J_4$  so gewählt, dass

$$D_J = \begin{pmatrix} A_{J_1} & & & \\ & * & & \\ -A_{J_2} & & & \\ & & I_{J_3} & 0 \\ & & & -I_{J_4} \end{pmatrix}$$

Gilt  $\text{rang}(D_J) = n$ , so besitzt  $A_{J_1 \cup J_2}$  vollen Rang und mit  $K := \{1, \dots, n\} \setminus (J_3 \cup J_4)$  ist  $A_B := A_{J_1 \cup J_2, K}$  regulär. Man verifiziert sofort, dass  $A_B$  zulässig ist, wenn  $N^- = J_3$ ,  $N^+ = J_4$  gewählt wird. Die Rückrichtung ist nun evident.  $\square$

Wir kennzeichnen also im Weiteren Basen, Nichtbasen und Basislösung von (12.2) mit einem E, um sie von den in Kapitel 9 eingeführten Begriffen zu unterscheiden.

**(12.7) Algorithmus (Upper-Bound-Technique zur Lösung von linearen Programmen der Form (12.2)).** Wir nehmen an, dass  $u \in (\mathbb{K}_+ \cup \{+\infty\})^n$  gilt und dass eine zulässige E-Basis  $A_B$  vorliegt. Wir beschreiben lediglich die Phase II.

**Eingabe:** Eine zulässige E-Basis  $A_B, A_B^{-1}, \bar{b} = A_B^{-1}b, c, B = (p_1, \dots, p_m)$  und  $\bar{N} = (\bar{q}_1, \dots, \bar{q}_{n-m})$  (die Mengen  $N, N^+, N^-$  können daraus bestimmt werden),  $\bar{b} = A_B^{-1}b - A_B^{-1}A_{N^-}u_{N^-}$ .

**Ausgabe:** Eine optimale Lösung des linearen Programms (12.2).

1. **BTRAN:**

Berechne  $\pi^T := c_B^T A_B^{-1}$ .

2. **PRICE:**

Berechne  $\bar{c}^T := c_N^T - \pi^T A_N$  und wähle ein  $s \in \{1, \dots, n-m\}$  mit

$$\bar{c}_s \begin{cases} > 0 & \text{falls } \bar{q}_s > 0 \\ < 0 & \text{falls } \bar{q}_s < 0 \end{cases}$$

mittels irgendeiner der Pivotspaltenauswahlregeln. Gibt es keinen solchen Index  $s$ , so ist die aktuelle E-Basislösung optimal.

**Begründung:** Offenbar gilt  $x \in \{y \in \mathbb{K}^n \mid Ay = b, 0 \leq y \leq u\} \iff x_B = A_B^{-1}b - A_B^{-1}A_{N^+}x_{N^+} - A_B^{-1}A_{N^-}x_{N^-}$  und  $0 \leq x_B, x_{N^+}, x_{N^-} \leq u$ . Also ist

$$\begin{aligned} c^T x &= c_B^T x_B + c_{N^+}^T x_{N^+} + c_{N^-}^T x_{N^-} \\ &= c_B^T A_B^{-1}b + (c_{N^+}^T - c_B^T A_B^{-1}A_{N^+})x_{N^+} + (c_{N^-}^T - c_B^T A_B^{-1}A_{N^-})x_{N^-} \\ &= \pi^T b + (c_{N^+}^T - \pi^T A_{N^+})x_{N^+} + (c_{N^-}^T - \pi^T A_{N^-})x_{N^-}. \end{aligned}$$

Da für die gegenwärtige E-Basislösung gilt  $x_{N^-} = u_{N^-}, x_{N^+} = 0$ , kann der Zielfunktionswert nur verbessert werden, wenn für eine Nichtbasisvariable  $q_s$  entweder gilt  $\bar{q}_s > 0$  und  $\bar{c}_s = (c_N^T - \pi^T A_N)_s > 0$  oder  $\bar{q}_s < 0$  und  $\bar{c}_s < 0$ .

3. **FTRAN:**

Berechne  $\bar{d} := A_B^{-1}A_{q_s} = \bar{A}_{\cdot s}$ .

4. **CHUZR:**

Setze  $\sigma = \text{sign}(\bar{q}_s)$  und berechne

$$\lambda_0 := \min \left\{ \frac{\bar{b}_i}{\sigma \bar{a}_{is}} \mid \sigma \bar{a}_{is} > 0, i = 1, \dots, m \right\},$$

$$\lambda_1 := \min \left\{ \frac{\bar{b}_i - u_{p_i}}{\sigma \bar{a}_{is}} \mid \sigma \bar{a}_{is} < 0, i = 1, \dots, m \right\},$$

$$\lambda_2 := u_{q_s}.$$

(a) Ist keine der Zahlen  $\lambda_0, \lambda_1, \lambda_2$  endlich, so ist das Programm (12.2) unbeschränkt (das kann natürlich nur vorkommen, wenn  $x_{q_s}$  nicht explizit beschränkt ist).

(b) Ist  $\lambda_0 = \min\{\lambda_0, \lambda_1, \lambda_2\}$ , so wähle einen Index

$$r \in \left\{ i \in \{1, \dots, m\} \mid \frac{\bar{b}_i}{\sigma \bar{a}_{is}} = \lambda_0, \sigma \bar{a}_{is} > 0 \right\}$$

mittels irgendeiner Pivotzeilenauswahlregel.

## 12 Varianten der Simplex-Methode

(c) Ist  $\lambda_1 = \min\{\lambda_0, \lambda_1, \lambda_2\}$ , so wähle

$$r \in \{i \in \{1, \dots, m\} \mid \frac{\bar{b}_i - u_{p_i}}{\sigma \bar{a}_{is}} = \lambda_1, \sigma \bar{a}_{is} < 0\}$$

mittels irgendeiner Pivotzeilenauswahlregel.

(d) Ist  $\lambda_2 = \min\{\lambda_0, \lambda_1, \lambda_2\}$ , so setze  $\bar{q}_s := -\bar{q}_s$ , berechne  $\bar{b}$  neu und gehe zurück zur PRICE-Routine 2.

### 5. WRETA:

Setze  $B' = (p_1, \dots, p_{r-1}, q_s, p_{r+1}, \dots, p_m)$ ,  $\bar{N} = (\bar{q}_1, \dots, \bar{q}_{s-1}, \bar{q}, \bar{q}_{s+1}, \dots, \bar{q}_{n-m})$ , wobei  $\bar{q} = p_r$ , falls  $\lambda_0 = \min\{\lambda_0, \lambda_1, \lambda_2\}$ , bzw.  $\bar{q} = -p_r$ , falls  $\lambda_1 = \min\{\lambda_0, \lambda_1, \lambda_2\}$ .

Berechne  $A_{B'}^{-1}$ ,  $\bar{b} = A_{B'}^{-1}b - A_{B'}^{-1}A_{N'}u_{N'}$ .  $\triangle$

**Begründung:** Die Pivotzeilenauswahl (CHUZR-Routine) dient dazu, die Zeilenauswahl so zu bestimmen, dass die transformierten Variablen nach erfolgter Pivotoperation wieder zulässig sind. Entsprechend der Transformationsregeln beim Pivotisieren mit dem Pivotelement  $\bar{a}_{rs}$  (vergleiche Satz (9.12)) muss gewährleistet sein, dass nach Ausführung des Pivotschrittes für die neue E-Basislösung folgendes gilt:

$$(i) \quad 0 \leq x'_{p_i} = \bar{b}_i - \frac{\bar{a}_{is}}{\bar{a}_{rs}} \bar{b}_r \leq u_{p_i}, \text{ für } i \neq r,$$

$$(ii) \quad 0 \leq x'_{q_s} = \frac{1}{\bar{a}_{rs}} \bar{b}_r \leq u_{q_s},$$

$$(iii) \quad x'_{p_r} \in \{0, u_{p_r}\},$$

$$(iv) \quad x'_{q_i} \in \{0, u_{q_i}\}, \text{ für } i \neq s.$$

Wollen wir nun den Wert der Variablen  $x_{q_s}$  um  $\lambda \geq 0$  ändern (d. h. erhöhen, falls  $\bar{q}_s > 0$ , bzw. um  $\lambda$  erniedrigen, falls  $\bar{q}_s < 0$ ), so können wir das so lange tun bis entweder

- eine Basisvariable den Wert ihrer oberen oder unteren Schranke annimmt oder
- die Nichtbasisvariable  $q_s$  den Wert ihrer anderen Schranke annimmt.

Aus letzterem folgt natürlich, dass  $\lambda \leq u_{q_s}$  gelten muss, ansonsten würden wir bei Erhöhung ( $\bar{q}_s > 0$ ) die obere Schranke bzw. bei Erniedrigung ( $\bar{q}_s < 0$ ) die untere Schranke für  $x_{q_s}$  verletzen. Dies erklärt die Bestimmung von  $\lambda_2$  in 4.

Ist  $\bar{q}_s > 0$ , so bewirkt eine Erhöhung von  $x_{q_s}$  um den Wert  $\lambda$ , dass  $x'_{p_i} = \bar{b}_i - \lambda \bar{a}_{is}$  gilt. Aus der Bedingung (i) erhalten wir daher die folgenden Schranken für  $\lambda$ :

$$\lambda \leq \frac{\bar{b}_i}{\bar{a}_{is}} \quad \text{für } \bar{a}_{is} > 0,$$

$$\lambda \leq \frac{\bar{b}_i - u_{p_i}}{\bar{a}_{is}} \quad \text{für } \bar{a}_{is} < 0.$$

Ist  $\bar{q}_s < 0$ , so bewirkt die Verminderung von  $x_{q_s}$  ( $= u_{q_s}$ ) um den Wert  $\lambda \geq 0$ , dass  $x'_{p_i} = \bar{b}_i + \lambda \bar{a}_{is}$  gilt. Aus (i) erhalten wir somit:

$$\lambda \leq -\frac{\bar{b}_i}{\bar{a}_{is}} \quad \text{für } \bar{a}_{is} < 0,$$

$$\lambda \leq \frac{u_{p_i} - \bar{b}_i}{\bar{a}_{is}} \quad \text{für } \bar{a}_{is} > 0.$$

Dies begründet die Definition von  $\lambda_0$  und  $\lambda_1$ . Sei nun  $\lambda_{\min} = \min\{\lambda_0, \lambda_1, \lambda_2\}$ . Gilt  $\lambda_{\min} = \lambda_2$ , so wird einfach der Wert einer Nichtbasisvariablen von der gegenwärtigen Schranke auf die entgegengesetzte Schranke umdefiniert. Die Basis ändert sich dadurch nicht, jedoch konnte aufgrund der Auswahl in PRICE eine Zielfunktionsverbesserung erreicht werden. Gilt  $\lambda_{\min} = \lambda_1$  oder  $\lambda_{\min} = \lambda_2$ , so führen wir einen üblichen Pivotschritt durch. Bei  $\lambda_{\min} = \lambda_1$  wird die neue Nichtbasisvariable  $x_{p_r}$  mit Null festgesetzt, da die untere Schranke von  $x_{p_r}$  die stärkste Einschränkung für die Veränderung von  $x_{q_s}$  darstellte. Bei  $\lambda_{\min} = \lambda_2$ , wird analog  $x_{p_r}$  eine Nichtbasisvariable, die den Wert ihrer oberen Schranke annimmt.

