Optimierung

Vorlesungsskriptum Sommersemester 2020

R. Verfürth

Fakultät für Mathematik, Ruhr-Universität Bochum

Inhaltsverzeichnis

Einleitung

Die folgenden stark vereinfachten Beispiele geben einige typische Optimierungsprobleme an, mit deren Lösung wir uns in den folgenden Kapiteln befassen.

Beispiel 1. Eine kleine Schuhfabrik stellt je ein Modell eines Damen- und Herrenschuhs her. Der Reingewinn pro Paar Damenschuh beträgt 16 € , pro Paar Herrenschuh 32 € . Der Lederbedarf pro Paar Damen- bzw. Herrenschuh beträgt $6\,\mathrm{dm}^2$ bzw. $15\,\mathrm{dm}^2$; monatlich stehen $4500\,\mathrm{dm}^2$ Leder zur Verfügung. Die Maschinen-Bearbeitungszeit beträgt 4 h bzw. 5 h pro Paar Damen- bzw. Herrenschuh; monatlich stehen höchstens 2000 h Maschinenzeit zur Verfügung. Die menschliche Arbeitszeit beträgt 20 h bzw. 10 h pro Paar Damen- bzw. Herrenschuh; die monatliche Gesamtarbeitszeit beträgt höchstens 8000 h. Die Firma möchte diese Ressourcen optimal einsetzen und einen maximalen monatlichen Gewinn erzielen.
Zur Lösung bezeichnen wir mit x die Zahl der produzierten Paar Damenschuhe und mit y die Zahl der produzierten Paar Herrenschuhe. Dann ist der Gewinn

$$16x + 32y.$$

Die vorhandenen Ressourcen ergeben folgende Bedingungen:

$$\begin{aligned} \text{Leder:} \quad & 6x + 15y \leq 4500, \\ \text{Maschinen:} \quad & 4x + 5y \leq 2000, \\ \text{Mensch:} \quad & 20x + 10y \leq 8000. \end{aligned}$$

Natürlich können keine negativen Schuhe produziert werden, d.h.

$$x \geq 0, \quad y \geq 0.$$

Diese 5 Ungleichungen beschreiben die in Abbildung 1 fett umrandete Fläche. Die dünnen Linien beschreiben Niveaulinien der Funktion $16x + 32y$.
Offensichtlich ist der Gewinn maximal im Punkt B. Dieser hat die Koordinaten $(250, 200)$, d.h. 250 Paar Damenschuhe und 200 Paar Herrenschuhe. Der Gewinn beträgt 10400 € .

Beispiel 2. Ein Transportunternehmen verfügt über 18 Güterwagen am Gahnhof A und 12 Güterwagen am Bahnhof B. Sie benötigt 11, 10 bzw. 9 Güterwagen an den Bahnhöfen R, S, T. Die folgende Tabelle gibt die Entfernungen in km zwischen den einzelnen Bahnhöfen an:

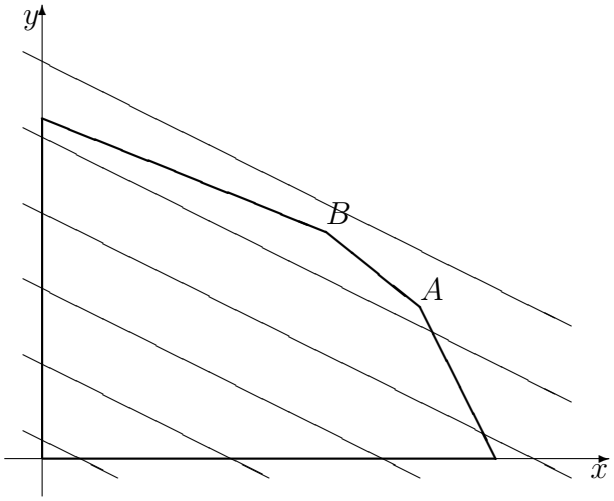

ABBILDUNG 1. Zulässiger Bereich und einige Niveaulinien der Gewinnfunktion in Beispiel 1

	R	S	T
A	5	4	9
B	7	8	10

Die Güterwagen sind nun so zu arrangieren, dass die Gesamtzahl der gefahrenen Kilometer minimal ist.
Da die Zahl der zur Verfügung stehenden Güterwagen gleich derjenigen der angeforderten ist, ist die Zahl der jeweils zu verschiebenden Waggons eindeutig festgelegt durch die Zahlen x und y der Waggons, die von A nach R bzw. von A nach S verschoben werden. Damit ergibt sich folgende „Verschiebetabelle“:

	R	S	T
A	x	y	$18-x-y$
B	$11-x$	$10-y$	$x+y-9$

Die gefahrenen Leerkilometer sind

$$\begin{aligned} &5x+4y+9(18-x-y)\\ &+7(11-x)+8(10-y)+10(x+y-9)\\ &=-x-3y+229. \end{aligned}$$

Diese Funktion ist zu minimieren unter denEinschränkungen

$$\begin{aligned} x\geq 0, &\qquad y\geq 0\\ 18-x-y\geq 0, &\qquad 11-x\geq 0\\ 10-y\geq 0, &\qquad x+y-9\geq 0. \end{aligned}$$

Diese 6 Ungleichungen beschreiben den fett umrandeten Bereich in Abbildung 2. Die dünnen Linien beschreiben Niveaulinien der Funktion $229-x-3y$.

Offensichtlich ist diese Funktion im Punkt A minimal. Dieser hat die Koordinaten $(8, 10)$. Es sind insgesamt 191 Leerkilometer zu fahren.

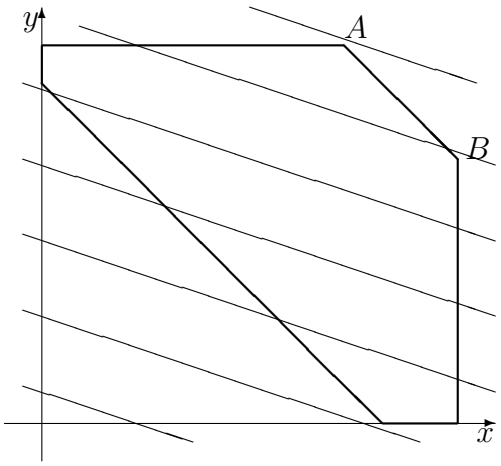

Abbildung 2. Zulässiger Bereich und einige Niveaulinien der Verlustfunktion in Beispiel 2

Beispiel 3. Ein Fürst will die drei wichtigsten Städte A, B, C seines Reiches besuchen. Er startet seine Reise in seiner Residenz R und will sie dort beenden. Damit niemand bevorzugt erscheint, darf er jede der Städte A, B und C nur genau einmal betreten. Die folgende Tabelle gibt die Fahrtzeiten zwischen den Städten an:

	R	A	B	C
R	–	8	6	15
A	8	–	2	4
B	6	2	–	5
C	15	4	5	–

Wie muss der Fürst seinen Weg gestalten, um eine möglichst kurze Strecke zurückzulegen?
Da offensichtlich Wege der Form R $\to$ A $\to$ B $\to$ C $\to$ R und R $\to$ C $\to$ B $\to$ A $\to$ R die gleiche Länge haben, hat er bis auf Umkehr der Reihenfolge folgende drei Möglichkeiten:

$$\begin{aligned} \mathrm{R} \to \mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{R}: &\quad 8+2+5+15=30, \\ \mathrm{R} \to \mathrm{B} \to \mathrm{C} \to \mathrm{A} \to \mathrm{R}: &\quad 6+5+4+8=23, \\ \mathrm{R} \to \mathrm{C} \to \mathrm{A} \to \mathrm{B} \to \mathrm{R}: &\quad 15+4+2+6=27. \end{aligned}$$

Er muss also die Städte in der Reihenfolge B, C, A aufsuchen.

Beispiel 4. Eine Ölfirma hat eine Quelle und eine Raffinerie, die wie in Abbildung 3 skizziert durch Pipelines miteinander verbunden sind. Die Zahlen geben die maximale Durchflussmenge jeder Pipeline an. Wie groß ist die maximale Ölmenge, die von der Quelle Q zur

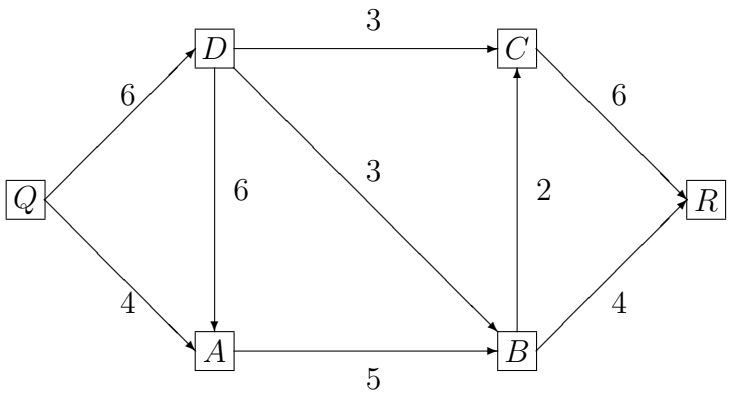

ABBILDUNG 3. Ölpipelines aus Beispiel 4 mit maximaler Durchflussmenge jeder Pipeline

Raffinerie R transportiert werden kann?
Durch „Hinsehen“ und Ausprobieren erhalten wir den maximalen Fluss 9, der sich wie in Abbildung 4 skizziert verteilt.

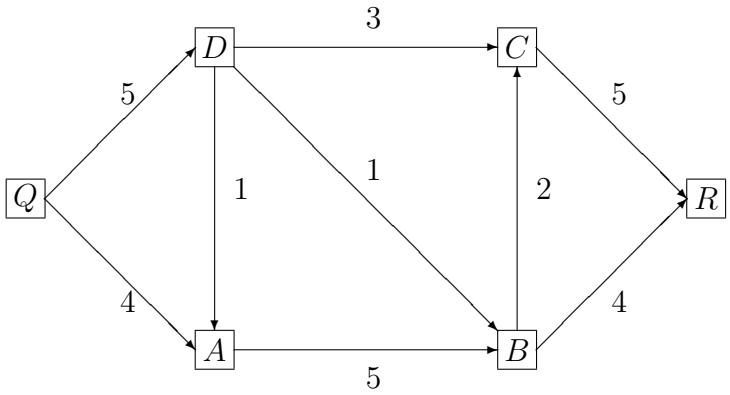

ABBILDUNG 4. Ölpipelines aus Beispiel 4 mit der Durchflussmenge jeder Pipeline, die den maximalen Gesamtfluss von Q nach R ergibt

BEISPIEL 5. Ein Paketdienst hat einen Sondertarif eingeführt. Er betrifft quaderförmige Pakete, die folgende Bedingungen erfüllen:

- Jede Kante ist höchstens 40 cm lang.
- Für jede Seitenfläche ist der Umfang plus die Länge der zur Seitenfläche senkrechten Kante höchsten 170 cm.

Wie groß ist das maximale Volumen eines Paketes, das zu dem Sondertarif befördert werden kann?

Wir bezeichnen mit x, y und z die Kantenlängen. Dann ist das Volumen

$$V = xyz.$$

Wir können die Bezeichnungen so wählen, dass

$$x \leq y \leq z$$

gilt. Dann ergeben sich die Einschränkungen

$$z \leq 40, \quad 2(z+y) + x \leq 170.$$

Offensichtlich erhalten wir das maximale Volumen, wenn wir x, y und z möglichst groß wählen und dabei die Bedingungen einhalten. Dies liefert die Kandidaten

$$z = 40, x = y = 30 \text{ mit } V = 36\,\mathrm{dm}^3$$

und

$$x = y = z = 34 \text{ mit } V = 39.304\,\mathrm{dm}^3\,.$$

Also ist das maximale Volumen $39.304\,\mathrm{dm}^3$.

Beispiel 6. In einem Frachtraum, der die Form eines Ellipsoides

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

hat, ist ein achsenparalleler Quader maximaler Länge unterzubringen. Wie groß ist das maximale Volumen?
Die Koordinaten der Quadereckpunkte sind $(\pm x, \pm y, \pm z)$. Dementsprechend ist das Volumen

$$V = 8xyz.$$

Dieses ist zu maximieren unter der Nebenbedingung

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Hierzu stellen wir die Lagrange-Funktion (s. Skriptum „Analysis I-III“, Satz VII.5.14, S. 253) auf

$$L(x, y, z, \lambda) = 8xyz + \lambda \left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 \right)$$

und bestimmen deren kritische Punkte aus den Gleichungen

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 8yz - 2\lambda \frac{x}{a^2} &&= 0 \\
\frac{\partial L}{\partial y} &= 8xz - 2\lambda \frac{y}{b^2} &&= 0 \\
\frac{\partial L}{\partial z} &= 8xy - 2\lambda \frac{z}{c^2} &&= 0 \\
\frac{\partial L}{\partial \lambda} &= \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 &&= 0.
\end{aligned}$$

Multiplizieren wir die erste, zweite und dritte Gleichung mit x, y und z und addieren die Ergebnisse, erhalten wir mit der vierten Gleichung

$$0 = 24xyz - 2\lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right) = 24xyz - 2\lambda$$
$$\Longrightarrow \quad \lambda = 12xyz.$$

Setzen wir dies in die ersten drei Gleichungen ein, erhalten wir

$$0 = 8yz\left(1 - 3\frac{x^2}{a^2}\right)$$
$$0 = 8xz\left(1 - 3\frac{y^2}{b^2}\right)$$
$$0 = 8xy\left(1 - 3\frac{z^2}{c^2}\right).$$

Damit ergeben sich die Lösungen

$$x = y = z = 0$$

und

$$x = \frac{a}{\sqrt{3}}\ ,\ y = \frac{b}{\sqrt{3}}\ ,\ z = \frac{c}{\sqrt{3}}.$$

Da die erste Lösung offensichtlich unsinnig ist, ist das maximale Volumen

$$V = \frac{8}{9}\sqrt{3}abc.$$

Alle Beispiele sind von der Form, dass eine Größe, genannt *Zielfunktion*, unter gewissen Nebenbedingungen, die durch Gleichungen oder Ungleichungen für weitere Funktionen beschrieben werden, minimiert oder maximiert werden soll.
In den ersten beiden Beispielen sind alle Funktionen linear. Dementsprechend spricht man von einem *linearen Optimierungsproblem*. Mit solchen Problemen befassen wir uns in Kapitel I.
Das dritte und vierte Beispiel unterscheiden sich von den anderen Beispielen dadurch, dass die auftretenden Größen nur diskrete Werte annehmen können. Dementsprechend spricht man von einem *diskreten Optimierungsproblem*. Diese sind Gegenstand des Kapitels II.
Beim fünften und sechsten Beispiel schließlich sind die Zielfunktion und / oder einige oder alle Nebenbedingungen nichtlineare Funktionen. Dementsprechend spricht man von einem *nichtlinearen Optimierungsproblem*. Solche Probleme behandeln wir in Kapitel III.
Probleme wie in Beispiel 1 können schnell auch zu diskreten Optimierungsproblemen werden. Dann nämlich, wenn das Optimum nicht ganzzahligen Werten entspricht und aus praktischen Gründen nur ganzzahlige Lösungen in Frage kommen. (Es macht keinen Sinn, $\frac{1}{3}$ Paar Damenschuhe zu produzieren!) Insofern waren die Zahlen in diesem

Beispiel glücklich gewählt. Mit derartigen Problemen befassen wir uns in Abschnitt II.1.

Die Kapitel I und III folgen im Wesentlichen der Monographie [2]. Kapitel II orientiert sich im Wesentlichen an dem Abschnitt „Ganzzahlige / Kombinatorische Optimierung“ in [1].

KAPITEL I

Lineare Optimierung

I.1. Problemstellung

DEFINITION I.1.1. Wir definieren eine Halbordnung „$\leq$“ auf $[\mathbb{R} \cup \{-\infty, \infty\}]^n$ durch

$$x \leq y \quad \Longleftrightarrow \quad x_i \leq y_i \; \forall 1 \leq i \leq n.$$

DEFINITION I.1.2. Gegeben seien ein Vektor $c \in \mathbb{R}^n$, eine Matrix $A \in \mathbb{R}^{m \times n}$, Vektoren $\underline{b}, \overline{b} \in [\mathbb{R} \cup \{-\infty, \infty\}]^m$ und Vektoren $\ell, u \in [\mathbb{R} \cup \{-\infty, \infty\}]^n$. Dann nennt man die Aufgabe:

$$c^t x \to \min \tag{I.1.1}$$

unter den Nebenbedingungen

$$x \in \mathbb{R}^n, \quad \underline{b} \leq Ax \leq \overline{b}, \quad \ell \leq x \leq u$$

ein *lineares Optimierungsproblem*, kurz *LP*, in allgemeiner Form. Für ein LP schreiben wir auch kurz

$$\min\{c^t x : \underline{b} \leq Ax \leq \overline{b}\ ,\ \ell \leq x \leq u\}.$$

Die Menge

$$\mathcal{P} = \{x \in \mathbb{R}^n : \underline{b} \leq Ax \leq \overline{b}\ ,\ \ell \leq x \leq u\}$$

heißt *Zulässigkeitsbereich* des LP.

BEMERKUNG I.1.3. Wegen

$$\max c^t x = -\min(-c^t x)$$

können wir uns auf Minimierungsprobleme beschränken.

BEISPIEL I.1.4. In Beispiel 1 (S. 5) ist $n = 2$, $m = 3$

$$A = \begin{pmatrix} 6 & 15 \\ 4 & 5 \\ 20 & 10 \end{pmatrix}, \quad \underline{b} = \begin{pmatrix} -\infty \\ -\infty \\ -\infty \end{pmatrix}, \quad \overline{b} = \begin{pmatrix} 4500 \\ 2000 \\ 8000 \end{pmatrix},$$

$$\ell = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad u = \begin{pmatrix} \infty \\ \infty \end{pmatrix} \qquad -c = \begin{pmatrix} 16 \\ 32 \end{pmatrix}.$$

In Beispiel 2 (S. 5) ist $n = 2$, $m = 1$

$$A = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad \underline{b} = 9, \qquad \overline{b} = 18,$$

$$\ell = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad u = \begin{pmatrix} 11 \\ 10 \end{pmatrix} \quad c = \begin{pmatrix} -1 \\ -3 \end{pmatrix}.$$

Man beachte, dass die additive Konstante 229 in der zu minimierenden Zielfunktion zwar den Wert der Zielfunktion nicht aber die Minimalstelle beeinflusst.

Für die konkrete Anwendung ist es hilfreich, statt der allgemeinen Form eines LP eine dazu äquivalente, standardisierte Form zu betrachten. Wir benutzen im Folgenden zwei standardisierte Formen: Die erste ist für theoretische Untersuchungen besonders geeignet; die zweite ist besonders günstig für die praktische Lösung mit dem Simplexalgorithmus.

DEFINITION I.1.5. Die *Standardform* eines LP ist gegeben durch

$$\min\{c^t x : Ax = b\ ,\ x \geq 0\} \tag{I.1.2}$$

mit der zugehörigen zulässigen Menge

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

Wir bezeichnen dieses Problem im Folgenden kurz mit *LP(P)* oder einfach P.

BEMERKUNG I.1.6. Jedes LP kann in ein LP(P) transformiert werden und umgekehrt. Die verschiedenen Formen sind ssquivalent in dem Sinne, dass die Zielfunktionen die gleichen Minimalwerte haben und dass jedes Minimum des einen LP in das des anderen LP transformiert werden kann.
Um dies einzusehen, betrachten wir zunächst die Richtung $P \to LP$. Offensichtlich entspricht Problem (I.1.2) dem Problem (I.1.1) mit

$$\underline{b} = b \quad \overline{b} = \overline{b},$$

$$\ell = 0, \quad u = \begin{pmatrix} \infty \\ \vdots \\ \infty \end{pmatrix}.$$

Für die Richtung $LP \to P$ beachten wir zuerst, dass Ungleichungen der Form

$$-\infty \leq (Ax)_i, \quad (Ax)_j \leq \infty$$

ignoriert werden können. Ungleichungen der Form

$$\underline{b}_i \leq (Ax)_i, \quad (Ax)_j \leq \overline{b}_j$$

werden durch Einführen von sog. *Schlupfvariablen* in die Standardform überführt:

$$\underline{b}_i = (Ax)_i - s_i, \quad s_i \geq 0,$$
$$\overline{b}_j = (Ax)_j + s_j, \quad s_j \geq 0.$$

Ungleichungen der Form

$$-\infty \leq x_i \leq \infty$$

gehen durch den Ansatz

$$x_i = z_i - y_i, \quad z_i \geq 0, \quad y_i \geq 0$$

in die Standardform über. Für endliche Werte ℓ_i, u_i lauten die entsprechenden Umformungen

$$\begin{aligned}
\ell_i \le x_i \le \infty \Longrightarrow \quad & y_i \ge 0 \quad \text{mit} \quad y_i = x_i - \ell_i, \\
-\infty \le x_i \le u_i \Longrightarrow \quad & z_i \ge 0 \quad \text{mit} \quad z_i = u_i - x_i, \\
\ell_i \le x_i \le u_i \Longrightarrow \quad & y_i \ge 0 \quad \text{und} \quad z_i \ge 0 \\
\text{mit} \quad & y_i = x_i - \ell_i, \quad z_i = u_i - x_i.
\end{aligned}$$

Man beachte, dass sich bei diesen Transformationen die beteiligten Vektoren und Matrizen, sowie die Dimensionen m und n ändern.

Beispiel I.1.7. Die Standardform von Beispiel 1 (S. 5) ist gegeben durch

$$n = 5, \qquad m = 3,$$

$$A = \begin{pmatrix} 6 & 15 & 1 & 0 & 0 \\ 4 & 5 & 0 & 1 & 0 \\ 20 & 10 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 4500 \\ 2000 \\ 8000 \end{pmatrix}, \quad -c = \begin{pmatrix} 16 \\ 32 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Die Erhöhung der Dimension n von 2 auf 5 ergibt sich durch die Einführung von 3 Schlupfvariablen, die sich in den 3 Null-Komponenten von c widerspiegeln.
Für Beispiel 2 (S. 5) erhalten wir analog

$$n = 6, \qquad m = 4,$$

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 18 \\ 11 \\ 10 \\ -9 \end{pmatrix}, \quad c = \begin{pmatrix} -1 \\ -3 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Definition I.1.8. Die *Simplexform* eines LP ist gegeben durch

$$\max \left\{ z \in \mathbb{R} : \begin{pmatrix} A & 0 \\ c^t & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}, \ x \ge 0 \right\} \tag{I.1.3}$$

mit der zulässigen Menge

$$\widehat{\mathcal{P}} = \left\{ (x, z) \in \mathbb{R}^{n+1} : \begin{pmatrix} A & 0 \\ c^t & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}, \ x \ge 0 \right\}.$$

Wir bezeichnen dieses Problem im Folgenden kurz mit *LP($\widehat{P}$)* oder einfach $\widehat{P}$. Ebenso setzen wir zur Abkürzung

$$\widehat{x} = \begin{pmatrix} x \\ z \end{pmatrix}, \quad \widehat{A} = \begin{pmatrix} A & 0 \\ c^t & 1 \end{pmatrix}, \quad \widehat{b} = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

BEMERKUNG I.1.9. Offensichtlich gilt

$$x \in \mathcal{P} \quad \Longleftrightarrow \quad \widehat{x} = \begin{pmatrix} x \\ -c^t x \end{pmatrix} \in \widehat{\mathcal{P}}$$

und die Minimierung von $c^t x$ ist äquivalent zur Maximierung von z unter der Nebenbedingung

$$c^t x + z = 0.$$

Daher sind die Probleme (I.1.2) und (I.1.3) äquivalent.

I.2. Geometrische Grundlagen

DEFINITION I.2.1. Eine Menge der Form

$$H = \{x \in \mathbb{R}^n : a^t x \leq b\}$$

mit $a \in \mathbb{R}^n$ und $b \in \mathbb{R}$ heißt ein *Halbraum*. Der Durchschnitt von endlich vielen Halbräumen heißt ein *Polyeder*.

BEMERKUNG I.2.2. Ein Polyeder ist stets von der Form

$$\{x \in \mathbb{R}^n : Ax \leq b\}$$

mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Da die Gleichung

$$Ax = b$$

offensichtlich äquivalent ist zu den beiden Ungleichungen

$$Ax \leq b, \quad (-A)x \leq -b$$

ist die zulässige Menge eines LP in Standardform ein Polyeder.

DEFINITION I.2.3. Eine Menge $M \subset \mathbb{R}^n$ heißt *konvex*, wenn für alle $x, y \in M$ und alle $\lambda \in [0, 1]$ gilt

$$\lambda x + (1 - \lambda) y \in M.$$

Eine Funktion $f : M \to \mathbb{R}$ heißt *konvex*, wenn M konvex ist und für alle $x, y \in M$ und alle $\lambda \in [0, 1]$ gilt

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y). \tag{I.2.1}$$

Die Funktion f heißt *strikt konvex*, wenn für alle $x, y \in M$ mit $x \neq y$ und alle $\lambda \in (0, 1)$ in (I.2.1) die strikte Ungleichung gilt.

BEMERKUNG I.2.4. Halbräume sind konvex. Lineare Funktionen sind konvex, aber nicht strikt konvex.

LEMMA I.2.5. *Der Durchschnitt von beliebig vielen konvexen Mengen ist konvex. Insbesondere ist jeder Polyeder konvex.*

BEWEIS. Ist offensichtlich. □

DEFINITION I.2.6. Eine Menge $A \subset \mathbb{R}^n$ heißt *affin* oder *affine Menge*, wenn für alle $x, y \in A$ und alle $\lambda \in \mathbb{R}$ gilt

$$\lambda x + (1-\lambda)y \in A.$$

Ist $x_0 \in A$, so ist

$$L = \{x - x_0 : x \in A\}$$

ein Untervektorraum von $\mathbb{R}^n$. Seine Dimension heißt die *Dimension* von A,

$$\dim A = \dim L.$$

Ist $A = \emptyset$ die leere Menge, setzt man $\dim A = -1$.

BEISPIEL I.2.7. Hyperebenen sind affin. Für

$$H = \{x \in \mathbb{R}^n : a^t x = b\}$$

gilt

$$\dim H = \begin{cases} n-1 & \text{falls } a \neq 0, \\ n & \text{falls } a = 0 \text{ und } b = 0, \\ -1 & \text{falls } a = 0 \text{ und } b \neq 0. \end{cases}$$

DEFINITION I.2.8. Für eine beliebige Menge $S \subset \mathbb{R}^n$ bezeichnet $\operatorname{aff}(S)$ die kleinste affine Menge, die S enthält:

$$\operatorname{aff}(S) = \bigcap_{\substack{C \supset S \\ C \text{affin}}} C.$$

Die Dimension von $\operatorname{aff}(S)$ nennt man die *(affine) Dimension* von S.

BEMERKUNG I.2.9. Ist $k = \dim \operatorname{aff}(S)$, so gibt es $k+1$ Punkte x_0, …, $x_k \in \operatorname{aff}(S)$ mit

$$\operatorname{aff}(S) = \left\{ \sum_{i=0}^{k} \lambda_i x_i : \lambda_i \in \mathbb{R} \ , \ \sum_{i=0}^{k} \lambda_i = 1 \right\}.$$

Man kann die Punkte sogar in S wählen. Denn ausgehend von einem beliebigen Punkt $x_0 \in S$ kann man durch Hinzunehmen weiterer Punkte $x_1, \ldots, x_k$ die Menge $\operatorname{aff}(S)$ sukzessive aufbauen:

$$\operatorname{aff}(\{x_0\}) \subset \operatorname{aff}(\{x_0, x_1\}) \subset \ldots \subset \operatorname{aff}(\{x_0, \ldots, x_k\}) = \operatorname{aff}(S).$$

BEISPIEL I.2.10. Für

$$P = \{x \in \mathbb{R}^3 : x_1 + x_2 + x_3 \leq 1 \ , \ x_1 \geq 0 \ , \ x_2 \geq 0 \ , \ x_3 \geq 0\}$$

erhalten wir

$$\operatorname{aff}(P) = \operatorname{aff}\left(\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \ , \ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \ , \ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \ , \ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \right)$$

und damit

$$\dim \operatorname{aff}(P) = 3.$$

BEMERKUNG I.2.11. Die zulässigen Mengen von LPs in Standardform sind spezielle Polyeder. Sie sind charakterisiert durch Gleichungen und besonders einfache Ungleichungen. Bemerkung I.1.6 (S. 14) zeigt andererseits, dass jeder Polyeder in dieser speziellen Form dargestellt werden kann. Wir verdeutlichen dies noch einmal für das spezielle Beispiel des Einheitsquadrates

$$Q = \{x \in \mathbb{R}^2 : 0 \leq x_i \leq 1 \ , \ i = 1, 2\}.$$

Analog zu Bemerkung I.1.6 setzen wir

$$y_1 = x_1, \quad y_2 = x_2, \quad y_3 = 1 - x_1, \quad y_4 = 1 - x_2.$$

Dann gilt

$$y_1 + y_3 = 1, \quad y_2 + y_4 = 1, \quad y_i \geq 0 \ i = 1, \ldots, 4.$$

D.h. das Einheitsquadrat ist die Projektion auf die (y_1, y_2)-Ebene des Polyeders

$$P = \{y \in \mathbb{R}^4 : Ay = b \ , \ y \geq 0\}$$

mit

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

Für die praktische Lösung eines LPs spielen die „*Ecken*“ des Zulässigkeitsbereiches eine wesentliche Rolle. Wir benötigen eine saubere mathematische Definition dieses intuitiven Begriffes. Dies leistet die folgende Definition. Sie besagt anschaulich, dass eine Ecke nicht als Konvexkombination verschiedener Punkte des Polyeders darstellbar ist.