Wir wollen nun anhand eines Beispiels die Ausführung von Algorithmus (12.7) in Tableautechnik demonstrieren. Dabei werden wir keine neuen Tableau-Update-Formeln definieren, sondern mit den bekannten Formeln für die Pivotschritte rechnen. Wir führen lediglich eine zusätzliche Spalte ein, in der jeweils aus  $\bar{b}$  und  $c_0$  der Wert der gegenwärtigen E-Basislösung berechnet wird. Ist also  $c_0$  der gegenwärtige Wert der Basislösung und  $x_B = A_B^{-1}b = \bar{b}$  so hat die zusätzliche Spalte in der ersten Komponente den Eintrag  $\bar{c}_0 := -c_0 - \bar{c}_{N-} u_{N-}$ , und die restlichen Komponenten berechnen sich durch

$$\bar{b} := \bar{b} - \bar{A}_{N-} u_{N-}.$$

**(12.8) Beispiel.** Wir betrachten das folgende LP, dessen Lösungsmenge in Abbildung 12.1 dargestellt ist:

$$\begin{aligned} \max \quad & 2x_1 + x_2 \\ & -x_1 + x_2 \leq \frac{1}{2} \\ & x_1 + x_2 \leq 2 \\ & x_1 \leq \frac{3}{2} \quad (= u_1) \\ & x_2 \leq 1 \quad (= u_2) \\ & x_1, x_2 \geq 0. \end{aligned}$$

Wir führen 2 Schlupfvariablen  $x_3, x_4$  ein (mit oberen Schranken  $+\infty$ ) und erhalten als Anfangstableau  $T_0$  (mit zusätzlicher Spalte  $\begin{pmatrix} \bar{c}_0 \\ \bar{b} \end{pmatrix}$ ), die mit der Spalte  $\begin{pmatrix} -c_0 \\ \bar{b} \end{pmatrix}$  identisch ist,

12 Varianten der Simplex-Methode

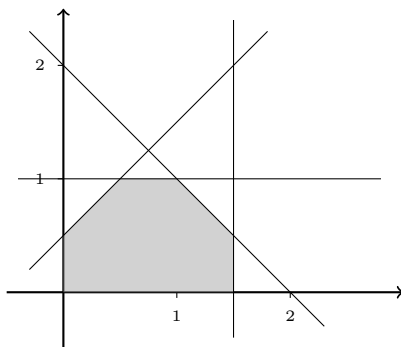


Abbildung 12.1: Lösungsmenge zum LP aus Beispiel (12.8)

da keine Nichtbasisvariable von Null verschieden ist):

$$T_0 : \begin{array}{cccc|cc} & 1 & 2 & 3 & 4 & & \\ & 2 & 1 & 0 & 0 & 0 & 0 \\ \hline 3 & -1 & 1 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 4 & 1 & 1 & 0 & 1 & 2 & 2 \end{array} = \begin{array}{c|c|c|c} \bar{c} & 0 & -c_0 & \bar{c}_0 \\ \hline A & I & \bar{b} & \bar{b} \end{array} \quad \begin{array}{l} B = (3, 4) \\ \bar{N} = (1, 2) \end{array}$$

Wir wählen die erste Spalte als Pivotspalte, d. h.  $\bar{q}_1 = 1$ . Schritt 4. CHUZR ergibt

$$\lambda_0 = \frac{\bar{b}_2}{\bar{a}_{21}} = \frac{2}{1} = 2, \quad \lambda_1 \text{ ist wegen } u_3 = \infty \text{ nicht definiert,} \quad \lambda_2 = u_1 = \frac{3}{2},$$

d. h.  $\lambda_{\min} = \lambda_2$ . Wir führen also Schritt 4.(d) aus, setzen  $\bar{q}_1 := -\bar{q}_1$ , d. h.  $\bar{q}_1 = -1$ , und berechnen  $\bar{b}$  neu. Die neue letzte Spalte, d. h. die neue E-Basislösung zur Basis  $A_B$  erhalten wir aus der (normalen) Basislösung wegen  $N^- = (1)$  wie folgt:

$$\bar{b} = \bar{b} - \bar{A}_{\cdot 1} u_1 = \begin{pmatrix} \frac{1}{2} \\ 2 \end{pmatrix} - \frac{3}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix},$$

und der Zielfunktionswert erhöht sich um  $\bar{c}_1 \cdot u_1 = 2 \cdot \frac{3}{2} = 3$ . Mithin erhalten wir als nächstes Tableau (mit  $\bar{q}_1 := -\bar{q}_1$ ):

$$T_1 : \begin{array}{cccc|cc} & -1 & 2 & 3 & 4 & & \\ & 2 & 1 & 0 & 0 & 0 & -3 \\ \hline 3 & -1 & 1 & 1 & 0 & \frac{1}{2} & 2 \\ 4 & 1 & 1 & 0 & 1 & 2 & \frac{1}{2} \end{array} \quad \begin{array}{l} B = (3, 4) \\ \bar{N} = (-1, 2) \end{array}$$

Als Pivotspalte kommt nur die 2. Spalte ( $\bar{q}_2 = 2$ ) in Frage. CHUZR ergibt

$$\lambda_0 = \frac{\bar{b}_2}{\bar{a}_{22}} = \frac{1}{2}, \quad \lambda_1 \text{ nicht definiert,} \quad \lambda_2 = u_2 = 1.$$

Die Wahl der Austauschzeile in 4.(b) ergibt ein eindeutig bestimmtes  $r = 2$ . Wir führen dann 5. in Tableautechnik aus. Wir führen also einen Standardpivotschritt auf dem Element  $\bar{a}_{22} = 1$  durch und korrigieren anschließend  $\bar{b}$  und  $c_0$ , um  $\bar{\bar{b}}$  und  $\bar{\bar{c}}_0$  zu erhalten.

$$T_2 : \begin{array}{cccc|c|c} & -1 & 2 & 3 & 4 & & \\ & 1 & 0 & 0 & -1 & -2 & -\frac{7}{2} \\ 3 & -2 & 0 & 0 & -1 & -\frac{3}{2} & \frac{3}{2} \\ 2 & 1 & 1 & 0 & 1 & 2 & \frac{1}{2} \end{array} \quad \begin{array}{l} B = (3, 2) \\ \bar{N} = (-1, 4) \end{array}$$

Dieses Tableau ist optimal, mit  $x_1 = \frac{3}{2}$ ,  $x_2 = \frac{1}{2}$ .

Zur Erläuterung der übrigen Fälle beginnen wir noch einmal mit Tableau  $T_0$ , wählen aber die 2. Spalte als Pivotspalte. CHUZR ergibt

$$\lambda_0 = \frac{\bar{\bar{b}}_1}{\bar{a}_{12}} = \frac{1}{2}, \quad \lambda_1 \text{ nicht definiert}, \quad \lambda_2 = 1.$$

In diesem Fall machen wir einen Standardpivotschritt mit dem Pivotelement  $\bar{a}_{rs} = \bar{a}_{12} = 1$ .

$$T_3 : \begin{array}{cccc|c|c} & 1 & 2 & 3 & 4 & & \\ & 3 & 0 & -1 & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 2 & -1 & 1 & 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 4 & 2 & 0 & -1 & 1 & \frac{3}{2} & \frac{3}{2} \end{array} \quad \begin{array}{l} B = (2, 4) \\ \bar{N} = (1, 3) \end{array}$$

Die gegenwärtige Basislösung wird wie üblich in eine E-Basislösung transformiert. Als Pivotspalte kommt nur die erste in Frage. CHUZR ergibt

$$\lambda_0 = \frac{\bar{\bar{b}}_2}{\bar{a}_{21}} = \frac{3}{4}, \quad \lambda_1 = \frac{\bar{\bar{b}}_1 - u_2}{\bar{a}_{11}} = \frac{1}{2}, \quad \lambda_2 = u_1 = \frac{3}{2} \quad (\lambda_{\min} = \lambda_1)$$

Wir führen nun Schritt 5 des Algorithmus (12.7) durch einen Pivotschritt auf  $\bar{a}_{rs} = \bar{a}_{11} = -1$  aus und berechnen das gesamte Tableau neu. Anschließend müssen wir die rechte Spalte des Tableaus korrigieren. Da  $\lambda_{\min} = \lambda_1$ , wird die neue Nichtbasisvariable  $x_2$  mit ihrer oberen Schranke  $u_2 = 1$  belegt (und  $\bar{q}_1 = -2$  gesetzt). Wir müssen daher vom Neuberechneten  $\bar{b} = (-\frac{1}{2}, \frac{5}{2})^T$  noch das  $u_2$ -fache der zum Index 2 gehörigen Spalte von  $\bar{A}$  (also  $\bar{A}_{\cdot 1} = (-1, 2)^T$ ) subtrahieren. Vom Zielfunktionswert  $c_0$  subtrahieren wir  $u_2 \cdot \bar{c}_2 = 1 \cdot 3$ . Daraus ergibt sich

$$T_4 : \begin{array}{cccc|c|c} & 1 & -2 & 3 & 4 & & \\ & 0 & 3 & 2 & 0 & 1 & -2 \\ 1 & 1 & -1 & -1 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 4 & 0 & 2 & 1 & 1 & \frac{5}{2} & \frac{1}{2} \end{array} \quad \begin{array}{l} B = (1, 4) \\ \bar{N} = (-2, 3) \end{array}$$

$$\bar{\bar{b}} = \begin{pmatrix} -\frac{1}{2} \\ \frac{5}{2} \end{pmatrix} - 1 \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

## 12 Varianten der Simplex-Methode

Als Pivotspalte müssen wir nun die 3. Spalte von  $T_4$ , also  $\bar{A}_{.2}$ , wählen.

$$\lambda_0 = \frac{\bar{b}_2}{\bar{a}_{22}} = \frac{1}{2} \quad \lambda_1 = \frac{\bar{b}_1 - u_1}{\bar{a}_{12}} = 1, \quad \lambda_2 = u_1 = \frac{3}{2}.$$

Somit führen wir einen Standardpivotschritt auf  $\bar{a}_{22} = 1$  aus

$$T_5 : \begin{array}{cccc|c|c} & 1 & -2 & 3 & 4 & & \\ & 0 & -1 & 0 & -2 & -4 & -3 \\ \hline 1 & 1 & 1 & 0 & 1 & 2 & 1 \\ 3 & 0 & 2 & 1 & 1 & \frac{5}{2} & \frac{1}{2} \end{array} \quad \begin{array}{l} B = (1, 3) \\ \bar{N} = (-2, 4) \end{array}$$

Als Pivotspalte kommt nur  $\bar{A}_{.1}$ , also die 2. Spalte von  $T_5$ , in Frage, d. h.  $q_s = q_1 = 2$ ,  $\bar{q}_1 = -2$ . Wir erhalten durch CHUZR

$$\lambda_0 \text{ undefiniert, } \lambda_1 = \frac{\bar{b}_1 - u_1}{\sigma \bar{a}_{11}} = \frac{1}{2}, \quad \lambda_2 = u_2 = 1.$$

Unser Pivotelement ist somit  $\bar{a}_{11} = 1$ . Wie üblich führen wir einen Standardpivotschritt durch. Da wir die neue Nichtbasisvariable  $x_1$  mit ihrer oberen Schranke  $u_1 = \frac{3}{2}$  belegen, ziehen wir wieder vom Neuberechneten  $\bar{b}$  das  $u_1$ -fache der zu 1 gehörigen Spalte von  $\bar{A}$  (das ist  $\bar{A}_{.1} = (1, -2)^T$ ) ab, um  $\bar{b}$  zu erhalten.

$$T_6 : \begin{array}{cccc|c|c} & -1 & 2 & 3 & 4 & & \\ & 1 & 0 & 0 & -1 & -2 & -\frac{7}{2} \\ \hline 2 & 1 & 1 & 0 & 1 & 2 & \frac{1}{2} \\ 3 & -2 & 0 & 1 & -1 & -\frac{3}{2} & \frac{3}{2} \end{array} \quad \begin{array}{l} B = (2, 3) \\ \bar{N} = (-1, 4) \end{array}$$

Wir haben wieder die Optimallösung erreicht. △

Analog zu oberen Schranken können natürlich auch untere Schranken behandelt werden. Ferner kann man auch allgemeine obere Schranken (generalized upper bounds, GUB), dies sind Restriktionen der Form  $\sum x_i \leq a$ , auf ähnliche Weise berücksichtigen. Bei solchen Restriktionen können ebenfalls Pivotschemata entworfen werden, die wesentlich weniger Speicher- und Rechenaufwand erfordern als die Standardmethode. Weiterhin gibt es effiziente Techniken zur Behandlung von sogenannten variable upper bounds (VUB) der Form  $0 \leq x \leq y$ .