DEFINITION I.2.12. Ein Punkt a einer konvexen Menge M heißt *Extremalpunkt* von M, falls mit $x, y \in M$ und $\lambda \in (0, 1)$ aus

$$a = \lambda x + (1 - \lambda)y$$

stets

$$a = x = y$$

folgt.

BEISPIEL I.2.13. M sei die abgeschlossene Einheitskugel in $\mathbb{R}^n$. Dann sind die Randpunkte von M genau die Extremalpunkte von M.

Man überzeugt sich leicht, dass die Ecken von Polyedern in $\mathbb{R}^2$ oder $\mathbb{R}^3$ genau die Extremalpunkte sind. Eine andere intuitive Definition von „*Ecken*“ ist diejenige als 0-dimensionaler Durchschnitt $m - 1$-dimensionaler „*Seitenflächen*“, wenn m die affine Dimension des Polyeders ist. Hierzu benötigt man aber eine Präzisierung des Begriffes „*Seitenfläche*“. Die leistet die folgende Definition.

DEFINITION I.2.14. Eine konvexe Teilmenge E einer konvexen Menge M heißt *Extremalmenge* von M, falls aus $a \in E$, $x, y \in M$, $\lambda \in (0, 1)$ und

$$a = \lambda x + (1 - \lambda)y$$

stets folgt

$$x, y \in E.$$

BEMERKUNG I.2.15. Extremalpunkte sind 0-dimensionale Extremalmengen. M selbst ist Extremalmenge von M maximaler Dimension. Die Kanten eines Polyeders in $\mathbb{R}^2$ oder $\mathbb{R}^3$ sind 1-dimensionale Extremalmengen. Die Seitenflächen eines Polyeders in $\mathbb{R}^3$ sind 2-dimensionale Extremalmengen.

SATZ I.2.16. *E sei Extremalmenge der konvexen Menge M. Dann gilt*

$$E = M \cap \operatorname{aff}(E).$$

BEWEIS. Offensichtlich ist

$$E \subset M \cap \operatorname{aff}(E).$$

Sei nun umgekehrt $x \in M \cap \operatorname{aff}(E)$. Gemäß Bemerkung I.2.9 gibt es dann Punkte $y_0, \ldots, y_k \in E$ und Gewichte $\omega_0, \ldots, \omega_k \in \mathbb{R}$ mit

$$x = \sum_{i=0}^{k} \omega_i y_i, \quad \sum_{i=0}^{k} \omega_i = 1.$$

Falls alle $\omega_i \geq 0$ sind, folgt $x \in E$ wegen der Konvexität von E. Seien also einige Gewichte negativ. Definiere

$$\alpha = -\sum_{\substack{0 \leq i \leq k \\ \omega_i < 0}} \omega_i, \quad z_1 = \frac{1}{1+\alpha} \sum_{\substack{0 \leq i \leq k \\ \omega_i \geq 0}} \omega_i y_i, \quad z_2 = -\frac{1}{\alpha} \sum_{\substack{0 \leq i \leq k \\ \omega_i < 0}} \omega_i y_i.$$

Dann ist

$$\sum_{\substack{0 \leq i \leq k \\ \omega_i \geq 0}} \omega_i = 1 + \alpha > 0$$

und daher $z_1 \in E$. Analog folgt $z_2 \in E$. Damit haben wir

$$x \in M, \quad z_2 \in E, \quad z_1 = \frac{1}{1+\alpha} x + \frac{\alpha}{1+\alpha} z_2 \in E.$$

Da E Extremalmenge ist, folgt $x \in E$ und damit $M \cap \operatorname{aff}(E) \subset E$. □

SATZ I.2.17. *Wenn ein allgemeines LP überhaupt Optimallösungen besitzt, ist die Menge E der Optimallösungen eine Extremalmenge des Polyeders $\mathcal{P}$ der zulässigen Lösungen von LP.*

BEWEIS. Das allgemeine LP besitzt die Form

$$\min\{c^t x : x \in \mathcal{P}\}$$

mit dem Polyeder $\mathcal{P}$ der zulässigen Lösungen. Falls es Optimallösungen gibt, ist

$$\alpha = \min\{c^t x : x \in \mathcal{P}\}$$

endlich und es ist

$$E = \{x \in \mathcal{P} : c^t x = \alpha\}$$

ein Polyeder in $\mathcal{P}$, also konvex.
Wir nehmen an, E sei keine Extremalmenge. Dann gibt es ein $a \in E$, $x, y \in \mathcal{P}$ und $\lambda \in (0,1)$ mit

$$a = \lambda x + (1-\lambda)y$$

und x und y sind nicht beide in E enthalten. O.E. sei $x \notin E$. Dann ist

$$c^t x > \alpha \quad \text{und} \quad c^t y \geq \alpha.$$

Damit folgt

$$\alpha = c^t a = \lambda c^t x + (1-\lambda) c^t y > \lambda\alpha + (1-\lambda)\alpha = \alpha.$$

Dies ist ein Widerspruch. □

Jede konvexe Menge M besitzt M und die leere Menge $\emptyset$ als Extremalmengen. Die Existenz weiterer Extremalmengen, insbesondere von Extremalpunkten dagegen ist nicht gesichert. Insbesondere besitzt kein Halbraum

$$H = \{x \in \mathbb{R}^n : a^t x \leq b\}$$

mit $n \geq 2$ und $a \neq 0$ Extremalpunkte. Anders sieht es aus für Polyeder, die keine Geraden enthalten.

Satz I.2.18. *Der Polyeder $\mathcal{P}$ sei nicht leer und enthalte keine Gerade. Dann enthält $\mathcal{P}$ mindestens einen Extremalpunkt.*

Beweis. Sei $x \in \mathcal{P}$. Falls x kein Extremalpunkt ist, gibt es ein $h \in \mathbb{R}^n \backslash \{0\}$ mit $x \pm h \in \mathcal{P}$. Da $\mathcal{P}$ keine Gerade enthält, gibt es ein $\lambda \in \mathbb{R}^*$, so dass $x + \lambda h = x'$ ein Randpunkt von $\mathcal{P}$ ist. Falls x' kein Extremalpunkt ist, können wir diesen Prozess mit einem von h linear unabhängigen Vektor h' wiederholen. Spätestens nach n Schritten haben wir dann einen Extremalpunkt gefunden. □

Satz I.2.19. *Das LP in Standardform besitze Optimallösungen. Dann gibt es unter ihnen auch Extremalpunkte des Zulässigkeitsbereiches $\mathcal{P}$.*

Beweis. Nach Voraussetzung ist $\mathcal{P} \neq \emptyset$. Wegen

$$\mathcal{P} \subset \{x \in \mathbb{R}^n : x \geq 0\}$$

enthält $\mathcal{P}$ keine Gerade. Also besitzt $\mathcal{P}$ gemäß Satz I.2.18 mindestens einen Extremalpunkt. Damit folgt die Behauptung aus Satz I.2.17. □

I.3. Algebraische Grundlagen

Im Folgenden ist stets

$$m \leq n, \quad A \in \mathbb{R}^{m \times n}, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^m, \quad N = \{1, \ldots, n\}. \tag{I.3.1}$$

DEFINITION I.3.1. (1) Sei $J \subset N$ nicht leer. Dann besteht A_J aus den Spaltenvektoren von A, die zu der Indexmenge J gehören; x_J besteht aus den Komponenten von x, die zu der Indexmenge J gehören. (2) Zwei nicht leere Teilmengen J und K von N heißen *komplementär*, wenn gilt

$$J \cap K = \emptyset \quad \text{und} \quad J \cup K = N.$$

In diesem Fall schreiben wir

$$N = J \oplus K.$$

(3) $J \subset N$ heißt eine *Basis*, wenn $\sharp J = m$ und A_J regulär ist. In diesem Fall heißt x_J ein *Basisvektor*. Ist J eine Basis und $N = J \oplus K$, so heißt K eine zugehörige *Nichtbasis* bzw. ein zugehöriges *Komplement*.

Es gelte

$$\operatorname{rang} A = m.$$

Dann ist das LGS $Ax = b$ lösbar. Daher gibt es eine Basis J. Für diese gilt mit $N = J \oplus K$

$$\begin{aligned} & b = A_J x_J + A_K x_K \\ \Longrightarrow \quad & A_J^{-1} b = x_J + A_J^{-1} A_K x_K \\ \Longrightarrow \quad & x_J = A_J^{-1} b - A_J^{-1} A_K x_K. \end{aligned}$$

DEFINITION I.3.2. Es gelte $\operatorname{rang} A = m$ und J sei eine Basis und $N = J \oplus K$. Der Vektor x mit

$$x_K = 0 \quad \text{und} \quad x_J = A_J^{-1} b$$

heißt eine *Basislösung* (zu J) und wird mit $x(J)$ bezeichnet.

DEFINITION I.3.3. Ein Paar $(J; (\overline{A}, \overline{b}))$ mit $J \subset N$, $\overline{A} \in \mathbb{R}^{m \times n}$ und $\overline{b} \in \mathbb{R}^m$ heißt ein *Tableau*, wenn gilt

$$\overline{A}_J = I.$$

Es heißt dem LGS $Ax = b$ *zugeordnet*, wenn die LGS $Ax = b$ und $\overline{A}x = \overline{b}$ dieselbe Lösungsmenge haben.

DEFINITION I.3.4. Das LGS $Ax = b$ gehöre zu einem LP in Standardform mit Zulässigkeitsbereich $\mathcal{P}$. Eine Basis J von A heißt dann *zulässig*, wenn die zughörige Basislösung $x(J)$ zulässig ist, d.h. in $\mathcal{P}$ liegt.

BEMERKUNG I.3.5. Man beachte, dass für $N = J \oplus K$ stets gilt

$$x(J)_K \geq 0.$$

Die Zulässigkeit von $x(J)$ ist also eine Bedingung an die Komponenten von $x(J)_J$.

Beispiel I.3.6. Betrachte

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

Da

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

regulär ist, ist

$$J = \{1, 4\}$$

eine Basis. Die entsprechende Nichtbasis ist

$$K = \{2, 3\}.$$

Um das entsprechende Tableau zu erhalten, müssen wir $A_J^{-1}A$ und $A_J^{-1}b$ berechnen:

$$A_J^{-1} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$$

$$A_J^{-1}A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 & 2 \\ -1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix}$$

$$A_J^{-1}b = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ -1 \end{pmatrix}.$$

Also lautet das entsprechende Tableau

$$(J; (\overline{A}, \overline{b})) = \left(\{1, 4\}; \left(\begin{array}{cccc|c} 1 & -2 & 1 & 0 & -3 \\ 0 & -1 & 1 & 1 & -1 \end{array} \right) \right).$$

Die zugehörige Basislösung ist

$$x(J) = \begin{pmatrix} -3 \\ 0 \\ 0 \\ -1 \end{pmatrix}.$$

Sie ist nicht zulässig. Dementsprechend ist J nicht zulässig.
Da

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

regulär ist, ist $J' = \{2, 3\}$ auch eine Basis. Die entsprechende Nichtbasis ist $K' = \{1, 4\}$. Für die zugehörige Basislösung erhalten wir

$$x_{J'} = A_{J'}^{-1}b = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

und damit

$$x(J') = \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}.$$

Diese ist zulässig. Dementsprechend ist J' zulässig.

Die folgende Definition und die nachfolgenden Sätze beschreiben den Zusammenhang zwischen der Standardform und der Simplexform eines LP und den zugehörigen Ecken.

DEFINITION I.3.7. Für ein LP in Standardform (I.1.2) (S. 14) und das zugehörige LP in Simplexform (I.1.3) (S. 15) setzen wir

$$\begin{aligned}
x_{n+1} &= z,\\
\widehat{N} &= N \cup \{n+1\} = \{1,\dots,n+1\},\\
\widehat{A} &= \begin{pmatrix} A & 0\\ c^t & 1\end{pmatrix},\\
\widehat{b} &= \begin{pmatrix} b\\ 0\end{pmatrix},\\
\widehat{x} &= \begin{pmatrix} x\\ z\end{pmatrix} = \begin{pmatrix} x\\ x_{n+1}\end{pmatrix},\\
\mathcal{P} &= \{x \in \mathbb{R}^n : Ax = b\ ,\ x \geq 0\},\\
\widehat{\mathcal{P}} &= \{\widehat{x} \in \mathbb{R}^{n+1} : \widehat{A}\widehat{x} = \widehat{b}\ ,\ \widehat{x}_N \geq 0\}.
\end{aligned}$$

SATZ I.3.8. *(1) J ist eine Basis von A genau dann, wenn*

$$\widehat{J} = J \oplus \{n+1\}$$

eine Basis von $\widehat{A}$ ist.
(2) $x(J)$ ist eine Basislösung von $Ax = b$ genau dann, wenn

$$\widehat{x}(\widehat{J}) = \begin{pmatrix} x(J)\\ -c^t x(J)\end{pmatrix}$$

eine Basislösung von $\widehat{A}\widehat{x} = \widehat{b}$ ist.

BEWEIS. *ad (1)*: Folgt aus

$$\det \widehat{A}_{\widehat{J}} = \det \begin{pmatrix} A_J & 0\\ c_J^t & 1\end{pmatrix} = \det A_J.$$

ad (2): Folgt aus

$$\widehat{A}_{\widehat{J}}\widehat{x}_{\widehat{J}} = \begin{pmatrix} A_J & 0\\ c_J^t & 1\end{pmatrix}\begin{pmatrix} x_J\\ -c^t x(J)\end{pmatrix} = \begin{pmatrix} A_J x_J\\ c_J^t x_J - c^t x(J)\end{pmatrix}$$

und

$$c_J^t x_J = c^t x(J).$$

□

SATZ I.3.9. *(1) Der Vektor x^* ist genau dann eine Ecke von $\mathcal{P}$, wenn der Vektor*

$$\widehat{x}^* = \begin{pmatrix} x^*\\ -c^t x^*\end{pmatrix}$$

eine Ecke von $\widehat{\mathcal{P}}$ ist.
(2) Der Vektor x^* ist genau dann eine Ecke von $\mathcal{P}$, wenn es eine zulässige Basis J von A gibt mit $x^* = x(J)$.