## 12.3 Das duale Simplexverfahren

Wir wollen nun eine Variante des Simplexverfahrens darstellen, die – wie wir noch sehen werden – bei bestimmten Problemstellungen von Vorteil ist. Sei  $A_B$  eine Basis von  $A$ , und betrachte das Paar dualer linearer Programme:



$$\begin{aligned} \max \quad & c^T x \\ & Ax = b \quad (\text{P}) \quad \text{und} \quad \min \quad u^T b \\ & x \geq 0 \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad u^T A \geq c^T \quad (\text{D}) \end{aligned}$$

bzw.

$$\begin{aligned} \max \quad & c^T x \\ & x_B = A_B^{-1}b - A_B^{-1}A_N x_N \quad \text{und} \quad \min \quad u^T b \\ & x_N, x_B \geq 0 \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad u^T A_B \geq c_B^T \\ & \quad u^T A_N \geq c_N^T \end{aligned}$$

**(12.9) Definition.** Die Basis  $A_B$  von  $A$  heißt primal zulässig, falls  $A_B^{-1}b \geq 0$  und dual zulässig, falls  $\bar{c}^T = c_N^T - c_B^T A_B^{-1} A_N \leq 0$ . Die zugehörigen Basislösungen  $x$  bzw.  $u$  mit  $x_B = A_B^{-1}b$ ,  $x_N = 0$  bzw.  $u^T = c_B^T A_B^{-1}$  heißen dann primal bzw. dual zulässig.  $\triangle$

**(12.10) Satz.** Sei  $P = \{u \in \mathbb{K}^m \mid u^T A \geq c^T\}$ . Der Vektor  $\bar{u}$  ist genau dann eine Ecke von  $P$ , wenn  $\bar{u}$  eine dual zulässige Basislösung ist.  $\triangle$

**Beweis.** Sei  $\bar{u}$  Ecke von  $P = P(-A^T, -c)$  und  $I = \text{eq}(\{\bar{u}\})$ . Mit Satz (8.8) folgt  $\text{rang}((-A^T)_I) = m$ , d. h. es existiert  $B \subseteq I$  mit  $A_B$  Basis von  $A$ , und es gilt  $\bar{u}^T A_B = c_B^T$ ,  $\bar{u}^T A_N \geq c_N^T$ . Also ist  $\bar{u}^T = c_B^T A_B^{-1}$  und  $c_B^T A_B^{-1} A_N \geq c_N^T$ , d. h.  $A_B$  ist dual zulässig. Ist umgekehrt  $A_B$  dual zulässig, so ist  $\bar{u}^T := c_B^T A_B^{-1}$  aufgrund von Satz (8.8) eine Ecke von  $P$ .  $\square$

**(12.11) Bemerkung.** Ist  $A_B$  eine Basis von  $A$ , und sind  $x$  bzw.  $u$  die zu  $A_B$  gehörenden primalen bzw. dualen (aber nicht notwendigerweise zulässigen) Basislösungen, so gilt:  $c^T x = u^T b$ .  $\triangle$

**Beweis.**  $u^T b = c_B^T A_B^{-1} b = c_B^T x_B = c^T x$ .  $\square$

**(12.12) Korollar.** Ist  $A_B$  eine Basis von  $A$ , so ist  $A_B$  optimal genau dann, wenn  $A_B$  primal und dual zulässig ist.  $\triangle$

**Beweis.** Dualitätssatz.  $\square$

**(12.13) Korollar.** Ist  $A_B$  eine optimale Basis für das Programm (P), dann ist  $c_B^T A_B^{-1}$  eine optimale Lösung des zu (P) dualen Programms (D).  $\triangle$

Der Vektor  $\pi := c_B^T A_B^{-1}$  (mit  $A_B$  dual zulässig) heißt der Vektor der Schattenpreise (vgl. ökonomische Interpretation der Dualität und Schritt 1 in (12.1)).

Wir wollen nun die dem dualen Simplexverfahren zugrunde liegende Idee entwickeln und bemerken, dass – im Prinzip – das duale Simplexverfahren das primale Simplexverfahren angewendet auf das duale Problem ist. Sei  $A_B$  eine Basis von  $A$ , so lässt sich (P) umformen in:

$$\begin{aligned} \max \quad & c_B^T A_B^{-1} b + \bar{c}^T x_N \\ & \bar{A} x_N + I x_B = A_B^{-1} b (= \bar{b}) \\ & x_B, x_N \geq 0 \end{aligned} \tag{12.14}$$

oder

$$\begin{aligned} c_B^T A_B^{-1} b + \max \quad & \bar{c}^T x_N \\ & \bar{A} x_N \leq \bar{b} \\ & x_N \geq 0 \end{aligned} \quad (12.15)$$

Das zu (12.15) duale Programm lautet (bei Weglassung des konstanten Terms  $c_B^T A_B^{-1} b$  der Zielfunktion):

$$\begin{aligned} \min \quad & u^T \bar{b} \\ & \bar{A}^T u \geq \bar{c} \\ & u \geq 0 \end{aligned} \quad (12.16)$$

Die Einführung von Schlupfvariablen  $y_N$  (und Variablenumbenennung  $y_B := u$ ) ergibt:

$$\begin{aligned} - \max \quad & -\bar{b}^T y_B \\ & -\bar{A}^T y_B + I y_N = -\bar{c} \\ & y \geq 0 \end{aligned} \quad (12.17)$$

Durch eine lange Kette von Umformungen ist es uns also gelungen, ein LP in Standardform (P) in ein anderes LP in Standardform (12.17) zu transformieren. Was haben wir gewonnen? Nicht viel, es sei denn, die Basis  $A_B$  ist *dual zulässig*. Denn in diesem Falle gilt, dass die rechte Seite von (12.17), also  $-\bar{c}$ , nichtnegativ ist. Also ist die Matrix  $I$  eine zulässige (Ausgangs-)basis für (12.17), und wir können auf (12.17) den Simplexalgorithmus (direkt mit Phase II beginnend) anwenden.

Eine besonders interessante Anwendung ergibt sich, wenn das ursprüngliche Problem die Form

$$\begin{aligned} \max \quad & c^T x \\ & Ax \leq b \\ & x \geq 0 \end{aligned}$$

hat, wobei  $c \leq 0$  und  $b \not\geq 0$  ist. Hier ist eine dual zulässige Basis direkt gegeben, während eine primal zulässige Basis erst mittels Phase I gefunden werden müsste.

### (12.18) Algorithmus (Duale Simplexmethode).

**Eingabe:**  $A' \in \mathbb{K}^{(m,n)}$ ,  $b' \in \mathbb{K}^m$ ,  $c \in \mathbb{K}^n$ .

**Ausgabe:** Optimale Lösung  $x$  des LP  $\max\{c^T x \mid A'x = b', x \geq 0\}$ .

**Phase I:** Bestimmung eines Subsystems  $Ax = b$ ,  $x \geq 0$  mit  $P=(A, b) = P=(A', b')$ , welches (9.2) erfüllt (falls möglich) und Bestimmung einer dual zulässigen Basis  $A_B$  von  $A$ . Berechne:  $\bar{A} = A_B^{-1} A_N$ ,  $\bar{b} = A_B^{-1} b$ ,  $\bar{c}^T = c_N^T - c_B^T A_B^{-1} A_N$ . (Dieser Teil des Algorithmus wird analog zur Phase I (9.38) der Grundversion des Simplexalgorithmus ausgeführt.)

**Phase II:** Optimierung

- (II.1)** (Optimalitätsprüfung)  
Gilt  $\bar{b}_i \geq 0$  ( $i = 1, \dots, m$ ), so ist die gegenwärtige Basislösung optimal ( $A_B$  ist primal und dual zulässig). Setze  $x_B = \bar{b}$  und  $x_N = 0$ , andernfalls gehe zu (II.2).
- (II.2)** (Bestimmung der Pivotzeile)  
Wähle einen Index  $r$  mit  $\bar{b}_r < 0$ .
- (II.3)** (Prüfung auf Beschränktheit des Optimums)  
Gilt  $\bar{A}_r \geq 0$ , so hat das duale Programm keine endliche Lösung, also ist  $P^=(A, b) = \emptyset$ . STOP!
- (II.4)** Berechne  $\lambda_0 := \min \left\{ \frac{\bar{c}_j}{\bar{a}_{rj}} \mid \bar{a}_{rj} < 0, j = 1, \dots, n - m \right\}$ .
- (II.5)** (Bestimmung der Pivotspalte)  
Wähle Index  $s \in \left\{ j \in \{1, \dots, n - m\} \mid \frac{\bar{c}_j}{\bar{a}_{rj}} = \lambda_0 \right\}$ .
- (II.6)** (Pivotoperation)  
Setze  $A_{B'}^{-1} := EA_B^{-1}$  mit  $E$  aus Satz (9.12), und berechne alle notwendigen Parameter neu. Gehe zu (II.1).  $\triangle$

Der duale Simplexalgorithmus wurde 1954 von Lemke entwickelt. Wie wir gerade gezeigt haben, ist er (nach Transformation) jedoch nichts anderes als die Anwendung des primalen Simplexalgorithmus auf das duale Problem. Aus diesem Grunde hat sich lange Zeit niemand die Mühe gemacht, eine „echte“ Implementation des dualen Verfahrens vorzunehmen. Als in den 90er Jahren die Anwendung der ganzzahligen Optimierung immer wichtiger und Schnittebenenverfahren (die wir später erklären werden) als die zentralen Methoden zur Lösung ganzzahliger Optimierungsprobleme erkannt wurden, sind erstmals duale Methoden programmiert worden. Zur Überraschung vieler erwiesen sich diese dualen Codes als (in der Praxis) schneller als die primalen, so dass sie heute bei praktischen Rechnungen dominieren. Ein Grund dafür ist, dass Goldfarbs „steepest edge“ Regel beim dualen Simplexalgorithmus sehr gut funktioniert, siehe Goldfarb and Reid (1977). In Bixby (2002) finden sich ein experimenteller Nachweis und heuristische Begründungen für die bessere Performanz des dualen Simplexverfahrens.

**(12.19) Bemerkung (Tableauform des dualen Simplexalgorithmus).** Wollen wir das duale Simplexverfahren auf das Problem (12.14) (mit dual zulässiger Basis  $A_B$ ) anwenden, so können wir das verkürzte Tableau

$$VT := \begin{array}{|c|c|} \hline -\bar{b}^T & -z_0 \\ \hline -\bar{A}^T & -\bar{c} \\ \hline \end{array}$$

verwenden und auf dieses Tableau die Update Formeln des verkürzten Simplextableaus (diese sind in Schritt (II.6) von (9.17) angegeben) anwenden. Jedoch zeigt eine einfache Überlegung, dass wir bei der Auswahl der Indizes  $r$  und  $s$  entsprechend (II.2) bzw. (II.5)

die Updateformeln (II.6) aus (9.17) auch direkt auf die Matrix  $\bar{A}$  bzw.  $\bar{b}$  und  $\bar{c}$  anwenden können und zwar genau in der Form wie sie in (9.17)(II.6) aufgeschrieben wurden (um das neue Tableau  $VT'$  zu erhalten).

$$VT' := \begin{array}{|c|c|} \hline -(\bar{b}')^T & -z'_0 \\ \hline -(\bar{A}')^T & -\bar{c}' \\ \hline \end{array}$$

Wir führen also die Transponierung bzw. den Vorzeichenwechsel nicht explizit durch, sondern beachten den Vorzeichenwechsel bzw. die Transponierung implizit bei der Bestimmung der Pivotzeile bzw. Pivotspalte.

Dann könnten wir aber auch wieder das primale (verkürzte) Tableau verwenden und anhand dieses Tableaus sowohl den primalen als auch den dualen Simplexalgorithmus durchführen.  $\triangle$

**(12.20) Bemerkung (Das Tucker-Tableau).** Für die nachfolgende spezielle Form linearer Programme erhalten wir primale und duale Optimallösungen direkt aus der Tableaurechnung. Wir betrachten:

$$\begin{array}{ll} \max & c^T x \\ \text{(P)} & Ax \leq b \\ & x \geq 0 \end{array} \quad \text{und} \quad \begin{array}{ll} \min & u^T b \\ \text{(D)} & u^T A \geq c^T \\ & u \geq 0 \end{array}$$

oder mit Schlupfvariablen

$$\begin{array}{ll} \max & z_0 \\ & c^T x - z_0 = 0 \\ & Ax - b = -u \\ & x, u \geq 0 \end{array} \quad \text{und} \quad \begin{array}{ll} \max & z_0 \\ & u^T b + z_0 = 0 \\ & -c^T + u^T A = x^T \\ & u, x \geq 0 \end{array}$$

oder in Tableauschreibweise

$c_1, c_2, \dots, c_n$	$z_0$
$A$	$b_1$
	$\vdots$
	$b_m$

$N$  = Nichtbasisvariable des primalen Programms

= Basisvariable des dualen Programms

$B$  = Basisvariable des primalen Programms

= Nichtbasisvariable des dualen Programms

Ein solches Tucker-Tableau heißt *primal zulässig*, wenn  $b \geq 0$ , und *dual zulässig*, wenn  $c \leq 0$ .

Führen wir den primalen oder dualen Simplexalgorithmus bis zum Optimaltableau durch, so sind die jeweiligen Optimallösungen gegeben durch:

$$\left. \begin{aligned} x_{B(i)} &= \bar{b}_i & i &= 1, \dots, m \\ x_{N(i)} &= 0 & i &= 1, \dots, n \end{aligned} \right\} \text{optimale Lösung des primalen Programms,}$$

$$\left. \begin{aligned} u_{N(i)} &= -\bar{c}_i & i &= 1, \dots, n \\ u_{B(i)} &= 0 & i &= 1, \dots, m \end{aligned} \right\} \text{optimale Lösung des dualen Programms.} \quad \triangle$$

Hat man eine Optimallösung von einem LP in Standardform ausgehend berechnet, so lässt sich die Optimallösung des dualen Programms nicht ohne Weiteres aus dem Tableau ablesen.

**(12.21) Beispiel.** Wir betrachten die folgenden zueinander dualen Programme (P) und (D). Die Lösungsmenge von (P) sei mit  $P$  bezeichnet, die von (D) mit  $D$ .  $P$  und  $D$  sind in Abbildung 12.2 dargestellt.

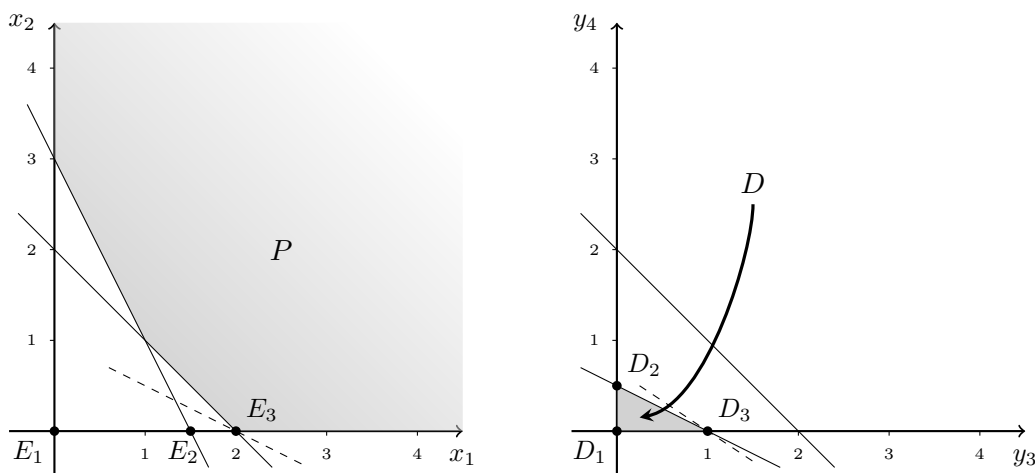


Abbildung 12.2: Primale und duale Lösungsmengen aus Beispiel (12.21)

$$\begin{aligned} \max \quad & -x_1 - 2x_2 \\ & -x_1 - x_2 \leq -2 \quad (x_3) \\ & -2x_1 - x_2 \leq -3 \quad (x_4) \\ & x_1, x_2 \geq 0 \end{aligned} \quad (\text{P})$$

$$\begin{aligned} \min \quad & -2y_3 - 3y_4 \\ & -y_3 - 2y_4 \geq -1 \quad (-y_1) \\ & -y_3 - y_4 \geq -2 \quad (-y_2) \\ & y_3, y_4 \geq 0 \end{aligned} \quad (\text{D})$$

Wir schreiben das Tucker-Tableau zu diesem Problem auf.

1	2		
-1	-2		0
3	-1	-1	-2
4	-2	-1	-3

$N = (1, 2)$   
 $B = (3, 4)$

## 12 Varianten der Simplex-Methode

$$\begin{array}{ll}
 x_1 = 0 & y_3 = 0 \\
 x_2 = 0 & y_4 = 0 \\
 x_3 = -2 & y_1 = 1 \\
 x_4 = -3 & y_2 = 2
 \end{array}$$

Die gegenwärtige Basis ist dual aber nicht primal zulässig. Wir machen einen Update mit dem Pivotelement  $\bar{a}_{rs} = \bar{q}_{21} = -2$ . Das neue Tableau hat folgende Form.

$$\begin{array}{ccc|c}
 & 4 & 2 & \\
 \hline
 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \\
 3 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
 1 & -\frac{1}{2} & \frac{1}{2} & \frac{3}{2} \\
 \hline
 & & & \\
 \end{array}
 \quad
 \begin{array}{l}
 N = (4, 2) \\
 B = (3, 1)
 \end{array}$$

$$\left. \begin{array}{l} x_1 = \frac{3}{2} \\ x_2 = 0 \end{array} \right\} = E_2
 \quad
 \left. \begin{array}{l} y_3 = 0 \\ y_4 = \frac{1}{2} \end{array} \right\} = D_2$$

Es folgt ein Pivotschritt mit  $\bar{a}_{rs} = \bar{a}_{11} = -\frac{1}{2}$ .

$$\begin{array}{ccc|c}
 & 3 & 2 & \\
 \hline
 & -1 & -1 & -2 \\
 4 & -2 & 1 & 1 \\
 1 & -1 & 1 & 2 \\
 \hline
 & & & \\
 \end{array}
 \quad
 \begin{array}{l}
 N = (3, 2) \\
 B = (4, 1)
 \end{array}$$

$$\left. \begin{array}{l} x_1 = 2 \\ x_2 = 0 \end{array} \right\} = E_3
 \quad
 \left. \begin{array}{l} y_3 = 1 \\ y_4 = 0 \end{array} \right\} = D_3$$

Dieses Tableau ist optimal. △

## 12.4 Postoptimierung und parametrische Programme

Wir haben bisher die Theorie (und Praxis) des Simplexverfahrens entwickelt unter der Annahme, dass die Daten,  $A$ ,  $b$ , und  $c$  fest vorgegeben sind. Bei praktischen Problemen können jedoch häufig nicht alle der benötigten Zahlen mit Sicherheit angegeben werden, d. h. es kann Unsicherheit über die Beschränkungen  $b$  oder den zu erwartenden Gewinn  $c^T x$  geben. Man muss also die Tatsache mit in Betracht ziehen, dass die Wirklichkeit nicht exakt in das lineare Modell abgebildet wurde bzw. werden konnte. Darüber hinaus können im Nachhinein zusätzliche Variablen oder Restriktionen auftreten, die bei Aufstellung des Modells übersehen wurden oder nicht berücksichtigt werden konnten.

Wir werden uns nun überlegen, wie wir gewisse auftretende Änderungen behandeln können, wenn wir z. B. die optimale Lösung des ursprünglichen Programms gefunden haben. Wir wollen uns auf die folgenden Fälle beschränken:

1. Variation der rechten Seite  $b$ .
2. Variation der Zielfunktion  $c$ .
3. Änderung einer Nichtbasisspalte.
4. Hinzufügung einer neuen Variablen.
5. Hinzufügung einer neuen Restriktion.

Im Prinzip könnten wir natürlich versuchen, eine Theorie zu entwerfen, bei der Schwankungen aller Daten berücksichtigt werden. Das ist jedoch außerordentlich kompliziert, und es gibt darüber kaum rechnerisch verwertbare Aussagen.

Wir gehen im weiteren davon aus, dass ein LP in Standardform (siehe (9.1))

$$\begin{aligned} \max \quad & c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned} \tag{12.22}$$

gegeben ist und dass wir eine optimale Basis  $A_B$  von  $A$  kennen.

### Änderung der rechten Seite $b$

Wir wollen uns zunächst überlegen, welchen Einfluss eine Änderung der rechten Seite auf die Optimallösung bzw. den Wert der Optimallösung hat.

**(12.23) Bemerkung.** Gegeben seien ein LP in Standardform (12.22) und eine optimale Basis  $A_B$  von  $A$ . Die neue rechte Seite des LP sei  $b' := b + \Delta$ . Wir berechnen die neue Basislösung zur Basis  $A_B$ . Diese ergibt sich aus:

$$x'_B = A_B^{-1}(b + \Delta) = x_B + A_B^{-1}\Delta, \quad x'_N = 0.$$

Gilt  $x'_B \geq 0$ , so ist die neue Basislösung optimal, da sich ja an den reduzierten Kosten  $\bar{c}^T = c_N^T - c_B^T A_B^{-1} A_N$  nichts geändert hat, d. h. es gilt weiterhin  $\bar{c} \leq 0$ .

Gilt  $x'_B \not\geq 0$ , so ist die neue Basislösung primal nicht zulässig. Die Optimalitätsbedingung  $\bar{c} \leq 0$  ist jedoch weiterhin erfüllt. Das aber heißt nach (12.9) dass die Basis dual zulässig ist. Folglich haben wir mit  $A_B$  eine zulässige Basis für die duale Simplexmethode (12.18), und wir können direkt mit Phase II von (12.18) mit der Neuberechnung der Optimallösung beginnen.  $\triangle$

**Exkurs.** In diesem Zusammenhang sei auf eine wichtige Funktion, die die Änderung des Zielfunktionswertes eines LP bei Variationen der rechten Seite angibt, hingewiesen. Diese Funktion hat interessante Eigenschaften, von denen wir nachfolgend einige aufführen wollen. Die Funktion

$$L : \{b \in \mathbb{K}^m \mid P^=(A, b) \neq \emptyset\} \longrightarrow \mathbb{K} \cup \{\infty\}$$

definiert durch

$$L(b) := \sup\{c^T x \mid Ax = b, x \geq 0\} \quad (12.24)$$

heißt *Perturbationsfunktion* bzgl. des LPs

$$\begin{aligned} \max \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Es bezeichne

$$\text{RP}(A) := \{b \in \mathbb{K}^m \mid P^=(A, b) \neq \emptyset\}$$

den Definitionsbereich von  $L$  (wegen  $0 \in \text{RP}(A)$  ist  $\text{RP}(A) \neq \emptyset$ ) und

$$\text{epi}(L) := \left\{ \begin{pmatrix} v \\ z \end{pmatrix} \in \mathbb{K}^{m+1} \mid v \in \text{RP}(A), z \in \mathbb{K}, L(v) \geq z \right\}$$

den Epigraphen von  $L$ .

**(12.25) Satz.** *Die Menge  $\text{epi}(L)$  ist ein polyedrischer Kegel.* △

**Beweis.** Offenbar ist  $\text{epi}(L)$  eine Projektion des polyedrischen Kegels  $\{(x^T, v^T, z)^T \mid Ax - v = 0, x \geq 0, c^T x - z \geq 0\}$  auf die  $(v, z)$ -Koordinaten. Also ist nach Satz (10.13)  $\text{epi}(L)$  ein Polyeder, das trivialerweise ein Kegel ist. □

**(12.26) Bemerkung.**

$$\exists b \in \text{RP}(A) \text{ mit } L(b) < \infty \iff \forall b \in \text{RP}(A) \text{ gilt } L(b) < \infty. \quad \triangle$$

**Beweis.** Sei  $V := \{x \in \mathbb{K}^n \mid Ax = 0\}$ , dann gibt es zu jedem  $b \in \mathbb{K}^m$  einen Vektor  $x(b) \in \mathbb{K}^n$  mit  $P^=(A, b) = (V + x(b)) \cap \mathbb{K}_+^n$ . Folglich gilt für alle  $b \in \text{RP}(A)$ :  $L(b) = \infty \iff L(0) = \infty$ . □

**(12.27) Satz.** *Ist  $L \neq \infty$ , so gibt es Vektoren  $g^i \in \mathbb{K}^m$ ,  $i = 1, \dots, k$ , so dass*

$$L(v) = \min\{v^T g^i \mid i = 1, \dots, k\}$$

für alle  $v \in \text{RP}(A)$  gilt. △

**Beweis.** Mit Satz (12.25) ist  $K := \text{epi}(L)$  ein polyedrischer Kegel. Also gibt es nach Bemerkung (2.8) eine  $(r, m+1)$ -Matrix  $B$ , so dass  $K = P(B, 0)$  gilt. Da  $L(0) \geq 0$ , ist  $\begin{pmatrix} 0 \\ z \end{pmatrix} \notin K$  für alle  $z > 0$ . Also kann die letzte Spalte von  $B$  kein Nullvektor sein, und wir können o. B. d. A. annehmen, dass  $B$  die Form

$$B = \begin{pmatrix} H & 0 \\ -G & h \end{pmatrix}$$



mit  $h \neq 0$  besitzt. Da  $L \neq \infty$ , folgt mit (12.26)  $L(0) < \infty$ , und daraus folgt direkt  $L(0) = 0$ . Also wissen wir, dass  $\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in K$ . Dies impliziert  $h > 0$ . Sei o. B. d. A.  $h = 1$ . Dann gilt:  $\text{epi}(L) = \{(v^T, z)^T \mid Hv \leq 0, Gv \geq z\}$ , und es ist  $L(v) = z \iff G_i v = z$  für ein  $i$ . Bezeichnen  $g^i, i = 1, \dots, k$  die Zeilen von  $G$ , so gilt

$$L(v) = \min\{v^T g^i \mid i = 1, \dots, k\}. \quad \square$$

**(12.28) Korollar.** Die Perturbationsfunktion  $L$  ist stückweise linear und konkav.  $\triangle$

Speziell folgt daraus, dass sich der Wert der Zielfunktion eines linearen Programms stetig mit einer Variation der rechten Seite des LP ändert. Die stetige Änderung lässt sich durch (12.27) explizit angeben. Sie ist „fast überall“ linear, bis auf einige „Knicke“.