BEWEIS. *ad (1)*: Seien $x, y, z \in \mathcal{P}$ und

$$\widehat{x} = \begin{pmatrix} x \\ -c^t x \end{pmatrix}, \quad \widehat{y} = \begin{pmatrix} y \\ -c^t y \end{pmatrix}, \quad \widehat{z} = \begin{pmatrix} z \\ -c^t z \end{pmatrix}.$$

Dann gilt

$$\begin{aligned} & x = \lambda y + (1-\lambda) z \quad \text{mit } \lambda \in [0,1] \\ \iff \quad & \widehat{x} = \lambda \widehat{y} + (1-\lambda)\widehat{z}. \end{aligned}$$

Hieraus folgt die Behauptung.
ad (2): „$\Longrightarrow$“: Sei x^* eine Ecke von $\mathcal{P}$. Setze

$$S = S(x^*) = \{\ell \in N : x^*_\ell > 0\}.$$

O.E. ist $S \neq \emptyset$, da sonst $x^* = 0$ und damit $b = 0$ ist und somit die Behauptung offensichtlich gilt. Wir behaupten, dass die Spalten von A_S linear unabhängig sind. Denn andernfalls gäbe einen Vektor $\lambda \in \mathbb{R}^{\sharp S}$ mit $\lambda \neq 0$ und

$$0 = \sum_{i \in S} \lambda_i A_i.$$

Definiere $z \in \mathbb{R}^n$ durch

$$z_i = \begin{cases} \lambda_i & \text{falls } i \in S, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt

$$z \neq 0 \quad \text{und} \quad Az = 0.$$

Daher gilt für alle $\varepsilon \in \mathbb{R}$

$$A(x^* \pm \varepsilon z) = b.$$

Für

$$0 \leq \varepsilon < \frac{\min_{i \in S} x^*_i}{\max_{i \in S} z_i}$$

ist offensichtlich

$$x \pm \varepsilon z \geq 0.$$

Daher ist für mindestens ein $\varepsilon > 0$

$$x^* = \frac{1}{2}\Big[(x^* + \varepsilon z) + (x^* - \varepsilon z)\Big] \quad \text{und} \quad x^* \pm \varepsilon z \in \mathcal{P}.$$

Dies ist ein Widerspruch zur Extremaleigenschaft von x^*.
Da die Spalten von A_S linear unabhängig sind, ist S eine Basis oder kann durch Hinzunahme weiterer Indizes zu einer Basis J ergänzt werden. Nach Konstruktion von S ist x^* eine Basislösung zu dieser Basis J.
„$\Longleftarrow$“: Sei nun J eine zulässige Basis von A und $x(J)$ eine zugehörige

Basislösung. Dann gilt $x(J) \geq 0$ und $x(J)_K = 0$ mit $N = J \oplus K$. Also gilt für

$$S = S(x(J)) = \{\ell \in N : x(J)_\ell > 0\}$$

die Inklusion $S \subset J$.
Wir nehmen nun an, dass

$$x(J) = \lambda y + (1 - \lambda)z$$

ist mit $y, z \in \mathcal{P}$ und $\lambda \in (0, 1)$ und zeigen $x = y = z$.
Für $i \notin S$ ist $x(J)_i = 0$. Wegen $y \geq 0$, $z \geq 0$ und $\lambda \in (0, 1)$ folgt $y_i = z_i = 0$.
Für $i \in S$ erhalten wir wegen $y, z \in \mathcal{P}$

$$0 = b - b = Ay - Az = A(y - z) = \sum_{i \in S} A_i(y_i - z_i).$$

Da $S \subset J$ ist, sind die Spalten von A_S linear unabhängig. Daher folgt aus obiger Gleichung

$$y_i = z_i \quad \forall i \in S.$$

Dies beweist $x = y = z$ und damit die Extremaleigenschaft von x. □

Beispiel I.3.10. Betrachte

$$\mathcal{P} = \{x \in \mathbb{R}^3 : Ax = b\ ,\ x \geq 0\}$$

mit

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

Dann ist

$$\mathcal{P} = \left\{ \begin{pmatrix} t \\ t \\ 1 \end{pmatrix} : t \geq 0 \right\}.$$

Die einzige Ecke ist

$$x^* = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

$J = \{1, 3\}$ ist eine Basis mit

$$A_J = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad x(J) = x^*.$$

$J' = \{2, 3\}$ ist ebenfalls eine Basis mit

$$A_{J'} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad x(J') = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x^*.$$

Dieses Beispiel zeigt, dass die Zuordnung „Ecke - Basis“ nicht eindeutig ist. Der Grund für die Nicht-Eindeutigkeit liegt daran, dass es zu viele aktive Ungleichungen gibt, d.h. für die Basen J und J' und die zugehörigen Basislösungen gibt es einen Index $j \in J$ bzw. $j' \in J'$ mit $x(J)_j = 0$ bzw. $x(J')_{j'} = 0$.
Gilt für eine Basis J dagegen $x(J)_J > 0$, so ist die Zuordnung „Ecke - Basis“ eindeutig. Dies führt auf folgende Definition.

DEFINITION I.3.11. Eine Basis J heißt *nicht entartet*, wenn für die zugehörige Basislösung gilt $x(J)_J > 0$.

SATZ I.3.12. *Sei J eine Basis zu A. Dann lautet das zugehörige Tableau $(J; (\overline{A}, \overline{b}))$ mit $\overline{A} = A_J^{-1}A$ und $\overline{b} = A_J^{-1}b$. Dann ist $\widehat{J} = J \oplus \{n+1\}$ eine Basis zu $\widehat{A}$. Das zugehörige Tableau lautet $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$ mit*

$$\overline{c}^t = -\pi A + c^t, \quad \beta = -\pi b, \quad \pi = c_J^t A_J^{-1}.$$

Die Komponenten von π nennt man Schattenpreise.

BEWEIS. Wir müssen nur noch die Aussagen zu $\widehat{J}$ beweisen. Es ist

$$\widehat{b} = \begin{pmatrix} b \\ 0 \end{pmatrix}, \quad \widehat{A} = \begin{pmatrix} A & 0 \\ c^t & 1 \end{pmatrix}, \quad \widehat{A}_{\widehat{J}} = \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}.$$

Daher ist $\widehat{A}_{\widehat{J}}^{-1}$ von der Form

$$\widehat{A}_{\widehat{J}}^{-1} = \begin{pmatrix} A_J^{-1} & 0 \\ -\pi & 1 \end{pmatrix}.$$

Aus

$$\begin{pmatrix} I & 0 \\ 0 & 1 \end{pmatrix} \overset{!}{=} \widehat{A}_{\widehat{J}}^{-1}\widehat{A}_{\widehat{J}} = \begin{pmatrix} A_J^{-1} & 0 \\ -\pi & 1 \end{pmatrix} \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\pi A_J + c_J^t & 1 \end{pmatrix}$$

folgt

$$\pi A_J = c_J^t \quad \Longleftrightarrow \quad \pi = c_J^t A_J^{-1}.$$

Damit erhalten wir

$$\widehat{A}_{\widehat{J}}^{-1}(\widehat{A}, \widehat{b}) = \begin{pmatrix} A_J^{-1} & 0 \\ -\pi & 1 \end{pmatrix} \begin{pmatrix} A & 0 & b \\ c^t & 1 & 0 \end{pmatrix} = \begin{pmatrix} I & 0 & A_J^{-1}b \\ -\pi A + c^t & 1 & -\pi b \end{pmatrix}.$$

Dies beweist die Behauptung. □

I.4. Der Simplexalgorithmus

Falls ein LP in Standardform überhaupt eine Lösung besitzt, wird gemäß Satz I.2.19 (S. 20) das Optimum an einer Ecke des zulässigen Bereiches angenommen. Daher kann man das Optimum durch systematisches Durchsuchen der endlich vielen Ecken finden. Auf dieser Beobachtung fußt der *Simplexalgorithmus*, der sich wie folgt beschreiben lässt:

- Finde eine Ecke von $\mathcal{P}$.

- Ausgehend von einer Ecke von $\mathcal{P}$ finde eine „Nachbarecke“ mit einem kleineren Wert der Zielfunktion. Falls keine solche Ecke existiert, ist das Optimum gefunden. Andernfalls setze das Verfahren mit der neuen Ecke fort.

Für die Durchführung des Verfahrens ersetzen wir wegen §I.3 den Begriff „Ecke“ durch den Begriff „Basis“. Dadurch ergibt sich allerdings die Schwierigkeit, dass bei entarteten Basen mehrere Basen zur gleichen Ecke gehören können. Eine weitere Schwierigkeit ist natürlich das Auffinden einer ersten Ecke bzw. Basis.

Wir gehen nun wie folgt vor:

- Zuerst beschreiben wir den Algorithmus unter der Annahme, dass eine erste Ecke bekannt ist und dass keine Entartung auftritt.
- Dann beschreiben wir das geeignete Vorgehen im Falle einer Entartung.
- Schließlich zeigen wir, wie eine erste Ecke bestimmt werden kann.

DEFINITION I.4.1. Zwei Basen J und J' zu dem LGS $Ax = b$ heißen *benachbart*, wenn sie sich um genau einen Index unterscheiden, d.h. $\sharp(J \cap J') = \sharp J - 1$. Zwei Basen $\widehat{J} = J \oplus \{n+1\}$ und $\widehat{J'} = J' \oplus \{n+1\}$ zu dem erweiterten Problem $\widehat{A}\widehat{x} = \widehat{b}$ heißen *benachbart*, wenn J und J' benachbart sind.

SATZ I.4.2. *Es sei $J = (i_1, \dots, i_m)$ eine Basis zu dem LGS $Ax = b$ und $s \notin J$. Weiter sei $\widehat{J} = J \oplus \{n+1\}$ und $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$ das zu $\widehat{J}$ und dem erweiterten System*

$$\widehat{A}\widehat{x} = \begin{pmatrix} A & 0 \\ c^t & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} = \widehat{b}$$

gehörige Tableau. Ferner sei

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \\ \alpha_{m+1} \end{pmatrix} = \begin{pmatrix} \overline{A}_s \\ \overline{c}_s \end{pmatrix}$$

die s-te Spalte von $\left(\begin{smallmatrix} \overline{A} \\ \overline{c}^t \end{smallmatrix}\right)$. Dann ist $J' = (i_1, \dots, i_{r-1}, s, i_{r+1}, \dots, i_m)$ genau dann eine Nachbarbasis von J, wenn $\alpha_r \neq 0$ ist. In diesem Fall ist $\widehat{J'} = J' \oplus \{n+1\}$ eine Nachbarbasis zu $\widehat{J}$. Das zu $\widehat{J'}$ gehörige Tableau $\left(\widehat{J'}; \left(\begin{smallmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{smallmatrix}\right)\right)$ ist gegeben durch

$$\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} = F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}$$

mit

$$F = \begin{pmatrix} 1 & & \overset{i_r\text{-te Spalte}}{-\frac{\alpha_1}{\alpha_r}} & & & 0 \\ & \ddots & \vdots & & & \vdots \\ & & -\frac{\alpha_{r-1}}{\alpha_r} & & & \vdots \\ & 1 & \frac{1}{\alpha_r} & & & \vdots \\ & & -\frac{\alpha_{r+1}}{\alpha_r} & 1 & & \vdots \\ & & \vdots & & \ddots & \vdots \\ & & -\frac{\alpha_m}{\alpha_r} & & 1 & 0 \\ & & -\frac{\alpha_{m+1}}{\alpha_r} & & & 1 \end{pmatrix}.$$

Es ist

$$F^{-1} = G = \begin{pmatrix} 1 & & \alpha_1 & & & 0 \\ & \ddots & \vdots & & & \vdots \\ & & \alpha_{r-1} & & & \vdots \\ & 1 & \alpha_r & & & \vdots \\ & & \alpha_{r+1} & 1 & & \vdots \\ & & \vdots & & \ddots & \vdots \\ & & \alpha_m & & 1 & 0 \\ & & \alpha_{m+1} & & & 1 \end{pmatrix}.$$

Beweis. Wie man leicht nachrechnet, gilt im Fall $\alpha_r \neq 0$

$$FG = GF = I.$$

Da J und J' sich in genau einem Index unterscheiden und da konstruktionsgemäß

$$\begin{pmatrix} \overline{A}_J & 0 \\ \overline{c}_J^t & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & 1 \end{pmatrix}$$

ist, gilt

$$G = \begin{pmatrix} \overline{A}_{J'} & 0 \\ \overline{c}_{J'}^t & 1 \end{pmatrix}.$$

Die zu J und J' gehörigen Tableaumatrizen sind definiert durch

$$\begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix} = \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A & 0 & b \\ c^t & 1 & 0 \end{pmatrix}$$

und

$$\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} = \begin{pmatrix} A_{J'} & 0 \\ c_{J'}^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A & 0 & b \\ c^t & 1 & 0 \end{pmatrix}.$$

Aus der ersten Gleichung folgt

$$\begin{pmatrix} \overline{A}_{J'} & 0 \\ (\overline{c}_{J'})^t & 1 \end{pmatrix} = \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A_{J'} & 0 \\ c_{J'}^t & 1 \end{pmatrix}$$

und damit

$$\begin{pmatrix} A_{J'} & 0 \\ c_{J'}^t & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \overline{A}_{J'} & 0 \\ (\overline{c}_{J'})^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1}.$$

Einsetzen in die zweite Gleichung liefert

$$\begin{aligned} \begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} &= \begin{pmatrix} \overline{A}_{J'} & 0 \\ (\overline{c}_{J'})^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A & 0 & b \\ c^t & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \overline{A}_{J'} & 0 \\ (\overline{c}_{J'})^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix} \\ &= F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}. \end{aligned}$$

Dies beweist die Behauptung. □

BEMERKUNG I.4.3. Die Berechnung von F erfordert 1 Division und m Multiplikationen. Wegen

$$\begin{pmatrix} \overline{A}'_{J'} & 0 \\ (\overline{c}')^t_{J'} & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & 1 \end{pmatrix} = I$$

müssen bei der Multiplikation $F \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix} \right)$ die zu $\widehat{J}' = J' \oplus \{n+1\}$ gehörenden Spalten nicht berechnet werden. Die Berechnung der restlichen $n+1-m$ Spalten erfordert $(m+1)(n+1-m)$ Additionen und Multiplikationen.