### Änderungen der Zielfunktion $c$

Wir gehen wieder davon aus, dass wir eine Optimalbasis  $A_B$  von (12.22) kennen und dass sich die Zielfunktion  $c$  um  $\Delta$  ändert. Da wir eine Basis  $A_B$  gegeben haben, können wir die neuen Kosten der Basislösung ausrechnen. Es gilt

$$\begin{aligned} (c^T + \Delta^T)x &= (c_B^T + \Delta_B^T)A_B^{-1}b + (c_N^T + \Delta_N^T - (c_B^T + \Delta_B^T)A_B^{-1}A_N)x_N \\ &= c_B^T \bar{b} + \bar{c}x_N + \Delta_B^T \bar{b} + \underbrace{(\Delta_N^T - \Delta_B^T \bar{A})}_{=: \bar{\Delta}} x_N. \end{aligned}$$

- (a) Wird keiner der Koeffizienten von  $c_B^T$  geändert (ist also  $\Delta_B = 0$ ), ändert sich wegen  $x_N = 0$  der Zielfunktionswert der Basislösung zur Basis  $A_B$  nicht.
- (b) Sind die neuen reduzierten Kosten  $\bar{c} + \bar{\Delta}$  nicht positiv, so ist das Optimalitätskriterium (9.15)(b) weiterhin erfüllt, d. h.  $A_B$  bleibt die optimale Basis, jedoch ändert sich u. U. der Wert der zugehörigen Basislösung wenn  $\Delta_B \neq 0$ .
- (c) Ist einer der Koeffizienten  $\bar{c}_j + \bar{\Delta}_j$  positiv, so wenden wir Phase II des primalen Simplexalgorithmus mit (zulässiger) Anfangsbasislösung  $A_B$  auf das neue Problem an.

(Zur Berechnung von  $\bar{\Delta}$  genügt die Kenntnis von  $\bar{A}$  oder  $A_B^{-1}$  und  $A_N$ .)

An den obigen Berechnungen kann man sehen, wie man eine Schar von linearen Programmen, bei denen entweder nur  $b$  oder nur  $c$  einer Änderung unterworfen wird, lösen kann.

### Änderungen des Spaltenvektors $A_{\cdot,j}, j = N(s)$

Wollen wir die  $s$ -te Nichtbasisspalte (d. h. die  $j$ -te Spalte von  $A$ ) um  $\Delta$  ändern und kennen wir die Basisinverse  $A_B^{-1}$ , so können wir die neue  $s$ -te Spalte von  $\bar{A}$  berechnen durch

$$\bar{A}'_{\cdot,s} = A_B^{-1}(A_{\cdot,j} + \Delta) = \bar{A}_{\cdot,s} + A_B^{-1}\Delta.$$

Die Basislösung  $\bar{x}$  zu  $A_B$  bleibt weiterhin primal zulässig, da diese Änderung keinen Einfluss auf  $\bar{x}_B = \bar{b}$ ,  $\bar{x}_N = 0$  hat. Jedoch bleibt  $\bar{x}$  i. A. nicht optimal, da sich der  $s$ -te Koeffizient der reduzierten Kosten ( $j = N(s)$ ) wie folgt ändert

$$\begin{aligned}\bar{c}'_s &= c_j - c_B^T(\bar{A}_{.s} + A_B^{-1}\Delta) \\ &= c_j - c_B^T\bar{A}_{.s} - c_B^T A_B^{-1}\Delta \\ &= \bar{c}_s - c_B^T A_B^{-1}\Delta = \bar{c}_s - u^T \Delta,\end{aligned}$$

wobei  $u$  eine optimale duale Lösung ist.

Die Optimalität der gegenwärtigen Basislösung  $\bar{x}$  bleibt genau dann erhalten, wenn  $\bar{c}_s \leq u^T \Delta$  gilt. Andernfalls können wir die neue Optimallösung mit Phase II des primalen Simplexalgorithmus berechnen, wobei wir von der zulässigen Basis  $A_B$  ausgehen (die revidierte Methode ist natürlich von Vorteil).

(Bei Änderung einer Basisspalte kann man kaum Vorhersagen machen. Es muss neu invertiert werden. Dabei kann sowohl primale als auch duale Zulässigkeit verloren gehen, und es kann passieren, dass  $A_B$  nach Änderung einer Basisspalte singulär wird und damit keine Basis mehr ist.)

**(12.29) Beispiel.** Wir betrachten nochmals unser Beispiel (12.21)

$$\begin{aligned}\max \quad & -x_1 - 2x_2 \\ & -x_1 - x_2 \leq -2 \\ & -2x_1 - x_2 \leq -3 \\ & x_1, x_2 \geq 0\end{aligned}$$

Das optimale (Tucker) Tableau ist:

-1	-1	-2
-2	1	1
-1	1	2

duale Lösung  $y_3 = 1$ ,  $y_4 = 0$ , primale Lösung  $x_1 = 2$ ,  $x_2 = 0$ .

Die Spalte  $A_{.2}$  ist eine Nichtbasisspalte. Wir untersuchen, um wieviel diese Spalte geändert werden kann, ohne Optimalität zu verlieren.

$$A_{.2} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \bar{c}_2 = -1, \quad u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

also  $\bar{c}'_2 = -1 - (1, 0) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = -1 - \alpha$ ,

$$\bar{c}'_2 \leq 0 \iff -1 - \alpha \leq 0 \iff \alpha \geq -1.$$

Die obige Basislösung ist also für alle linearen Programme der Form

$$\begin{aligned}\max \quad & -x_1 - 2x_2 \\ & -x_1 + (\alpha - 1)x_2 \leq -2 \\ & -2x_1 + (\beta - 1)x_2 \leq -3 \\ & x_1, x_2 \geq 0\end{aligned} \quad \triangle$$

optimal, wenn nur  $\alpha \geq -1$  gilt. Zur geometrischen Interpretation siehe Abbildung 12.3.

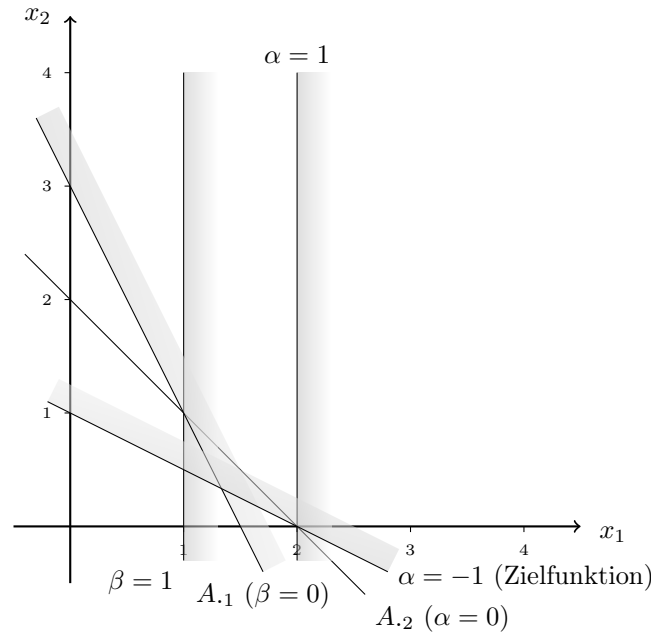


Abbildung 12.3: Änderung der Nichtbasisspalte  $A_2$

### Hinzufügung einer neuen Variablen

Sei  $A_{.n+1}$  die neue Spalte von  $A$  und  $c_{n+1}$  der Zielfunktionswert der neuen Variablen  $x_{n+1}$ . Wir verlängern den Vektor  $N$  der Nichtbasisindizes um eine Komponente und setzen  $N(n - m + 1) := n + 1$ . Wir berechnen

$$\bar{c}_{n-m+1} := c_{n+1} - \underbrace{c_B^T A_B^{-1}}_{u^T} A_{.n+1}.$$

Gilt  $\bar{c}_{n-m+1} \leq 0$ , so bleibt aufgrund von (9.15)(b) die alte Lösung optimal. Andernfalls führen wir mit der Anfangsbasis  $A_B$  den primalen Simplexalgorithmus aus. Dabei ist

$$\bar{A}_{.n-m+1} = A_B^{-1} A_{.n+1}.$$

Im ersten Schritt wird die neue Variable in die Basis aufgenommen.

### Hinzufügung einer neuen Restriktion

Zu den bestehenden Nebenbedingungen fügen wir die neue Restriktion

$$A_{m+1}.x = \sum_{i=1}^n a_{m+1,i} x_i = b_{m+1}$$

hinzu.

**1. Fall:** Die Basislösung  $\bar{x}$  zu  $A_B$  erfüllt  $A_{m+1}\bar{x} = b_{m+1}$ .

In diesem Fall ist  $\bar{x}$  auch die optimale Lösung des erweiterten Problems. Ist die neue Zeile linear abhängig von den übrigen Zeilen von  $A$ , so ist die neu hinzugefügte Gleichung irrelevant für das Problem und kann gestrichen werden. Ist die neue Zeile linear unabhängig von den übrigen, so ist  $\bar{x}$  eine entartete Basislösung. Eine der Nichtbasisvariablen wird mit Wert 0 in die Basis aufgenommen und die Basis entsprechend erweitert.

**2. Fall:** Die Basislösung  $\bar{x}$  erfüllt die neue Gleichung nicht.

Wir führen eine neue Schlupfvariable  $x_{n+1} \geq 0$  (mit dem Zielfunktionswert  $c_{n+1} = 0$ ) ein und verändern die neue Gleichung in

$$\sum_{i=1}^n a_{m+1,i}x_i \pm x_{n+1} = b_{m+1},$$

wobei wir ein  $+$  Zeichen schreiben, falls  $A_{m+1}\bar{x} > b_{m+1}$  gilt, andernfalls ziehen wir  $x_{n+1}$  ab. Wir verlängern  $B$  um eine Komponente und setzen  $B(m+1) := n+1$ , d. h. wir nehmen  $x_{n+1}$  mit dem Wert  $\bar{b}_{m+1} = -|b_{m+1} - A_{m+1}\bar{x}|$  in die Basis auf. Da  $\bar{b}_{m+1}$  negativ ist, ist die neue Basis primal nicht zulässig; wir haben jedoch keine Änderung an der Zielfunktion vorgenommen, d. h. die reduzierten Kosten sind weiterhin nicht positiv. Also ist die gegenwärtige neue Basis dual zulässig. Die neue Basis hat die Form

$$\begin{pmatrix} A_B & 0 \\ A_{m+1,B} & 1 \end{pmatrix}.$$

Die Basisinverse lässt sich wie folgt schreiben:

$$\begin{pmatrix} A_B^{-1} & 0 \\ -A_{m+1,B}A_B^{-1} & 1 \end{pmatrix}.$$

Die neue  $(m+1)$ -te Zeile von  $\bar{A}$  lautet:

$$-A_{m+1,B}\bar{A} + A_{m+1,N} = \bar{A}_{m+1}.$$

Wir führen nun einen dualen Pivotschritt auf der  $(m+1)$ -ten Zeile von  $\bar{A}$  durch, wobei wir versuchen, die Variable  $x_{n+1}$  aus der Basis zu entfernen.

Ergibt der duale Beschränktheitstest ( $\bar{A}_{m+1} \geq 0$ , (12.18)(II.3)), dass das duale Problem unbeschränkt ist, so ist das neue primale Problem unzulässig, d. h. die Hyperebene  $A_{m+1}x = b_{m+1}$  hat einen leeren Schnitt mit der Menge der zulässigen Lösungen des alten Problems, und wir können den Algorithmus beenden.

Andernfalls führen wir Phase II des dualen Simplexalgorithmus (12.18) aus. Endet der Algorithmus mit einer primal und dual zulässigen Lösung, so gibt es eine optimale primale Lösung  $x^*$  mit  $x_{n+1}^* = 0$ . Diese kann wie folgt konstruiert werden, falls  $x_{n+1}^*$  nicht bereits Null ist:

Sei  $x'$  die primale zulässige Basislösung nach Beendigung des dualen Verfahrens und  $\bar{x}$

die Anfangsbasislösung bei Beginn des dualen Programms, d. h.  $\bar{x}_{n+1} = \bar{b}_{m+1} < 0$ . Setzen wir

$$\lambda := \frac{x'_{n+1}}{x'_{n+1} - \bar{x}_{n+1}},$$

so gilt  $\lambda > 0$ , da  $x'_{n+1} > 0$  und  $\bar{x}_{n+1} < 0$ . Sei  $x^* := \lambda \bar{x} + (1 - \lambda)x'$ , dann gilt

$$\begin{aligned} x_i^* &\geq 0 \quad \text{für } i = 1, \dots, n, \text{ da } \bar{x}_i \geq 0, x'_i \geq 0, \\ x_{n+1}^* &= \frac{x'_{n+1}}{x'_{n+1} - \bar{x}_{n+1}} \bar{x}_{n+1} + x'_{n+1} - \frac{x'_{n+1}}{x'_{n+1} - \bar{x}_{n+1}} x'_{n+1} \\ &= \frac{1}{x'_{n+1} - \bar{x}_{n+1}} (x'_{n+1} \bar{x}_{n+1} - x'_{n+1} x'_{n+1} + x'_{n+1} x'_{n+1} - \bar{x}_{n+1} x'_{n+1}) \\ &= 0. \end{aligned}$$

Also gilt  $x^* \geq 0$  und ferner

$$c^T x^* = \lambda \underbrace{c^T \bar{x}}_{\geq c^T x'} + (1 - \lambda) c^T x' \geq c^T x'.$$

Daraus folgen die Zulässigkeit und die Optimalität von  $x^*$ .