SATZ I.4.4. *Sei J eine zulässige Basis, K das zugehörige Komplement und $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix} \right)\right)$ das zugehörige Tableau.*

(1) *Es sei $\overline{c}_K \geq 0$. Dann ist die Basislösung $x^* = x(J)$ eine Lösung des LP in Standardform.*
(2) *Es gebe ein $s \in K$ mit $\overline{c}_s < 0$. Für die s-te Spalte $\overline{A}_s$ von $\overline{A}$ gelte $\overline{A}_s \leq 0$. Dann ist der zulässige Bereich des LP unbeschränkt und das LP besitzt keine Lösung.*
(3) *Es gebe ein $s \in K$ mit $\overline{c}_s < 0$ und ein $r \in \{1, \ldots, m\}$ mit $\overline{A}_{rs} > 0$. Dann führt Satz I.4.2 auf eine benachbarte Basis $J' = (J \backslash \{i_r\}) \oplus \{s\}$. Für die zugehörige Basislösung $x^{**} = x(J')$ gilt $c^t x^{**} \leq c^t x^*$. Dabei gilt Gleichheit genau dann, wenn J entartet ist.*

BEWEIS. *ad (1)*: Wegen $A_J = I$ und $\overline{c}_J = 0$ reduzieren sich die Gleichungen des Tableaus auf

$$x_J + \overline{A}_K x_K = \overline{b}, \quad z + \overline{c}_K^t x_K = \beta. \tag{I.4.1}$$

Wegen $\overline{c}_K \geq 0$ gilt daher für jede zulässige Lösung (x, z) des erweiterten Problems

$$z = \beta - \overline{c}_K^t x_K \leq \beta.$$

Wegen $x^*_K = 0$ gilt andererseits für das zugehörige z^*

$$z^* = \beta.$$

Also ist x^* eine Lösung des LP.
ad (2): Sei nun $s \in K$ so, dass $\overline{c}_s < 0$ ist. Bezeichne mit $\overline{a}$ die s-te Spalte von $\overline{A}$. Aus (I.4.1) erkennen wir, dass sich wegen $\overline{c}_s < 0$ der Wert z der Zielfunktion vergrößert, wenn man die Komponenten x_k mit $k \in K\backslash\{s\}$ bei Null behält, x_s vergrößert und die Komponenten x_J so bestimmt, dass die restlichen Tableaugleichungen erfüllt bleiben. Für $\theta \geq 0$ setzen wir daher

$$\begin{aligned} x_s(\theta) &= \theta, & x_{K\backslash\{s\}}(\theta) &= 0, \\ x_J(\theta) &= \overline{b} - \theta\overline{a}, & z(\theta) &= \beta - \overline{c}_s\theta. \end{aligned}$$

Mit dieser Wahl sind die Gleichungen (I.4.1) für alle θ erfüllt. Da nach Voraussetzung J zulässig ist und $x^*_J = \overline{b}$ ist, gilt $\overline{b} \geq 0$. Da die Komponenten von $\overline{a}$ nach Voraussetzung nicht positiv sind, gilt für alle $\theta \geq 0$

$$x_J(\theta) \geq \overline{b} \geq 0.$$

Also ist $x(\theta)$ für alle $\theta \geq 0$ zulässig. Daher ist der zulässige Bereich unbeschränkt und $z(\theta)$ wächst unbeschränkt. Daher besitzt das LP keine Lösung.
ad (3): Wir gehen wie in Teil (2) vor. Dann gibt es ein maximales θ^*, so dass $x(\theta^*)$ zulässig ist. Diese θ^* ist bestimmt durch

$$\begin{aligned} \theta^* &= \max\left\{\theta : \overline{b} - \theta\overline{a} \geq 0\right\} \\ &= \max\left\{\theta : \overline{b}_j - \theta\overline{a}_j \geq 0 \text{ für alle } j \text{ mit } \overline{a}_j > 0\right\} \\ &= \min\left\{\frac{\overline{b}_j}{\overline{a}_j} : \overline{a}_j > 0\right\}. \end{aligned}$$

Wir wählen nun r in Satz I.4.2 so, dass dieses Minimum angenommen wird. Dann gilt

$$x_r(\theta^*) = 0, \quad x_s(\theta^*) \geq 0, \quad x_{K\backslash\{s\}}(\theta^*) = 0.$$

Dabei ist $x_s(\theta^*) > 0$, falls J nicht entartet ist. Daher ist $x(\theta^*)$ eine Basislösung zu $J' = (J\backslash\{i_r\}) \oplus \{s\}$ und für das zugehörige $z(\theta^*)$ gilt

$$z(\theta^*) \geq \beta \quad \text{d.h.} \quad c^t x(\theta^*) \leq c^t x^*.$$

□

Die Sätze I.4.2 und I.4.4 führen auf Algorithmus I.4.1.

Beispiel I.4.5. Wir betrachten das LP aus Beispiel 1 (S. 5)

$$16x + 32y \to \max$$

Algorithmus I.4.1 Simplexschritt

Gegeben: zulässige Basis J und Komplement K
Gesucht: neue Basis J' und zugehöriges Tableau

1: $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right) \leftarrow$ Tableau zu $\widehat{J} = J \oplus \{n+1\}$
2: $\overline{x}_J \leftarrow \overline{b}$, $\overline{x}_K \leftarrow 0$, $\overline{z} \leftarrow \beta$
3: **if** $\overline{c}_K \geq 0$ **then**
4: **stop** ▷ $\overline{x}$ ist Optimallösung
5: **end if**
6: $s \leftarrow$ Index in K mit $\overline{c}_s < 0$
7: $\overline{a} \leftarrow \overline{A}_s$ ▷ s-te Spalte von $\overline{A}$
8: **if** $\overline{a} \leq 0$ **then**
9: **stop** ▷ LP besitzt keine Lösung
10: **end if**
11: $r \leftarrow \operatorname{argmin}_j \left\{ \frac{\overline{b}_j}{\overline{a}_j} : j \in \{1, \ldots, m\}, \overline{a}_j > 0 \right\}$
12: $J' \leftarrow (J \backslash \{i_r\}) \oplus \{s\} = \{i_1, \ldots, i_{r-1}, s, i_{r+1}, \ldots, i_m\}$
13: $\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} \leftarrow F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}$ ▷ neues Tableau

unter den Nebenbedingungen

$$\begin{aligned} 6x + 15y &\leq 4500, \\ 4x + 5y &\leq 2000, \\ 20x + 10y &\leq 8000. \end{aligned}$$

Zunächst müssen wir dieses LP in die Standardform bringen. Dazu schreiben wir x_1 statt x und x_2 statt y und führen Schlupfvariable x_3, x_4, x_5 ein. Dann lauten die Nebenbedingungen

$$Ax = b$$

mit

$$A = \begin{pmatrix} 6 & 15 & 1 & 0 & 0 \\ 4 & 5 & 0 & 1 & 0 \\ 20 & 10 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 4500 \\ 2000 \\ 8000 \end{pmatrix}.$$

Unter diesen Nebenbedingungen ist $-16x_1 - 32x_2$ zu minimieren, d.h.

$$c = \begin{pmatrix} -16 \\ -32 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Wegen der Einführung der Schlupfvariablen ist $(0, 0, 4500, 2000, 8000)^t$ eine zulässige Ecke, d.h. $J = \{3, 4, 5\}$ ist eine zulässige Basis. Es ist

$$A_J = I, \quad c_J = 0.$$

Aus Satz I.3.12 (S. 26) erhalten wir das zu $\widehat{J} = \{3,4,5,6\}$ gehörige Tableau:

$$\begin{aligned}
\overline{A} = A_J^{-1}A &= A,\\
\overline{b} = A_J^{-1}b &= b,\\
\pi = c_J^t A_J^{-1} &= 0,\\
\overline{c}^t = -\pi A + c^t &= c^t,\\
\beta = -\pi b &= 0,
\end{aligned}$$

d.h.

$$\begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix} = \begin{pmatrix} 6 & 15 & 1 & 0 & 0 & 0 & 4500 \\ 4 & 5 & 0 & 1 & 0 & 0 & 2000 \\ 20 & 10 & 0 & 0 & 1 & 0 & 8000 \\ -16 & -32 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Es ist

$$K = \{1,2\}$$

und

$$\overline{c}_K = \begin{pmatrix} -16 \\ -32 \end{pmatrix} < 0.$$

Wir wählen $s = 1$. Dann ist

$$\overline{a} = \begin{pmatrix} 6 \\ 4 \\ 20 \end{pmatrix} > 0.$$

Die Größen $\frac{\overline{b}_j}{\overline{a}_j}$ lauten

$$\frac{4500}{6} = 750, \quad \frac{2000}{4} = 500, \quad \frac{8000}{20} = 400.$$

Also ist $r = 1$ und damit

$$J' = \{3,4,1\}.$$

Für die Matrix F aus Satz I.4.2 erhalten wir

$$F = \begin{pmatrix} 1 & 0 & -\frac{6}{20} & 0 \\ 0 & 1 & -\frac{4}{20} & 0 \\ 0 & 0 & \frac{1}{20} & 0 \\ 0 & 0 & \frac{16}{20} & 1 \end{pmatrix}.$$

Damit ergibt sich das neue Tableau zu

$$\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} = F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & -\frac{6}{20} & 0 \\ 0 & 1 & -\frac{4}{20} & 0 \\ 0 & 0 & \frac{1}{20} & 0 \\ 0 & 0 & \frac{16}{20} & 1 \end{pmatrix} \begin{pmatrix} 6 & 15 & 1 & 0 & 0 & 0 & 4500 \\ 4 & 5 & 0 & 1 & 0 & 0 & 2000 \\ 20 & 10 & 0 & 0 & 1 & 0 & 8000 \\ -16 & -32 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 12 & 1 & 0 & -\frac{6}{20} & 0 & 2100 \\ 0 & 3 & 0 & 1 & -\frac{4}{20} & 0 & 400 \\ 1 & \frac{1}{2} & 0 & 0 & \frac{1}{20} & 0 & 400 \\ 0 & -24 & 0 & 0 & \frac{16}{20} & 1 & 6400 \end{pmatrix}.$$

Also ist

$$\overline{x}(J') = \begin{pmatrix} 400 \\ 0 \\ 2100 \\ 400 \\ 0 \end{pmatrix}, \quad \overline{z}(J') = 6400.$$

Bezogen auf das ursprüngliche LP sind wir also von der Ecke $(0,0)^t$ mit dem Zielfunktionswert 0 in die Ecke $(400,0)^t$ mit dem Zielfunktionswert 6400 gewandert.

Der Beweis von Satz I.4.4 (3) zeigt, dass Algorithmus I.4.1 genau dann zu einer echten Reduzierung der Zielfunktion führt, wenn

$$\min\left\{\frac{\overline{b}_j}{\overline{a}_j} : \overline{a}_j > 0\right\} > 0$$

ist. Dies ist äquivalent dazu, dass die aktuelle Basis nicht entartet ist. Falls Entartung vorliegt, nimmt die Zielfunktion nicht ab und es kann passieren, dass die neue Basis J' zu derselben Ecke gehört wie die alte Basis J. In diesem Fall kann man bei fortgesetzter Durchführung von Algorithmus I.4.1 in einen Zyklus geraten, d.h. man konstruiert eine Folge von Basen $J_1, \ldots, J_\ell$ mit $J_\ell = J_1$, die alle zur selben Ecke gehören. Um dies zu vermeiden, müssen wir Schritt (4) von Algorithmus I.4.1 modifizieren.

Definition I.4.6. Ein Zeilenvektor $u^t \in \mathbb{R}^n$ heißt *lexiko-positiv*, kurz $u^t >_\ell 0$, wenn gilt

$$u^t = (0, \ldots, 0, u_i, \ldots, u_n)$$

mit $i \geq 1$ und $u_i > 0$, d.h. die erste von Null verschiedene Komponente von u^t ist positiv. Es ist $u^t >_\ell v^t$ genau dann, wenn $(u-v)^t >_\ell 0$ ist.

BEMERKUNG I.4.7. (1) Zwei Vektoren sind bezüglich der lexikographischen Ordnung genau dann gleich groß, wenn sie identisch sind.
(2) Sei $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$ ein Tableau zu einer zulässigen Basis $\widehat{J} = J \oplus \{n+1\}$. Dann kann man die Variablen $x_1, \ldots, x_n$ so umnummerieren, dass die ersten m-Zeilen des permutierten Tableaus $\left(\begin{smallmatrix} \overline{b} & \overline{A} & 0 \\ \beta & \overline{c}^t & 1 \end{smallmatrix}\right)$ lexikopositiv sind, d.h.

$$e_j^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

für $1 \leq j \leq m$. Um dies einzusehen, nummeriere man die Unbekannten so um, dass $J = \{1, \ldots, m\}$ gilt. Dann ist

$$(\overline{b}, \overline{A}, 0) = (\overline{b}, I, \overline{a}_{m+1}, \ldots, \overline{a}_n)$$

mit geeigneten Vektoren $\overline{a}_{m+1}$, …, $\overline{a}_n$ und einem Vektor $\overline{b} \geq 0$. Daher ist in jeder Zeile entweder das erste Element positiv oder die Zeile hat die Form $(0, \ldots, 0, 1, *, \ldots, *)$.

Dies führt auf Algorithmus I.4.2.

Algorithmus I.4.2 Lexikographischer Simplexschritt

Gegeben: zulässige Basis J und Komplement K
Gesucht: neue Basis J' und zugehöriges Tableau

1: $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right) \leftarrow$ Tableau zu $\widehat{J} = J \oplus \{n+1\}$
2: Nummeriere die Variablen so um, dass die ersten m Zeilen von $(\overline{b}, \overline{A}, 0)$ lexiko-positiv sind.
3: $\overline{x}_J \leftarrow \overline{b}$, $\overline{x}_K \leftarrow 0$, $\overline{z} \leftarrow \beta$
4: **if** $\overline{c}_K \geq 0$ **then**
5: **stop** ▷ $\overline{x}$ ist Optimallösung
6: **end if**
7: $s \leftarrow$ Index in K mit $\overline{c}_s < 0$
8: $\overline{a} \leftarrow \overline{A}_s$ ▷ s-te Spalte von $\overline{A}$
9: **if** $\overline{a} \leq 0$ **then**
10: **stop** ▷ LP besitzt keine Lösung
11: **end if**
12: $r \leftarrow \operatorname{argmin}_{j;>_\ell}\{\frac{1}{\overline{a}_j}(\overline{b}, \overline{A}, 0) : j \in \{1, \ldots, m\}, \overline{a}_j > 0\}$
13: $J' \leftarrow (J \backslash \{i_r\}) \oplus \{s\} = \{i_1, \ldots, i_{r-1}, s, i_{r+1}, \ldots, i_m\}$
14: $\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} \leftarrow F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}$ ▷ neues Tableau

BEMERKUNG I.4.8. Falls

$$\min\left\{\frac{\overline{b}_j}{\overline{a}_j} : \overline{a}_j > 0\right\} > 0$$

ist, stimmen die Zeilen 11 und 12 der Algorithmen I.4.1 und I.4.2 überein. D.h., für eine nicht entartete Ecke ändert sich nichts. Falls aber

die Ecke entartet ist, wird die Wahl von r durch das Minimum bzgl. der lexikographischen Ordnung eindeutig. Daher können bei entarteten Ecken keine Zyklen auftreten.