## 12.5 Zur Numerik des Simplexverfahrens

Dieser Abschnitt ist nur äußerst cursorisch und oberflächlich ausgearbeitet. Eine sorgfältige Behandlung des Themas würde eine gesamte Spezialvorlesung erfordern. Wir empfehlen dem Leser die Bemerkungen zur Numerik in Chvátal (1983) oder das Buch von Bastian (1980) zu lesen, das einem Spezialaspekt dieses Themas gewidmet ist und diesen einigermaßen erschöpfend behandelt. Das Buch von Murtagh (1981) ist ganz Rechen- und Implementierungstechniken der linearen Optimierung gewidmet. Generelle Methoden zur Behandlung spezieller (z. B. dünn besetzter oder triangulierter oder symmetrischer) Matrizen im Rahmen numerischer Verfahren (etwa Gauß-Elimination, Matrixinvertierung) finden sich in Pissanetsky (1984).

Generell wird bei Implementationen die revidierte Simplexmethode verwendet. Es gibt zwei grundsätzliche Varianten für die Reinverson: **Produktform der Inversen (PFI)**, **Eliminationsform der Inversen (EFI)**.

### Produktform der Inversen

$B^{-1}$  liegt in Produktform vor:  $B^{-1} = E_k \cdot E_{k-1} \cdot \dots \cdot E_1$  mit  $E_i$  aus Satz (9.12).

Prinzip: Gauß-Jordan Verfahren.

Freiheitsgrade:

- a) Positionen der Pivotelemente,
- b) Reihenfolge der Pivots.

- a) hat Einfluss auf die numerische Stabilität  
 b) hat Einfluss auf die Anzahl NNE (nicht Null Elemente) im Etafile (Speicherung der Spalten  $\eta$  von  $E_i$ ).

(12.30) Beispiel.

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

besitzt folgende Produktform-Darstellungen der Inversen:

- a) Pivotalisieren auf der Hauptdiagonalen von „links oben“ nach „rechts unten“ ergibt

$$B^{-1} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{3} & & \\ & 1 & -\frac{1}{3} & \\ & & \frac{2}{3} & \\ & & -\frac{1}{3} & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & & \\ & \frac{1}{2} & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

Zu speichern sind 16 Werte (+Positionen).

- b) Pivotalisieren auf der Hauptdiagonalen von „rechts unten“ nach „links oben“ ergibt:

$$B^{-1} = \begin{pmatrix} \frac{1}{4} & & & \\ \frac{1}{4} & 1 & & \\ \frac{1}{4} & & 1 & \\ \frac{1}{4} & & & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & & \\ & 1 & & \\ & & 0 & 1 \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & & \\ & 1 & 0 & \\ & & 1 & \\ & & & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & & -1 & \\ & 1 & & 0 \\ & & 1 & 0 \\ & & & 1 & 1 \end{pmatrix}$$

Zu speichern sind 10 Werte (+Positionen)

△

Folgende Gegebenheiten sollten bei der Pivotauswahl berücksichtigt werden:

(12.31) **Bemerkung.** Es ist günstig, in einer Spalte mit wenigen NNE zu pivotisieren, da dann die Etavektoren dünn besetzt sind. △

(12.32) **Bemerkung.** Es ist günstig, in einer Zeile mit wenigen NNE zu pivotisieren, da dann in anderen Zeilen der umgeformten Restmatrix potentiell weniger neue NNE entstehen. △

(12.33) **Bemerkung.** Es ist aus Gründen der Stabilität nicht günstig, ein Pivotelement zu wählen, dessen Betrag sehr klein ist. △

In den Inversionsroutinen, die bei den ersten Implementationen des Simplexverfahrens benutzt wurden, beachtete man nur die numerische Stabilität, nahm sich die Spalten der Basismatrix in der Reihenfolge, in der sie abgespeichert waren, multiplizierte sie mit den bereits bestimmten Elementarmatrizen und pivotisierte auf der Komponente mit dem größten Absolutbetrag. Später nahm man eine Vorsortierung der Basisspalten in der Weise vor, so dass die dünn-besetzten zuerst bearbeitet wurden: dies war ohne großen Aufwand möglich, da aufgrund der spaltenweisen Speicherung der NNE die „column counts“ (Anzahl NNE einer bestimmten Spalte) bekannt waren.

Heutzutage wird die Inversion üblicherweise in 2 Phasen zerlegt:



Diese Methode wird z. B. verwendet in der Preassigned Pivot Procedure (Hellerman und Rarick) und ist implementiert in OPTIMA (CDC) und MPS/III (IBM).

### Eliminationsform der Inversen (EFI)

Die Grundlage dieser Methode bildet die folgende Beobachtung:

**(12.37) Satz (LU-Zerlegung).** *Ist  $\bar{B}$  eine reguläre  $(m, m)$ -Matrix, so gibt es eine Matrix  $B$ , die durch Spaltenvertauschungen aus  $\bar{B}$  hervorgeht und sich in der Form  $B = LU$  darstellen lässt, wobei  $L$  eine untere und  $U$  eine obere Dreiecksmatrix ist. Es gilt  $B^{-1} = U^{-1}L^{-1}$ .  $\triangle$*

**Prinzipielles Vorgehen:** Gauß'sches Eliminationsverfahren.

Die Festlegung der Pivotpositionen und ihrer Reihenfolge erfolgt wie bei der PFI in einer vorgeschalteten Booleschen Phase. Da  $U^{-1} = U_2 \cdot U_3 \cdot \dots \cdot U_m$  mit

$$U_i = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & -u_{1i} & & & \\ & & & \ddots & & \\ & & & & -u_{i-1,i} & \\ & & & & & 1 & \\ & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix}$$

gilt, ist der Rechenaufwand bei der Inversion geringer als bei der PFI, was sofort aus der Tatsache folgt, dass die  $U_i$  bereits durch die Spalten von  $U$  gegeben sind und die  $L_i$  sich durch weniger Rechenoperationen ergeben als die entsprechenden Elementarmatrizen bei der PFI.

Die nachfolgende Beschreibung der Inversionsroutine des LP-Codes MPSX/370 der Firma IBM ist dem oben genannten Buch von Bastian (1980) (Seite 31ff) entnommen.

**(12.38) Bemerkung (Die Inversionsroutine von MPSX/370).** Die Routine INVERT von MPSX/370 beruht auf der im vorhergehenden besprochenen LU-Zerlegung der Basismatrix. Sie ist zur Inversion großer dünn-besetzter Matrizen entwickelt worden; es wird daher versucht, durch eine sehr weitgehende Listenverarbeitung den Aufwand proportional zur Anzahl der von Null verschiedenen Elemente (NNE) in der Darstellung der Inversen zu halten.

#### (a) Die Boolesche Phase

In der Booleschen Phase dieser Inversionsroutine werden die Pivotpositionen so vorbestimmt, dass die Anzahl der NNE in der Darstellung der Inversen möglichst gering wird (vgl. Bénichou et al. (1977) u. a.).

Zunächst wird die Besetzungsstruktur der Ausgangsmatrix in die Form zweier Listen übertragen: die eine Liste enthält spaltenweise die Zeilenindices der NNE, und in der anderen werden entsprechend zeilenweise die Spaltenindices gespeichert. Darüber hinaus werden noch column und row counts (Anzahl der NNE in einer Spalte bzw. Zeile) festgehalten. Auf diese Weise wird ein rascher Zugriff auf Spalten oder Zeilen



nach dem Kriterium der Besetzungsdichte ermöglicht.

Die Bestimmung der Pivotpositionen erfolgt in drei Schritten, wobei die beiden ersten Schritte der Identifizierung des „Bumps“ in den oben besprochenen Triangularisierungsverfahren entsprechen:

1. Wähle Spalten mit row count 1; dadurch sind die zugehörigen Pivotzeilen ebenfalls eindeutig festgelegt.

Die Spalten- und Zeilenindices der NNE dieser Spalten werden aus den beiden Listen entfernt und die row und column counts angepasst. Dadurch entstehen möglicherweise wieder Zeilen mit row count 1, die anschließend ausgewählt werden.

2. Entsprechend wie im Schritt 1 werden dann Spalten mit column count 1 ausgewählt, Pivotspalten und -zeilen aus den Indices-Listen gestrichen und row sowie column counts angepasst.

Das Ergebnis der beiden ersten Schritte ist die Umordnung von Spalten der Ausgangsmatrix so, dass die vorbestimmten Pivotelemente auf der Hauptdiagonale liegen. Durch symmetrisches Vertauschen der Zeilen und Spalten gemäß der vorgesehenen Pivotfolge erhält man eine Matrix  $B$  der folgenden Gestalt:

$$B = \begin{array}{|c|c|c|} \hline & & \\ \hline & u^1 & u^3 \\ \hline L^j & & N \\ \hline \end{array}$$

Der Nukleus  $N$  entspricht dem Bump beim Verfahren von Hellerman und Rarick und in der Tat könnte dies Verfahren auch zur Zerlegung des Nukleus benutzt werden.

3. Stattdessen beruht der Schritt 3 der Reversionsroutine von MPSX/370 auf einer LU-Zerlegung des Nukleus in der Booleschen Matrix, wobei die folgenden einfachen Auswahlregeln angewandt werden:

- wähle unter den Spalten mit minimalem column count eine solche als Pivotspalte, die ein NNE in einer Zeile mit möglichst kleinem row count besitzt;
- wähle als Pivotzeile eine Zeile mit minimalem row count unter den Zeilen mit NNE in der Pivotspalte.

Ist eine Pivotposition bestimmt worden, so werden die beiden Indices-Listen aktualisiert; die Anpassung der row und column counts erfolgt entsprechend.

Durch besondere Maßnahmen wird versucht, das Durchsuchen von Listen sowie die Suche nach NNE in Vektoren einzuschränken (eine detaillierte Beschreibung der eingesetzten Methoden findet man bei Gustavson (1972)). So wird die Pivot-

wahl z. B. dadurch vereinfacht, dass zu jeder Spalte  $j$ , in der noch nicht pivotisiert wurde, nicht nur der column count sondern auch der minimale row count unter den Zeilen mit NNE in Spalte  $j$  mitgeführt wird. Spalten mit gleichem column count und minimalem row count werden verkettet. Zusätzlich werden die Listen der Zeilen- und Spaltenindices von NNE zu der vorbestimmten Pivotposition für die numerische Phase aufgehoben:

- die Liste der Spaltenindices in der Pivotzeile gibt die Spalten an, die aktualisiert werden müssen (durch Umspeicherung erhält man zu jeder Spalte die Adressen derjenigen Etavektoren, die zur Aktualisierung dieser Spalte herangezogen werden müssen);
- die Liste der Zeilenindices in der Pivotspalte entspricht der Besetzung mit NNE in der aktualisierten Pivotspalte, und ihre Speicherung verhindert, dass die gesamte Spalte nach NNE durchsucht werden muss.

Werden die Listen im Verlaufe zu umfangreich und damit auch die Aktualisierung zu aufwendig, so wird die Boolesche Matrix explizit als Matrix umgespeichert (ein Byte pro Element). Ist diese Matrix voll-besetzt oder erreicht sie eine vom Benutzer zu spezifizierende Dichte, so wird die Boolesche Phase abgebrochen.

#### (b) Die Numerische Phase

In der Numerischen Phase werden die Pivotspalten in der Reihenfolge, die in der Booleschen Phase vorbestimmt wurde, aktualisiert und die Etavektoren gebildet. Die Etavektoren von  $L^{-1}$  und  $U^{-1}$  werden dabei auf getrennten Files, dem  $L$ -File und dem  $U$ -File, abgespeichert. Die Etavektoren zu  $L^1$  und zu  $U^1$  ergeben sich direkt aus den Ausgangsdaten und werden sofort extern gespeichert. Bei der anschließenden Zerlegung der Nucleusspalten hält man die neu entstehenden Etaspalten des  $L$ -Files zunächst im Hauptspeicher, sofern genug Speicherplatz vorhanden ist.

Sei  $d$  diejenige Basisspalte, aus der die Elementarmatrizen  $L_k$  und  $U_k$  berechnet werden sollen. Ausgangspunkt ist die Spalte

$$\hat{d} := L_{k-1} \cdot \dots \cdot L_2 \cdot L_1 \cdot d,$$

die man sich in einem Arbeitsbereich erzeugt (man beachte, dass die Etavektoren von  $L_1, \dots, L_j$  nicht benötigt werden, wenn  $L^1$  aus  $j$  Spalten besteht).

Diejenigen Komponenten von  $\hat{d}$ , die in Zeilen liegen, in denen bereits pivotisiert wurde, liefern nun den Etavektor von  $U_k$ ; aus den anderen ergibt sich der Etavektor von  $L_k$ .

Wegen der in der Booleschen Phase geleisteten Vorarbeiten sind hierbei nur sehr wenige Suchvorgänge erforderlich:

- die Pivotzeile ist bereits bekannt;
- diejenigen vorausgegangenen Pivots, die für die Berechnung von  $\hat{d}$  relevant sind, sind bekannt; das  $L$ -File muss also nicht durchsucht werden, sondern auf die benötigten  $L_j$  kann direkt zugegriffen werden;

- die Indices der NNE in der aktualisierten Spalte sind von vornherein bekannt; dies erleichtert nicht nur die Abspeicherung der Etavektoren sondern auch das Löschen des Arbeitsbereichs für die nachfolgenden Operationen.

Ist ein vorbestimmtes Pivotelement dem Betrag nach zu klein bzw. durch Differenzbildung tatsächlich gleich Null, so wird die betreffende Spalte zunächst zurückgestellt bis alle anderen vorbestimmten Pivots durchgeführt worden sind. Die Verarbeitung dieser Spalten ist wesentlich aufwendiger, da keine Informationen aus der Booleschen Phase verwendet werden können. Die Pivot-Wahl in den zurückgestellten Spalten sowie in denjenigen Spalten, für die in der Booleschen Phase kein Pivotelement bestimmt wurde, erfolgt nach dem Gesichtspunkt der numerischen Stabilität: man entscheidet sich für die Komponente mit dem größten Absolutbetrag in einer Zeile, in der noch nicht pivotisiert worden ist (Pivoting for Size).  $\triangle$

## Spezialliteratur zum Simplexalgorithmus

Das folgende Literaturverzeichnis enthält eine (unvollständige) Sammlung einiger wichtiger Paper zu den in diesem Kapitel behandelten Themen.

### Literaturverzeichnis

- D. Avis and V. Chvátal. Notes on Bland's pivoting rule. *Mathematical Programming Study*, 8:24–34, 1978.
- R. H. Bartels. A stabilization of the simplex method. *Numerische Mathematik*, 16: 414–434, 1971.
- R. H. Bartels and G. H. Golub. The simplex method of linear programming using LU decomposition. *Communications of the Association for Computing Machinery*, 12: 266–268, 275–278, 1969.
- M. Bastian. *Lineare Optimierung großer Systeme*. Athenäum/Hain/Skriptor/Hanstein, Königstein, 1980.
- E. M. L. Beale and J. A. Tomlin. Special facilities in a general mathematical programming system for non-convex problems using sets of variables. In J. Lawrence, editor, *Proceedings of the fifth IFORS Conference*, pages 447–454, Tavistock, London, 1970.
- M. Bénichou, J. M. Gauthier, P. Girodet, G. Hentgès, G. Ribière, and O. Vincent. Experiments in mixed integer linear programming. *Mathematical Programming*, 1:76–94, 1971.
- M. Bénichou, J. M. Gauthier, G. Hentgès, and G. Ribière. The efficient solution of large-scale linear programming problems – some algorithmic techniques and computational results. *Mathematical Programming*, 13:280–322, 1977.

- R. E. Bixby. Solving real-world linear programs: A decade and more of progress. *Operations Research*, 50:3–15, 2002.
- R. Bland. New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2:103–107, 1977.
- R. K. Brayton, F. G. Gustavson, and R. A. Willoughby. Some results on sparse matrices. *Mathematics of Computation*, 24:937–954, 1970.
- V. Chvátal. *Linear Programming*. Freeman, New York, 1983.
- H. Crowder and J. M. Hattingh. Partially normalized pivot selection in linear programming. *Mathematical Programming Study*, 4:12–25, 1975.
- A. R. Curtis and J. K. Reid. On the automatic scaling of matrices for Gaussian elimination. *Journal of the Institute of Mathematics and its Applications*, 10:118–124, 1972.
- G. B. Dantzig, R. P. Harvey, R. D. McKnight, and S. S. Smith. Sparse matrix techniques in two mathematical programming codes. In R. A. Willoughby, editor, *Sparse Matrix Proceedings RA-1*, pages 85–99. IBM Research Center, Yorktown Heights, New York, 1969.
- R. Fletcher and S. P. J. Matthews. Stable modification of explicit LU factors for simplex updates. *Mathematical Programming*, 30:267–284, 1984.
- J. J. H. Forrest and J. A. Tomlin. Updated triangular factors of the basis to maintain sparsity in the product form simplex method. *Mathematical Programming*, 2:263–278, 1972.
- A. M. Geoffrion and R. E. Marsten. Integer programming algorithms: a framework and state-of-the-art survey. *Management Science*, 18:465–491, 1972.
- D. Goldfarb. Using the steepest edge simplex algorithm to solve sparse linear programs. In J. Bunch and D. Rose, editors, *Sparse Matrix Computations*, pages 227–240. Academic Press, New York, 1976.
- D. Goldfarb and J. K. Reid. A practicable steepest-edge simplex algorithm. *Mathematical Programming*, 12:361–371, 1977.
- F. G. Gustavson. Some basic techniques for solving sparse systems of linear equations. In D. J. Rose and R. A. Willoughby, editors, *Sparse Matrices and Their Applications*, pages 41–52. Plenum Press, New York, 1972.
- P. M. J. Harris. Pivot selection methods of the Devex LP code. *Mathematical Programming Studies*, 4:30–57, 1975.
- R. G. Jeroslow. The simplex algorithm with the pivot rule of maximizing improvement criterion. *Discrete Mathematics*, 4:367–377, 1973.

- V. Klee and G. J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities – III*, pages 159–175. Academic Press, New York, 1972.
- H. Kuhn and R. E. Quandt. An experimental study of the simplex method. In N. C. Metropolis, editor, “*Experimental Arithmetic, High-Speed Computing and Mathematics*”, *Proceedings of Symposia on Applied Mathematics XV*, pages 107–124, American Mathematical Society, Providence, RI, 1963.
- B. A. Murtagh. *Advanced Linear Programming: Computation and Practice*. McGraw-Hill, New York, 1981.
- S. Pissanetsky. *Sparse Matrix Technology*. Academic Press, London, 1984.
- M. A. Saunders. The complexity of LU updating in the simplex method. In R. S. Anderssen and R. P. Brent, editors, *The Complexity of Computational Problem Solving*, pages 214–230. Queensland University Press, Queensland, 1976a.
- M. A. Saunders. A fast, stable implementation of the simplex method using bartels-golub updating. In J. Bunch and D. Rose, editors, *Sparse Matrix Computations*, pages 213–226. Academic Press, New York, 1976b.
- M. J. Todd. Modifying the Forrest-Tomlin and Saunders updates for linear programming problems with variable upper bounds. Technical report, Cornell University, School of OR/IE, 1984. URL <http://hdl.handle.net/1813/8502>. Technical report 617.
- J. A. Tomlin. On scaling linear programming problems. *Mathematical Programming Studies*, 4:146–166, 1975.



# 13 Ganzzahligkeit von Polyedern und Totale Unimodularität

In der Vorlesung haben wir viele Beispiele von Optimierungsproblemen gesehen, die sich als ganzzahliges lineares Programm formulieren lassen. Mit dem Simplex-Algorithmus haben wir eine in der Praxis sehr effektive Methode zur Lösung linearer Programme. Wenn wir also wissen, dass die gesuchte Optimallösung einem ganzzahligen Punkt entspricht, können wir den Simplex-Algorithmus zur Lösung des Problems nutzen. Es ist daher interessant zu untersuchen, welche Polyeder die Eigenschaft haben, (nur) ganzzahlige Ecken bzw. Seitenflächen mit ganzzahligen Punkten zu besitzen.

**(13.1) Definition.** Ein Polyeder  $P$  heißt ganzzahlig, wenn jede nichtleere Seitenfläche einen ganzzahligen Punkt enthält.  $\triangle$

Da jede nichtleere Seitenfläche eine minimale nichtleere Seitenfläche enthält, ist ein Polyeder ganzzahlig genau dann, wenn jede minimale nichtleere Seitenfläche einen ganzzahligen Punkt enthält. Das leere Polyeder betrachten wir auch als ganzzahliges Polyeder. Wir nehmen in diesem Kapitel an, dass  $P \subseteq \mathbb{Q}^n$  eine Darstellung  $P = P(A, b)$  hat, bei der  $A$  eine rationale Matrix und  $b$  ein rationaler Vektor ist.

Die folgende Klasse von Matrizen spielt eine wichtige Rolle bei Ganzzahligkeitsüberlegungen.

**(13.2) Definition.** Eine Matrix  $A$  heißt total unimodular, wenn jede Subdeterminante von  $A$  entweder 0, 1 oder  $-1$  ist.  $\triangle$

**(13.3) Bemerkung.** Die folgenden Operationen erhalten totale Unimodularität:

1. Vertauschen von Zeilen oder Spalten
2. Transponieren
3. Multiplikation einer Zeile oder Spalte mit  $-1$
4. Pivotalisieren, d. h. Ersetzen der Matrix

$$\begin{pmatrix} a & c \\ b & D \end{pmatrix} \text{ durch } \begin{pmatrix} -a & ac \\ ab & D - abc \end{pmatrix}$$

für  $a \in \{-1, 1\}$ , Zeilenvektor  $c$ , Spaltenvektor  $b$  und Matrix  $D$

5. Hinzufügen einer Nullspalte oder Nullzeile sowie Hinzufügen einer Zeile oder Spalte mit genau einem Nichtnullelement mit Wert  $\pm 1$
6. Wiederholen einer Zeile oder einer Spalte  $\triangle$

**(13.4) Bemerkung.** Insbesondere folgt aus (13.3), dass, wenn  $A$  total unimodular ist, auch die folgenden Matrizen total unimodular sind:

$$-A, \quad A^T, \quad \begin{pmatrix} A \\ -A \end{pmatrix}, \quad \begin{pmatrix} I \\ A \end{pmatrix}, \quad (I \ A). \quad \triangle$$

Bekannte Resultate aus der linearen Algebra spielen bei den nachfolgenden Argumentationen eine wichtige Rolle. Ist  $A$  eine nichtsinguläre  $(n, n)$ -Matrix, dann kann man ihre inverse Matrix wie folgt darstellen:

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A),$$

wobei  $\operatorname{adj}(A)$  die transponierte Matrix der Matrix  $M$  mit den Elementen

$$m_{ij} = (-1)^{i+j} \det A_{ij}$$

ist. Hierbei ist  $A_{ij}$  die Matrix, die aus  $A$  durch Streichen der  $i$ -ten Zeile und  $j$ -ten Spalte entsteht. Aus dieser Formel können wir schließen, dass, wenn  $A$  eine ganzzahlige Matrix ist,  $\operatorname{adj}(A)$  ebenfalls eine ganzzahlige Matrix ist. Darüber hinaus folgt damit aus  $\det A = \pm 1$  die Ganzzahligkeit von  $A^{-1}$ . Folglich ist die inverse Matrix einer total unimodularen Matrix ganzzahlig.

Wir erinnern hier ferner an die Cramersche Regel, die besagt, dass für eine nichtsinguläre Matrix  $A$  und einen Vektor  $b$  die eindeutig bestimmte Lösung des Gleichungssystems  $Ax = b$  durch die Formel

$$x_i = \frac{1}{\det A} \det A_i$$

gegeben ist, wobei  $A_i$  die Matrix ist, die aus  $A$  durch Ersetzen von Spalte  $i$  durch den Vektor  $b$  entsteht. Ist  $b$  ein ganzzahliger Vektor und  $A$  total unimodular, so ist der Lösungsvektor  $x$  offenbar ganzzahlig.