LEMMA I.4.9. *Die Basis J sei zulässig und das zugehörige Tableau $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$ lexiko-positiv im Sinne von Zeile 2 von Algorithmus I.4.2. Dann führt Zeile 12 von Algorithmus I.4.2 auf eine benachbarte Basis J', so dass das zugehörige Tableau $\left(\widehat{J}'; \left(\begin{smallmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{smallmatrix}\right)\right)$ lexiko-positiv ist im Sinne von Zeile 2 von Algorithmus I.4.2. Insbesondere wächst die letzte Zeile des permutierten Tableaus $\left(\begin{smallmatrix} \overline{b} & \overline{A} & 0 \\ \beta & \overline{c}^t & 1 \end{smallmatrix}\right)$ im Sinne der lexikographischen Ordnung streng monoton an.*

BEWEIS. Das permutierte Tableau zu J' ist gegeben durch

$$\begin{pmatrix} \overline{b}' & \overline{A}' & 0 \\ \beta' & (\overline{c}')^t & 1 \end{pmatrix} = F \begin{pmatrix} \overline{b} & \overline{A} & 0 \\ \beta & \overline{c}^t & 1 \end{pmatrix}$$

mit der Matrix F aus Satz I.4.2. Wegen

$$e_j^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

für alle j und $\overline{a}_r > 0$ ist

$$e_r^t(\overline{b}', \overline{A}', 0) = \frac{1}{\overline{a}_r} e_r^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0.$$

Für $j \neq r$ gilt

$$e_j^t(\overline{b}', \overline{A}', 0) = e_j^t(\overline{b}, \overline{A}, 0) - \frac{\overline{a}_j}{\overline{a}_r} e_r^t(\overline{b}, \overline{A}, 0)$$

und daher

$$e_j^t(\overline{b}', \overline{A}', 0) \underset{\ell}{>} 0$$

genau dann, wenn gilt

$$\frac{1}{\overline{a}_j} e_j^t(\overline{b}, \overline{A}, 0) - \frac{1}{\overline{a}_r} e_r^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

falls $\overline{a}_j > 0$ bzw.

$$e_j^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

und

$$e_r^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

falls $\overline{a}_j \leq 0$.
Die zweite Forderung ist erfüllt, da nach Voraussetzung

$$e_j^t(\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0$$

ist für alle j. Die erste Forderung ist wegen der Wahl von r erfüllt, da das Minimum bzgl. der lexikographischen Ordnung eindeutig ist. Also

bleibt die lexikographische Ordnung des Tableaus erhalten.
Da

$$\overline{a}_{m+1} = \overline{c}_s < 0$$

ist, gilt für die letzte Zeile des Tableaus

$$(\beta', (\overline{c}')^t, 1) = (\beta, \overline{c}^t, 1) - \frac{\overline{a}_{m+1}}{\overline{a}_r} e_r^t (\overline{b}, \overline{A}, 0) \underset{\ell}{>} 0.$$

Dies beweist die Monotonie dieser Zeile bzgl. der lexikographischen Ordnung. □

Wir wenden uns nun dem Problem zu, eine zulässige Basis bzw. Ecke für den Start von Algorithmus I.4.1 zu finden.
Falls wie in Beispiel I.4.5 die Nebenbedingungen des ursprünglichen LP von der Form sind

$$Ax \leq b \quad \text{mit} \quad b \geq 0,$$

ist die Antwort einfach. Für die Standardform müssen wir Schlupfvariablen $y \in \mathbb{R}^m$ einführen, so dass die Nebenbedingungen übergehen in

$$Ax + y = b, \quad x \geq 0, \quad y \geq 0.$$

Offensichtlich ist dann $x = 0$, $y = b$ eine zulässige Ecke und $J = \{n+1, \ldots, n+m\}$ eine zulässige Basis.
Betrachte also die Nebenbedingung

$$Ax = b, \quad x \geq 0.$$

O.E. ist $b \neq 0$, da wir für $b = 0$ sofort die zulässige Ecke $x = 0$ sehen und dementsprechend z.B. $J = \{1, \ldots, m\}$ eine zulässige Basis ist.
Ebenso können wir o.E. $b \geq 0$ annehmen. Denn ist $b_j < 0$ für ein j, multiplizieren wir einfach die j-te Zeile des LGS mit -1, ohne die zulässige Menge zu ändern.

Satz I.4.10. *Betrachte das LP in Standardform* (I.1.2) *(S. 14) mit* $b \geq 0$ *und* $b \neq 0$. *Setze* $e = (1, \ldots, 1)^t \in \mathbb{R}^m$ *und betrachte das Hilfs-LP in Standardform*

$$\min\{e^t s : Ax + s = b\,,\ x \geq 0\,,\ s \geq 0\}. \tag{I.4.2}$$

(1) *Die zulässige Menge* $\mathcal{P}'$ *von* (I.4.2) *ist nicht leer. Das Problem* (I.4.2) *besitzt ein Optimum. Der Optimalwert von* (I.4.2) *ist nicht negativ.*
(2) *Der Optimalwert von* (I.4.2) *ist genau dann positiv, wenn die zulässige Menge* $\mathcal{P}$ *von* (I.1.2) *leer ist.*

Beweis. *ad (1)*: Offensichtlich ist $x = 0$, $s = b$ eine Ecke von $\mathcal{P}'$. Für alle $(x, s) \in \mathcal{P}'$ gilt wegen $s \geq 0$

$$e^t s \geq 0.$$

Also ist die Zielfunktion nach unten beschränkt. Da $\mathcal{P}' \neq \emptyset$ ist, besitzt das Problem (I.4.2) eine Lösung und der Optimalwert ist nicht negativ.

ad (2): Der Optimalwert von (I.4.2) sei positiv. Dann gilt für alle $(x, s) \in \mathcal{P}'$

$$e^t s > 0$$

d.h. $s \neq 0$. Für mindestens ein j ist also $s_j > 0$. Für die j-te Gleichung von (I.4.2) gilt dann

$$b_j = e_j^t Ax + s_j > e_j^t Ax.$$

Also kann nicht gelten $Ax = b$. Dies zeigt, dass die zulässige Menge von (I.1.2) leer ist.
Sei umgekehrt der Optimalwert von (I.4.2) gleich 0 und $(x^*, s^*) \in \mathcal{P}'$ eine Optimallösung. Wegen $s^* \geq 0$ und $e^t s^* = 0$ folgt $s^* = 0$ und damit $Ax^* = b$. Also ist x^* eine zulässige Ecke von $\mathcal{P}$. □

Wegen Satz I.4.10 (1) kann Algorithmus I.4.2 auf das Problem (I.4.2) angewandt werden und liefert nach endlich vielen Simplexschritten ein Optimum. Falls der Optimalwert gleich Null ist und die zugehörige Basis Komponenten des s-Vektors enthält, können endlich viele Austauschschritte gemäß Satz I.4.2 so durchgeführt werden, dass man eine Basis bestehend aus Komponenten des x-Vektors erhält. Dies führt auf Algorithmus I.4.3.

Algorithmus I.4.3 Simplex-Anlaufrechnung

Gegeben: LP in Standardform
Gesucht: zulässige Basis J' oder Information, dass LP nicht lösbar ist

1: **if** $b = 0$ **then**
2: $\quad J' \leftarrow \{1, \ldots, m\}$, **stop**
3: **end if**
4: Erzeuge äquivalentes Problem mit $b \geq 0$.
5: Stelle Hilfsproblem (I.4.2) auf.
6: Ausgehend von $J' = \{n+1, \ldots, n+m\}$ wende Algorithmus I.4.2 auf Problem (I.4.2) an; Ergebnis: Basis J', Optimalwert γ.
7: **if** $\gamma > 0$ **then**
8: $\quad$ **stop** ▷ LP besitzt keine Lösung
9: **end if**
10: **if** J' enthält Indizes größer als n **then**
11: $\quad$ Tausche Indizes $> n$ gegen Indizes $\leq n$; Ergebnis: Basis J'
▷ Algorithmus I.4.1 angewandt auf Problem (I.4.2)
12: **end if**

Fassen wir die Algorithmen I.4.1, I.4.2 und I.4.3 zusammen, erhalten wir Algorithmus I.4.4.

Da eine nicht leere zulässige Menge endlich viele Ecken hat, folgt aus den bisherigen Ergebnissen:

Satz I.4.11. *Der Simplexalgorithmus liefert nach endlich vielen Schritten entweder eine Optimallösung oder die Information, dass das LP nicht lösbar ist.*

Algorithmus I.4.4 Simplexalgorithmus

Gegeben: LP in allgemeiner Form
Gesucht: Optimallösung oder Information, dass LP nicht lösbar ist

```
1: Bringe LP auf äquivalente Standardform.
2: Bestimme zulässige Basis.
3: if keine Basis zulässige Basis gefunden then
4:     Wende Algorithmus I.4.3 an.
5:     if Algorithmus I.4.3 bricht mit Fehlermeldung ab then
6:         stop                                  ▷ LP nicht lösbar
7:     end if
8:     Wende Algorithmen I.4.1 oder I.4.2 (im Entartungsfall) an.
9: end if
```

SATZ I.4.12. *Der Aufwand für den Simplexalgorithmus für ein LP in Standardform mit n Variablen und m Gleichungen beträgt höchstens* $O(\binom{n}{m}(m+1)(n+1-m))$ *Operationen. Für festes n sind dies höchstens* $O(2^{\frac{n}{2}}(\frac{n}{2})^2)$ *Operationen.*

BEWEIS. Die zulässige Menge hat höchstens $\binom{n}{m}$ Ecken. Daher benötigt der Simplexalgorithmus höchstens $O(\binom{n}{m})$ Schritte. Jeder dieser Schritte erfordert gemäß Bemerkung I.4.3 $O((m+1)(n+1-m))$ Operationen. Die Ausdrücke $\binom{n}{m}$ und $(m+1)(n+1-m)$ sind jeweils für $m = \frac{n}{2}$ maximal und liefern die Werte $2^{\frac{n}{2}}$ bzw. $(\frac{n}{2})^2$. □

BEMERKUNG I.4.13. (1) Die Abschätzung $\binom{n}{m}$ für die Zahl der Ecken ist scharf. Um dies einzusehen, betrachte man den k-dimensionalen Einheitswürfel $\{z \in \mathbb{R}^k : 0 \le z_i \le 1\ ,\ 1 \le i \le k\}$. Er hat 2^k Ecken. Durch Einführen von Schlupfvariablen erhalten wir das äquivalente LP in Standardform mit $m = k$, $n = 2k$.
(2) Die Abschätzung von Satz I.4.12 über die Zahl der erforderlichen Schritte ist extrem pessimistisch. In der Praxis beobachtet man eine lineare Laufzeit des Simplexalgorithmus, d.h. man benötigt $O(n)$ Schritte entsprechend $O(n^3)$ Operationen. Von Klee und Minty wurden aber Beispiele konstruiert, in denen der Simplexalgorithmus tatsächlich eine exponentielle Laufzeit benötigt.

I.5. Dualität

Wir erinnern an unsere allgemeine Voraussetzung (I.3.1) (S. 20) und betrachten ein LP in Standardform (I.1.2) (S. 14). O.E. setzen wir voraus, dass $b \ge 0$ ist, sonst multiplizieren wir entsprechende Gleichungen mit -1.

Es ist relativ leicht, obere Schranken für den Optimalwert von (I.1.2) zu erhalten: Jeder zulässige Punkt $x \in \mathcal{P}$ liefert mit $c^t x$ eine solche obere Schranke.

Wir möchten aber gerne auch eine möglichst gute untere Schranke für den Optimalwert von (I.1.2) haben. Ist $d \in \mathbb{R}^n$ irgendein Vektor mit $d \leq c$ und $x \in \mathcal{P}$ gilt wegen $x \geq 0$ natürlich

$$d^t x \leq c^t x.$$

Aber wie erhält man gute Kandidaten für d?

Sei dazu J eine zulässige Basis für (I.1.2), $x = x(J)$ die zugehörige Basislösung und K das zugehörige Komplement. Dann ist

$$x_K = 0 \quad \text{und} \quad x_J = A_J^{-1} b.$$

Damit folgt

$$c^t x = c_J^t x_J = c_J^t A_J^{-1} b = b^t A_J^{-t} c_J \geq b^t y$$

für jeden Vektor $y \in \mathbb{R}^m$ mit

$$y \leq A_J^{-t} c_J.$$

Man beachte, dass wir an dieser Stelle die Voraussetzung $b \geq 0$ ausgenutzt haben.
Falls die Elemente von A_J nicht negativ sind, folgt aus

$$y \leq A_J^{-t} c_J$$

sofort

$$A_J^t y \leq c_J.$$

Wir betrachten daher Vektoren $d \in \mathbb{R}^n$ von der Form

$$d = A^t y \quad \text{mit} \quad A^t y \leq c.$$

Falls es solche Vektoren überhaupt gibt, gilt für alle diese Vektoren

$$c^t x \geq d^t x = y^t A x = y^t b = b^t y,$$

d.h. $b^t y$ liefert eine untere Schranke für den Optimalwert von (I.1.2).

Diese Überlegungen führen auf folgende Definition:

Definition I.5.1. Das LP

$$\max\{b^t y : A^t y \leq c\} \tag{I.5.1}$$

heißt das *duale Programm* zu dem LP (I.1.2) (S. 14) in Standardform.