**(13.5) Satz.** *Seien  $A$  eine total unimodulare Matrix und  $b$  ein ganzzahliger Vektor. Dann ist das Polyeder  $P(A, b)$  ganzzahlig.*  $\triangle$

**Beweis.** Sei  $F = \{x \in P \mid A'x = b'\}$  eine minimale Seitenfläche von  $P$ , wobei  $A'x = b'$  als Teilsystem von  $A_{\operatorname{eq}(F)} \cdot x = b_{\operatorname{eq}(F)}$  so gewählt sei, dass  $A'$  vollen Zeilenrang besitzt. Dann können wir (nach Zeilen- und Spaltentausch)  $A'$  schreiben als  $A' = (U \ V)$  für eine Teilmatrix  $U$  mit  $\det U = \pm 1$ . Nach den obigen Überlegungen ist  $U^{-1}$  ganzzahlig. Damit ist

$$x = \begin{pmatrix} U^{-1}b' \\ 0 \end{pmatrix}$$

ein ganzzahliger Punkt, der das System  $A'x = b'$  löst. Wegen Satz (8.18) ist  $F = \{x \in \mathbb{Q}^n \mid A'x = b'\}$  und  $x$  daher in  $F$ .  $\square$



**(13.6) Folgerung.** Seien  $A$  eine total unimodulare Matrix und  $b$  und  $c$  ganzzahlige Vektoren. Dann haben beide LPs des primal-dualen Paares

$$\max\{c^T x \mid Ax \leq b\}, \quad \min\{u^T b \mid u^T A = c, u \geq 0\}$$

optimale ganzzahlige Lösungen. △

**Beweis.** Folgt direkt aus Satz (13.5) und Bemerkung (13.4), da mit  $A$  auch die Matrix

$$\begin{pmatrix} I \\ A \\ -A \end{pmatrix}$$

total unimodular ist. □

Satz (13.5) zeigt, dass Polyeder der Form  $P(A, b)$  ganzzahlig sind, wenn  $A$  total unimodular ist. Ein ähnliches Resultat gilt für Polyeder der Form  $P^=(A, b)$ , wofür wir aber eine weitere Definition benötigen.

**(13.7) Definition.** Eine ganzzahlige reguläre quadratische Matrix  $A$  heißt unimodular, wenn ihre Determinante 1 oder  $-1$  ist. Allgemeiner heißt eine ganzzahlige Matrix  $A \in \mathbb{Q}^{(m,n)}$  mit vollem Zeilenrang unimodular, wenn jede Basis von  $A$  unimodular ist, also Determinante 1 oder  $-1$  hat. △

Man kann zeigen, dass  $A$  genau dann total unimodular ist, wenn die Matrix  $\begin{pmatrix} I & A \end{pmatrix}$  unimodular ist.

**(13.8) Satz.** Sei  $A$  eine ganzzahlige Matrix mit vollem Zeilenrang. Das Polyeder  $P^=(A, b)$  ist ganzzahlig für jeden ganzzahligen Vektor  $b$  genau dann, wenn  $A$  unimodular ist. △

**Beweis.** Sei  $A \in \mathbb{Q}^{(m,n)}$  und nehmen wir zunächst an, dass  $A$  unimodular ist. Sei  $x'$  eine Ecke des Polyeders  $P^=(A, b)$  für einen ganzzahligen Vektor  $b$ . Nach Satz (8.9) sind die Spalten  $A_{.j}$ ,  $j \in \text{supp}(x')$  linear unabhängig und können daher zu einer Basis  $B$  von  $A$  erweitert werden. Dann ist  $x'_B = B^{-1}b$  und wegen  $\det B = \pm 1$  ganzzahlig; da  $x'_i = 0$  für  $i \in \{1, \dots, n\} \setminus B$ , ist  $x'$  ganzzahlig.

Sei nun  $P^=(A, b)$  ganzzahlig für jeden ganzzahligen Vektor  $b$ , und  $B$  sei eine beliebige Basis von  $A$ . Um Unimodularität von  $A$  zu zeigen, genügt es aufgrund der Cramerschen Regel zu zeigen, dass  $B^{-1}t$  für jeden ganzzahligen Vektor  $t$  ganzzahlig ist. Sei also  $t$  ganzzahlig. Dann existiert ein ganzzahliger Vektor  $y$ , sodass  $z := y + B^{-1}t \geq 0$ . Weil  $A, t$  und  $y$  ganzzahlig sind, ist auch  $Bz = By + t =: b$  ganzzahlig. Wir erweitern  $z$  zu  $z'$ , indem wir Nullkomponenten hinzufügen, so dass  $Az' = Bz = b$  gilt. Dann ist  $z'$  eine Ecke von  $P^=(A, b)$ , also nach Voraussetzung ganzzahlig. Daher ist auch  $z$  und damit  $B^{-1}t = z - y$  ganzzahlig. □

**(13.9) Folgerung (Satz von Hoffman und Kruskal).** Sei  $A$  eine ganzzahlige Matrix. Dann ist  $A$  total unimodular genau dann, wenn für jeden ganzzahligen Vektor  $b$  das Polyeder  $\{x \in \mathbb{Q}^n \mid x \geq 0, Ax \leq b\}$  ganzzahlig ist. △

**Beweis.** Wir haben bereits erwähnt, dass  $A$  genau dann total unimodular ist, wenn die Matrix  $(I \ A)$  unimodular ist. Für jeden ganzzahligen Vektor  $b$  sind die Ecken des Polyeders  $\{x \in \mathbb{Q}^n \mid x \geq 0, Ax \leq b\}$  genau dann ganzzahlig, wenn die Ecken des Polyeders

$$\left\{ \begin{pmatrix} y \\ x \end{pmatrix} \in \mathbb{Q}^{m+n} \mid (I \ A) \begin{pmatrix} y \\ x \end{pmatrix} = b, x \geq 0, y \geq 0 \right\}$$

ganzzahlig sind. Damit ergibt sich die Folgerung direkt aus Satz (13.8).  $\square$

Da totale Unimodularität eine sehr nützliche Eigenschaft ist, ist sie sehr gut untersucht, und es gibt eine ganze Reihe weiterer Charakterisierungen von total unimodularen Matrizen, die in verschiedenen Zusammenhängen nützlich sind.

**(13.10) Satz (Charakterisierungen total unimodularer Matrizen).** Sei  $A$  eine Matrix mit Einträgen  $0, 1$  und  $-1$ . Dann sind äquivalent:

- (a)  $A$  ist total unimodular.
- (b) Für jeden ganzzahligen Vektor  $b$  hat das Polyeder  $\{x \in \mathbb{Q}^n \mid Ax \leq b, x \geq 0\}$  nur ganzzahlige Ecken.
- (c) Für alle ganzzahligen Vektoren  $a, b, c, d$  hat das Polyeder  $\{x \in \mathbb{Q}^n \mid a \leq Ax \leq b, c \leq x \leq d\}$  nur ganzzahlige Ecken.
- (d) Jede Teilmenge der Spalten von  $A$  kann in zwei Teilmengen  $M^+$  und  $M^-$  geteilt werden, sodass die Summe der Spalten aus  $M^+$  minus die Summe der Spalten aus  $M^-$  einen Vektor ergibt, dessen Komponenten  $0, 1$  oder  $-1$  sind.
- (e) Jede nichtsinguläre Teilmatrix von  $A$  hat eine Zeile mit einer ungeraden Anzahl von Nichtnull-Komponenten.
- (f) Die Summe der Einträge jeder quadratischen Teilmatrix von  $A$  mit gerader Zeilen- und Spaltensumme ist durch  $4$  teilbar.
- (g) Keine quadratische Teilmatrix von  $A$  hat Determinante  $2$  oder  $-2$ .  $\triangle$

**(13.11) Beispiel (Inzidenzmatrizen bipartiter Graphen).** Sei  $G = (V, E)$  ein ungerichteter bipartiter Graph und sei  $M$  die Knoten-Kanten-Inzidenzmatrix von  $G$ , also

$$M_{v,e} = \begin{cases} 1 & v \in e, \\ 0 & \text{sonst.} \end{cases}$$

Um zu zeigen, dass  $M$  total unimodular ist, wollen wir (13.10)(d) verwenden. Wir nutzen aus, dass eine Matrix  $A$  genau dann total unimodular ist, wenn  $A^T$  total unimodular ist, d. h. es genügt zu zeigen, dass wir die Zeilen von  $M$  so in zwei Gruppen einteilen können, dass die Komponenten des Differenzvektors der Gruppensummen nur die Werte  $1, -1$  oder  $0$  annehmen. Nach Voraussetzung ist  $G$  bipartit, d. h. es existiert eine Zerlegung von

$V$  in Knotenmengen  $V_1$  und  $V_2$  sodass jede Kante einen Knoten in  $V_1$  und einen Knoten in  $V_2$  enthält. Summieren wir alle Zeilen von  $M$ , die Knoten in  $V_1$  entsprechen, so erhalten wir den Vektor  $\mathbf{1} \in \mathbb{Q}^E$ , da jede Kante mit genau einem Knoten aus  $V_1$  inzident ist. Analoges gilt für die  $V_2$ -Zeilensumme, sodass sich als Differenz der Nullvektor ergibt. Damit ist die totale Unimodularität von  $M$  gezeigt.  $\triangle$

**(13.12) Beispiel (Inzidenzmatrizen gerichteter Graphen).** Sei  $D = (V, A)$  ein gerichteter Graph und sei  $M$  die Knoten-Bogen-Inzidenzmatrix von  $G$ , d. h.

$$M_{v,a} = \begin{cases} 1 & a = (u, v), \\ -1 & a = (v, u), \\ 0 & \text{sonst.} \end{cases}$$

Dann ist  $M$  total unimodular, was man wieder über (13.10)(d) zeigen kann.  $\triangle$

**(13.13) Beispiel (Netzwerkmatrizen).** Seien  $D = (V, A)$  ein gerichteter Graph und  $T = (V, A_0)$  eine aufspannende Arboreszenz in  $D$ . Wir definieren die  $(A_0 \times A)$ -Matrix für  $a' \in A_0$  und  $a = (u, v)$  über

$$M_{a',a} := \begin{cases} 1 & \text{der eindeutige } [u, v]\text{-Weg in } T \text{ nutzt } a' \text{ vorwärts,} \\ -1 & \text{der eindeutige } [u, v]\text{-Weg in } T \text{ nutzt } a' \text{ rückwärts,} \\ 0 & \text{der eindeutige } [u, v]\text{-Weg in } T \text{ nutzt } a' \text{ nicht.} \end{cases}$$

Matrizen, die so dargestellt werden können, heißen *Netzwerkmatrizen* und sind total unimodular.  $\triangle$

Netzwerkmatrizen sind die Basis aller total unimodularen Matrizen, was durch das im Folgenden skizzierte Resultat von Seymour deutlich wird. Es sind jedoch nicht alle total unimodularen Matrizen Netzwerkmatrizen oder Transponierte davon; insbesondere sind die beiden folgenden Matrizen keine Netzwerkmatrizen, aber dennoch total unimodular:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & -1 \\ -1 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 \\ -1 & 0 & 0 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}. \quad (13.14)$$

Hier besteht ein tiefer Bezug zur Theorie regulärer Matroide, der z. B. in Schrijver (1998) dargestellt ist, auf den wir an dieser Stelle leider nicht eingehen können.

Seymour zeigte, dass jede total unimodulare Matrix auf eine bestimmte Weise aus Netzwerkmatrizen und den Matrizen (13.14) hervorgeht.

## Literaturverzeichnis

Neben den in (13.3) aufgeführten Operationen erhalten auch die folgenden Kombinationsoperationen für Matrizen totale Unimodularität:

$$A \oplus_1 B := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \quad (13.15)$$

$$(A \ a) \oplus_2 \begin{pmatrix} b \\ B \end{pmatrix} := \begin{pmatrix} A & ab \\ 0 & B \end{pmatrix}, \quad (13.16)$$

$$\begin{pmatrix} A & a & a \\ c & 0 & 1 \end{pmatrix} \oplus_3 \begin{pmatrix} 1 & 0 & b \\ d & d & B \end{pmatrix} := \begin{pmatrix} A & ab \\ dc & B \end{pmatrix}, \quad (13.17)$$

wobei  $A$  und  $B$  Matrizen,  $a$  und  $d$  Spaltenvektoren sowie  $b$  und  $c$  Zeilenvektoren passender Dimension sind. Dass diese Kombinationsoperationen totale Unimodularität erhalten, kann man unter Verwendung der Charakterisierung (13.10)(d) sehen.

**(13.18) Satz (Dekompositionssatz von Seymour für total unimodulare Matrizen).** *Eine Matrix  $A$  ist total unimodular genau dann, wenn sie aus Netzwerkmatrizen und den Matrizen in (13.14) durch die Operationen 1 bis 6 aus (13.3) bzw. durch die Kombinationsoperationen (13.15)–(13.17) erzeugt werden kann. Die Kombinationsoperationen (13.15)–(13.17) werden dabei nur angewendet, wenn sowohl für  $A$  als auch für  $B$  die Summe aus Spalten- und Zeilenanzahl mindestens 4 ist.*  $\triangle$

Aufbauend auf Satz (13.18) kann ein Polynomialzeitalgorithmus für die Erkennung total unimodularer Matrizen entwickelt werden.

## Literaturverzeichnis

A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, New York, 1998.