Satz I.5.2. *Mit der Konvention*

$$\max \emptyset = -\infty, \quad \min \emptyset = \infty$$

gilt

$$\min\{c^t x : Ax = b\ ,\ x \geq 0\} = \max\{b^t y : A^t y \leq c\},$$

sofern mindestens eines der LP (I.1.2) *(S. 14) und* (I.5.1) *eine zulässige Lösung besitzt.*

BEWEIS. Bezeichne mit $\mathcal{P}$ bzw. $\mathcal{D}$ die zulässige Menge von (I.1.2) bzw. (I.5.1) und setze

$$P_{\inf} = \inf\{c^t x : x \in \mathcal{P}\}, \quad D_{\sup} = \sup\{b^t y : y \in \mathcal{D}\}.$$

Ist $x \in \mathcal{P}$ und $y \in \mathcal{D}$, folgt

$$c^t x \geq (A^t y)^t x = y^t A x = y^t b = b^t y.$$

Also ist

$$D_{\sup} \leq P_{\inf}.$$

Insbesondere folgt

$$\mathcal{D} = \emptyset \quad \text{falls} \quad P_{\inf} = -\infty,$$
$$\mathcal{P} = \emptyset \quad \text{falls} \quad D_{\sup} = \infty.$$

Also ist in diesen beiden Fällen die Behauptung bewiesen, und wir müssen nur noch die Fälle betrachten, dass $P_{\inf}$ oder $D_{\sup}$ endlich sind. Sei also $P_{\inf}$ endlich. Dann folgt aus §I.4, dass das LP (I.1.2) (S. 14) eine Optimallösung x^* besitzt und dass der Simplexalgorithmus eine zugehörige Basis J mit Komplement K liefert. Für das zugehörige Tableau $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$ gilt $\widehat{J} = J \oplus \{n+1\}$ und

$$\begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix} = \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} \begin{pmatrix} A & 0 & b \\ c^t & 1 & 0 \end{pmatrix}$$

und

$$\begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} = \begin{pmatrix} A_J^{-1} & 0 \\ -\pi & 1 \end{pmatrix}$$

mit

$$\pi = c_J^t A_J^{-1}.$$

Wegen Satz I.4.4(1) (S. 29) ist

$$\overline{c}_K \geq 0.$$

Weiter ist

$$\overline{c}_J^t = -\pi A + c_J^t = 0.$$

Definiere den Vektor y^* durch

$$y^* = \pi^t.$$

Dann folgt

$$A_K^t y^* = A_K^t \pi^t = (\pi A_K)^t = (\pi A_K)^t - c_K + c_K = -\overline{c}_K + c_K \leq c_K$$

und

$$A_J^t y^* = A_J^t \pi^t = (\pi A_J)^t = c_J.$$

Also ist $Ay^* \leq c$, d.h. $y^* \in \mathcal{D}$. Weiter gilt

$$b^t y^* = b^t \pi^t = \pi b = -[-\pi b + 1 \cdot 0] = -\beta = c^t x^* = P_{\inf}.$$

Also ist in diesem Fall $D_{\sup} = P_{\inf}$ und y^* ist eine Optimallösung von (I.5.1).

Sei nun $D_{\sup}$ endlich. Wir schreiben das LP (I.5.1) um in ein äquivalentes LP (I.5.2) in Standardform und wenden auf das neue LP (I.5.2) das soeben Gezeigte an.
Da jeder Vektor y in der Form $y = u - v$ mit $u \geq 0$ und $v \geq 0$ geschrieben werden kann und da

$$\max b^t y = -\min(-b)^t y$$

ist, ist gemäß Bemerkung I.1.6 (S. 14) das LP (I.5.1) äquivalent zu

$$-\min\{-b^t(u-v) : A^t(u-v)+s = c\ ,\ u \geq 0\ ,\ v \geq 0\ ,\ s \geq 0\}. \tag{I.5.2}$$

Bis auf das Vorzeichen ist dieses LP von der Form

$$\min\{\widetilde{b}^t\widetilde{x} : \widetilde{A}\widetilde{x} = \widetilde{c}\ ,\ \widetilde{x} \geq 0\}$$

mit

$$\widetilde{b} = \begin{pmatrix} -b \\ b \\ 0 \end{pmatrix}, \quad \widetilde{x} = \begin{pmatrix} u \\ v \\ s \end{pmatrix}, \quad \widetilde{c} = c, \quad \widetilde{A} = \begin{pmatrix} A^t & -A^t & I \end{pmatrix}.$$

Das zugehörige duale Problem ist

$$\max\{\widetilde{c}^t\widetilde{y} : \widetilde{A}^t\widetilde{y} \leq \widetilde{b}\} \tag{I.5.3}$$

mit $\widetilde{y} \in \mathbb{R}^n$. Die Nebenbedingungen lauten

$$A\widetilde{y} \leq -b \quad \text{und} \quad -A\widetilde{y} \leq b$$

also

$$A\widetilde{y} = -b.$$

Also ist (I.5.3) unser ursprüngliches Problem (I.1.2). Daher beweist das bisher Gezeigte die Behauptung auch in diesem Fall. □

Bemerkung I.5.3. Im letzten Teil des obigen Beweises haben wir implizit gezeigt, dass das duale LP zum Problem (I.5.1) das ursprüngliche LP (I.1.2) (S. 14) ist.

Im Beweis von Satz I.5.2 haben wir gesehen, dass man für eine beliebige Basis J des LP (I.1.2) den Vektor

$$y(J) = A_J^{-t} c_J$$

definieren kann. Dieser ist für das duale Problem (I.5.1) genau dann zulässig, wenn gilt

$$\overline{c}_K = c_K - A_K^t y(J) \geq 0.$$

In diesem Fall ist der Wert der dualen Zielfunktion

$$b^t y(J) = b^t A_J^{-t} c_J = \overline{b}^t c_J = x(J)_J^t c_J = c^t x(J).$$

Falls $x(J)$ für (I.1.2) zulässig ist, hat man Lösungen für (I.5.1) und (I.1.2) gefunden. Andernfalls kann man versuchen, den Wert der dualen Zielfunktion unter Beibehaltung der Zulässigkeit für (I.5.1) so lange zu vergrößern, bis man ein für (I.1.2) zulässiges $x(J)$ gefunden hat.

Diese Idee führt auf Definition I.5.4 und Algorithmus I.5.1.

DEFINITION I.5.4. Eine Basis J für das LP (I.1.2) (S. 14) mit Komplement K heißt *dual zulässig*, wenn $\overline{c}_K \geq 0$ ist.

Algorithmus I.5.1 Dualer Simplexschritt

Gegeben: dual zulässige Basis J und Tableau $\left(\widehat{J}; \left(\begin{smallmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{smallmatrix}\right)\right)$
Gesucht: neue dual zulässige Basis J' und Tableau

1: **if** $\overline{b} \geq 0$ **then**
2: **stop** ▷ J ist primal zulässig und optimal
3: **end if**
4: $r \leftarrow$ Index in $\{1, \ldots, m\}$ mit $\overline{b}_r < 0$, $\overline{a}_{r,K} \leftarrow r$-te Zeile von $\overline{A}_K$
5: **if** $\overline{a}_{r,K} \geq 0$ **then**
6: **stop** ▷ LP besitzt keine Lösung
7: **end if**
8: $s \leftarrow \operatorname{argmax}\left\{\dfrac{\overline{c}_s}{\overline{a}_{r,s}} : s \in K \text{ und } \overline{a}_{r,s} < 0\right\}$
9: $J' \leftarrow (J \backslash \{i_r\}) \oplus \{s\}$, berechne Tableau zu $J' \oplus \{n+1\}$

BEMERKUNG I.5.5. Algorithmus I.5.1 sucht eine Lösung des dualen Problems (I.5.1), arbeitet aber mit den Daten des primalen Problems (I.1.2).

SATZ I.5.6. *Falls Algorithmus I.5.1 eine neue Basis J' liefert, ist diese wieder dual zulässig und es gilt*

$$c^t x(J') \geq c^t x(J).$$

BEWEIS. Das Tableau zu $\widehat{J'} = J' \oplus \{n+1\}$ erhält man durch Multiplikation mit der Matrix F aus Satz I.4.2 (S. 27):

$$\begin{pmatrix} \overline{A}' & 0 & \overline{b}' \\ (\overline{c}')^t & 1 & \beta' \end{pmatrix} = F \begin{pmatrix} \overline{A} & 0 & \overline{b} \\ \overline{c}^t & 1 & \beta \end{pmatrix}.$$

Insbesondere folgt für die letzte Zeile des Tableaus mit den Notationen aus Satz I.4.2

$$\begin{aligned} \overline{c}' &= \overline{c} - \frac{\alpha_{m+1}}{\alpha_r} e_r^t \overline{A} = \overline{c} - \frac{\overline{c}_s}{\alpha_r} e_r^t \overline{A}, \\ \beta' &= \beta - \frac{\overline{c}_s}{\alpha_r} \overline{b}, \end{aligned} \tag{I.5.4}$$

wobei

$$\alpha_{m+1} = \overline{c}_s \geq 0$$

ist, da J nach Voraussetzung dual zulässig ist. Aufgrund der Auswahlregeln für r und s gilt weiter

$$\overline{b}_r < 0, \quad \alpha_r = \overline{A}_{r,s} < 0, \quad \frac{\overline{c}_s}{\alpha_r} = \frac{\overline{c}_s}{\overline{A}_{r,s}} = \max\left\{\frac{\overline{c}_{s'}}{\overline{A}_{r,s'}} : \overline{A}_{r,s'} < 0\right\}.$$

Wir müssen nun zeigen, dass $\overline{c}'_\ell \geq 0$ ist für alle $\ell \in K' = (K \oplus \{i_r\}) \backslash \{s\}$. Für $\ell = i_r$ ist $\overline{c}_{i_r} = 0$ und $\overline{A}_{r,i_r} = 1$, weil $i_r \in J$ ist. Damit folgt aus der ersten Gleichung von (I.5.4)

$$\overline{c}'_{i_r} = \overline{c}_{i_r} - \frac{\overline{c}_s}{\overline{A}_{r,s}} \overline{A}_{r,i_r} = -\frac{\overline{c}_s}{\overline{A}_{r,s}} \geq 0.$$

Für $\ell \in K$ mit $\ell \neq s$ und $\overline{A}_{r,\ell} \geq 0$ ist

$$\overline{c}'_\ell = \overline{c}_\ell - \frac{\overline{c}_s}{\overline{A}_{r,s}} \overline{A}_{r,\ell} \geq \overline{c}_\ell \geq 0.$$

Für $\ell \in K$ mit $\ell \neq s$ und $\overline{A}_{r,\ell} < 0$ folgt

$$\overline{c}'_\ell = \overline{c}_\ell - \frac{\overline{c}_s}{\overline{A}_{r,s}} \overline{A}_{r,\ell} = \overline{c}_\ell + \frac{\overline{c}_s}{\overline{A}_{r,s}} (-\overline{A}_{r,\ell}) \geq \overline{c}_\ell + \frac{\overline{c}_\ell}{\overline{A}_{r,\ell}} (-\overline{A}_{r,\ell}) = 0.$$

Also ist J' dual zulässig.
Die Behauptung für die primale Zielfunktion folgt aus der zweiten Gleichung von (I.5.4). □

Der duale Simplexalgorithmus ist besonders dann von Vorteil, wenn eine dual zulässige Basis leicht erkennbar ist, während eine zulässige Basis für das primale Problem nicht offensichtlich ist.

Beispiel I.5.7. Wir betrachten das Problem

$$x_1 + x_2 \to \min$$

unter

$$\begin{aligned} -2x_1 - x_2 &\leq -3, \\ -x_1 - 2x_2 &\leq -3, \\ x &\geq 0. \end{aligned}$$

Wir führen Schlupfvariablen x_3 und x_4 ein. Dann lauten die Daten des LP

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

Als erste Basis wählen wir $J = \{3, 4\}$. Das zugehörige Tableau ergibt sich zu

$$\begin{array}{rrrr|c|r} -2^* & -1 & 1 & 0 & 0 & -3 \\ -1 & -2 & 0 & 1 & 0 & -3 \\ \hline 1 & 1 & 0 & 0 & 1 & 0 \end{array}.$$

Es ist primal nicht zulässig. Allerdings ist es dual zulässig. In Zeile 4 von Algorithmus I.5.1 wählen wir $r = 1$, d.h. $i_1 = 3$ soll die Basis verlassen. Dies führt wegen

$$\frac{1}{-2} > \frac{1}{-1}$$

auf $s = 1$ und das mit * gekennzeichnete Element des Tableaus. Die neue Basis ist $\{1, 4\}$ und das zugehörige Tableau ergibt sich zu

$$\begin{array}{cccc|c|c} 1 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{3}{2} \\ 0 & -\frac{3}{2}^* & -\frac{1}{2} & 1 & 0 & -\frac{3}{2} \\ \hline 0 & \frac{1}{2} & \frac{1}{2} & 0 & 1 & -\frac{3}{2} \end{array}.$$

Diese Basis ist nach wie vor primal nicht zulässig. Zeile 4 von Algorithmus I.5.1 führt jetzt zu $r = 2$ und $s = 2$ und das mit * markierte Element im Tableau. Die neue Basis ist $\{1, 2\}$, das zugehörige Tableau lautet

$$\begin{array}{cccc|c|c} 1 & 0 & -\frac{2}{3} & \frac{1}{3} & 0 & 1 \\ 0 & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 1 \\ \hline 0 & 0 & \frac{1}{3} & \frac{1}{3} & 1 & -2 \end{array}.$$

Dieses Tableau ist primal zulässig. Die zugehörige Basislösung

$$x^* = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

liefert das Optimum für das LP, der Wert der Zielfunktion ist 2.

I.6. Sensitivitätsanalyse

In diesem Abschnitt zeigen wir an Hand einiger Beispiele wie man aus dem Tableau zu einer Optimallösung eines LP ablesen kann, wie sich die Optimallösung und der Optimalwert ändern, wenn sich die Daten des LP ändern.

Beispiel I.6.1. Wir greifen Beispiel 1 (S. 5) auf. Wie wir in Beispiel I.4.5 (S. 30) gesehen haben, führt die Transformation auf Standardform auf die Ausgangsbasis $J = \{3, 4, 5\}$ und das Tableau

$$\begin{array}{rrccc|c|r} 6 & 15 & 1 & 0 & 0 & 0 & 4500 \\ 4 & 5 & 0 & 1 & 0 & 0 & 2000 \\ 20 & 10 & 0 & 0 & 1 & 0 & 8000 \\ \hline -16 & -32 & 0 & 0 & 0 & 1 & 0 \end{array}.$$

Aus der graphischen Lösung von Beispiel 1 wissen wir, dass $J = \{1, 2, 5\}$ die optimale Basis sein sollte. Wir wollen dies verifizieren, indem wir das zugehörige Tableau bestimmen. Mit etwas Rechnung

erhalten wir

$$\begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 6 & 15 & 0 & 0 \\ 4 & 5 & 0 & 0 \\ 20 & 10 & 1 & 0 \\ -16 & -32 & 0 & 1 \end{pmatrix}^{-1} = \frac{1}{30} \begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 48 & 48 & 0 & 30 \end{pmatrix}.$$

Damit ergibt sich das zugehörige Tableau zu

$$\begin{array}{cccccc|c} 1 & 0 & -\frac{1}{6} & \frac{1}{2} & 0 & 0 & 250 \\ 0 & 1 & \frac{2}{15} & -\frac{1}{5} & 0 & 0 & 200 \\ 0 & 0 & 2 & -8 & 1 & 0 & 1000 \\ \hline 0 & 0 & \frac{8}{5} & \frac{8}{5} & 0 & 1 & 10400 \end{array}.$$

Wegen $\overline{c}_K \geq 0$ haben wir ein Optimum gefunden und lesen die Optimallösung ab: 250 Paar Damenschuhe, 200 Paar Herrenschuhe, 10400 € Gewinn.

Beispiel I.6.2. Der Fabrikant überlegt, ob er zusätzlich Damenstiefel produzieren soll. Ein Paar benötigt $24\,\mathrm{dm}^2$ Leder, 16 Stunden Maschinenzeit und 100 Stunden menschliche Arbeitszeit und bringt einen Gewinn von 60 € .
Wir müssen eine neue Variable x_6 für die Damenstiefel einführen und eine zusätzliche Spalte

$$z = \begin{pmatrix} 24 \\ 16 \\ 100 \\ -60 \end{pmatrix}$$

in die Matrix $\left(\begin{smallmatrix} A & 0 & b \\ c^t & 1 & 0 \end{smallmatrix}\right)$ aufnehmen. Wegen

$$\begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} z = \begin{pmatrix} * \\ * \\ * \\ c_6 \end{pmatrix}$$

mit

$$c_6 = \frac{1}{30}(48 \cdot 24 + 48 \cdot 16 - 30 \cdot 60) = \frac{8}{5} \cdot 40 - 60 = 4 > 0$$

ist die Basis $J = \{1, 2, 5\}$ nach wie vor optimal. Es lohnt daher nicht, die Produktion umzustellen und die Damenstiefel in das Programm aufzunehmen.

Der Fabrikant fragt sich nun, wie groß der Gewinn pro Paar Damenstiefel sein muss, damit sich eine Umstellung lohnt.
Bezeichnen wir mit g diesen Gewinn, müssen wir die Spalte

$$z_g = \begin{pmatrix} 24 \\ 16 \\ 100 \\ -g \end{pmatrix}$$

in die Matrix $\left(\begin{smallmatrix} A & 0 & b \\ c^t & 1 & 0 \end{smallmatrix}\right)$ aufnehmen. Dies führt auf

$$c_6 = e_4^t \begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} z_g = \frac{8}{5}(24+16) - g = 64 - g.$$

Damit sich die Umstellung lohnt, muss $c_6 < 0$ also $g > 64$ € sein.
Der Fabrikant handelt jetzt einen Gewinn von 66 € pro Paar Damenstiefel aus. Wie lautet die optimale Produktion jetzt?
Wir erhalten

$$\begin{pmatrix} A_J & 0 \\ c_J^t & 1 \end{pmatrix}^{-1} z_{66} = \frac{1}{30} \begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 48 & 48 & 0 & 30 \end{pmatrix} \begin{pmatrix} 24 \\ 16 \\ 100 \\ -66 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \\ 20 \\ -2 \end{pmatrix}$$

Wegen

$$\frac{250}{4} = 62.5 > 50 = \frac{1000}{20}$$

ist im Simplexschritt $s = 6$ und $r = 3$, d.h. die Schlupfvariable x_5 wird gegen die Variable x_6 für die Damenstiefel ausgetauscht. Die Matrix F aus Satz I.4.2 (S. 27) lautet daher

$$F = \begin{pmatrix} 1 & 0 & -\frac{1}{5} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{20} & 0 \\ 0 & 0 & \frac{1}{10} & 1 \end{pmatrix}.$$

Damit ergibt sich das neue Tableau zu

$$\begin{array}{cccccc|c|c} 1 & 0 & * & * & * & 0 & 0 & 50 \\ 0 & 1 & * & * & * & 0 & 0 & 200 \\ 0 & 0 & * & * & * & 1 & 0 & 50 \\ \hline 0 & 0 & \frac{9}{5} & \frac{4}{5} & \frac{1}{10} & 0 & 1 & 10500 \end{array}.$$

Diese Tableau ist optimal und wir lesen die Lösung ab: 50 Paar Damenschuhe, 200 Paar Herrenschuhe, 50 Paar Damenstiefel und 10500 € Gewinn.

Beispiel I.6.3. Durch Rationalisierungsmaßnahmen will der Fabrikant die monatliche Arbeitszeit auf 6000 Stunden reduzieren und die Maschinenlaufzeit auf 2500 Stunden erhöhen. Ist dies machbar?
Die letzte Spalte von $\left(\begin{smallmatrix} A & 0 & b \\ c^t & 1 & 0 \end{smallmatrix}\right)$ ändert sich zu

$$\begin{pmatrix} 4500 \\ 2500 \\ 6000 \\ 0 \end{pmatrix}.$$

Damit lautet die neue letzte Spalte des Tableaus

$$\frac{1}{30}\begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 48 & 48 & 0 & 30 \end{pmatrix}\begin{pmatrix} 4500 \\ 2500 \\ 6000 \\ 0 \end{pmatrix} = \begin{pmatrix} 1500 \\ 100 \\ -5000 \\ 11200 \end{pmatrix}.$$

Da die dritte Komponente negativ wird, ist die Basis $J = \{1, 2, 5\}$ nicht mehr zulässig und das Anliegen scheitert.
Der Fabrikant überlegt nun, dass er die Maschinenlaufzeit um m Stunden erhöhen und dadurch die menschliche Arbeitszeit um $2m$ Stunden verringern will. Er fragt sich, ob dies möglich ist und wie groß er ggf. m wählen kann und welchen Gewinn die Maßnahme ggf. bringt.
Die neue letzte Spalte von $\left(\begin{smallmatrix} A & 0 & b \\ c^t & 1 & 0 \end{smallmatrix}\right)$ lautet

$$\begin{pmatrix} 4500 \\ 2000 + m \\ 8000 - 2m \\ 0 \end{pmatrix}$$

und ergibt als letzte Spalte des Tableaus

$$\frac{1}{30}\begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 48 & 48 & 0 & 30 \end{pmatrix}\begin{pmatrix} 4500 \\ 2000 + m \\ 8000 - 2m \\ 0 \end{pmatrix} = \begin{pmatrix} 250 + \frac{m}{2} \\ 200 - \frac{m}{5} \\ 1000 - 10m \\ 10400 + \frac{8}{5}m \end{pmatrix}.$$

Falls die Maßnahme durchführbar ist, bringt sie also auf jeden Fall einen Gewinn. Sie ist durchführbar, wenn die ersten drei Komponenten der letzten Tableauspalte alle nicht negativ sind. Dies gibt die Bedingungen

$$\begin{aligned} 200 - \frac{m}{5} \geq 0 &\implies m \leq 1000 \\ 1000 - 10m \geq 0 &\implies m \leq 100. \end{aligned}$$

Also ist $m = 100$ maximal möglich. Diese Wahl liefert 300 Paar Damenschuhe, 180 Paar Herrenschuhe und 10560 € Gewinn.

Beispiel I.6.4. Der Hauptabnehmer des Fabrikanten möchte den Preis für Damenschuhe reduzieren, so dass der Gewinn pro Paar nur noch 14 € beträgt. Als Ausgleich schlägt er vor, für ein Paar Herrenschuhe 2 € mehr zu zahlen. Soll sich der Fabrikant auf diesen Vorschlag

einlassen?
Die neuen Gewinne führen auf den neuen Vektor

$$\widetilde{c} = \begin{pmatrix} -14 \\ -34 \\ 0 \\ 0 \end{pmatrix},$$

und wir müssen

$$\begin{pmatrix} A_J & 0 \\ \widetilde{c}_J^t & 1 \end{pmatrix}^{-1}$$

berechnen. Aus dem Beweis von Satz I.3.12 (S. 26) wissen wir, dass gilt

$$\begin{pmatrix} A_J & 0 \\ \widetilde{c}_J^t & 1 \end{pmatrix}^{-1} = \begin{pmatrix} A_J^{-1} & 0 \\ -\widetilde{\pi} & 1 \end{pmatrix}$$

mit

$$\widetilde{\pi} = \widetilde{c}_J^t A_J^{-1} = \begin{pmatrix} -14 & -34 & 0 \end{pmatrix} \begin{pmatrix} -5 & 15 & 0 \\ 4 & -6 & 0 \\ 60 & -240 & 30 \end{pmatrix} \frac{1}{30} = \begin{pmatrix} -\frac{11}{5} & -\frac{1}{5} & 0 \end{pmatrix}.$$

Also ist

$$\begin{pmatrix} A_J & 0 \\ \widetilde{c}_J^t & 1 \end{pmatrix}^{-1} = \frac{1}{30} \begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 66 & 6 & 0 & 30 \end{pmatrix}.$$

Damit ergibt sich das neue Tableau zu

$$\begin{array}{ccccc|c|c} 1 & 0 & * & * & 0 & 0 & 250 \\ 0 & 1 & * & * & 0 & 0 & 200 \\ 0 & 0 & * & * & 1 & 0 & 1000 \\ \hline 0 & 0 & \frac{11}{5} & \frac{1}{5} & 0 & 1 & 10300 \end{array}.$$

Der Handel wäre also für den Fabrikanten nachteilig.

Für das letzte Beispiel benötigen wir das folgende Ergebnis:

Satz I.6.5 (Sherman-Morrison-Woodbury-Formel). *Sei $A \in \mathbb{R}^{n\times n}$ invertierbar, $u, v \in \mathbb{R}^n$ und $v^t A^{-1} u \neq -1$. Dann ist $A+uv^t$ invertierbar und*

$$(A + uv^t)^{-1} = A^{-1} - \frac{1}{1 + v^t A^{-1} u} A^{-1} u v^t A^{-1}.$$

Beweis. Wegen $v^t A^{-1} u \neq -1$ ist

$$B = A^{-1} - \frac{1}{1 + v^t A^{-1} u} A^{-1} u v^t A^{-1}$$

wohl definiert. Ausmultiplizieren liefert

$$\begin{aligned} B(A+uv^t) &= I + A^{-1}uv^t - \frac{1}{1+v^tA^{-1}u}A^{-1}uv^t \\ &\quad - \frac{1}{1+v^tA^{-1}u}\underbrace{A^{-1}uv^tA^{-1}u}_{=(v^tA^{-1}u)A^{-1}u}v^t \\ &= I \end{aligned}$$

und

$$\begin{aligned} (A+uv^t)B &= I - \frac{1}{1+v^tA^{-1}u}uv^tA^{-1} + uv^tA^{-1} \\ &\quad - \frac{1}{1+v^tA^{-1}u}\underbrace{uv^tA^{-1}uv^t}_{=(v^tA^{-1}u)uv^t}A^{-1} \\ &= I. \end{aligned}$$

□

Beispiel I.6.6. Krankheitsbedingt kommt es zu Engpässen in der Produktion, die begrenzt durch höhere Maschinenlaufzeiten ausgeglichen werden können. Dies äußert sich in einer Störung der Matrix A der Form

$$A_\varepsilon = \begin{pmatrix} 6 & 15 \\ 4+2\varepsilon & 5+5\varepsilon \\ 20-4\varepsilon & 10-10\varepsilon \end{pmatrix}$$

mit kleinem positiven ε. Wie groß darf die Störung sein, ohne das Optimum zu beeinflussen, und wie wirkt sich die Störung auf den Gewinn aus?
Wir müssen ε maximal so bestimmen, dass die Basis $J = \{1,2,5\}$ optimal bleibt. Wir müssen daher die gestörte Matrix

$$B_\varepsilon = \begin{pmatrix} 6 & 15 & 0 & 0 \\ 4+2\varepsilon & 5+5\varepsilon & 0 & 0 \\ 20-4\varepsilon & 10-10\varepsilon & 1 & 0 \\ -16 & -32 & 0 & 1 \end{pmatrix}$$

invertieren. Mit

$$B = \begin{pmatrix} 6 & 15 & 0 & 0 \\ 4 & 5 & 0 & 0 \\ 20 & 10 & 1 & 0 \\ -16 & -32 & 0 & 1 \end{pmatrix}$$

ist B_ε von der Form

$$B_\varepsilon = B + \varepsilon uv^t$$

mit

$$u = \begin{pmatrix} 0 \\ 1 \\ -2 \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} 2 \\ 5 \\ 0 \\ 0 \end{pmatrix}.$$

Es ist

$$\begin{aligned} v^t B^{-1} u &= \frac{1}{30} \begin{pmatrix} 2 & 5 & 0 & 0 \end{pmatrix} \begin{pmatrix} -5 & 15 & 0 & 0 \\ 4 & -6 & 0 & 0 \\ 60 & -240 & 30 & 0 \\ 48 & 48 & 0 & 30 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -2 \\ 0 \end{pmatrix} \\ &= \frac{1}{30} \begin{pmatrix} 2 & 5 & 0 & 0 \end{pmatrix} \begin{pmatrix} 15 \\ -6 \\ * \\ * \end{pmatrix} \\ &= 0. \end{aligned}$$

Daher können wir Satz I.6.5 anwenden und erhalten

$$\begin{aligned} B_\varepsilon^{-1} &= B^{-1} - \varepsilon B^{-1} u v^t B^{-1} \\ &= B^{-1} - \varepsilon \frac{1}{30} \begin{pmatrix} 15 \\ -6 \\ -300 \\ 48 \end{pmatrix} \begin{pmatrix} 10 & 0 & 0 & 0 \end{pmatrix} \frac{1}{30}. \end{aligned}$$

Die letzte Spalte $\begin{pmatrix} \bar{b} \\ \beta \end{pmatrix}$ des Tableaus ergibt sich damit zu

$$\begin{aligned} B_\varepsilon^{-1} \begin{pmatrix} 4500 \\ 2000 \\ 8000 \\ 0 \end{pmatrix} &= \begin{pmatrix} 250 \\ 200 \\ 1000 \\ 10400 \end{pmatrix} - \frac{\varepsilon}{900} \begin{pmatrix} 15 \\ -6 \\ -300 \\ 48 \end{pmatrix} 45000 \\ &= \begin{pmatrix} 250 \\ 200 \\ 1000 \\ 10400 \end{pmatrix} - \varepsilon \begin{pmatrix} 750 \\ -300 \\ -15000 \\ 2400 \end{pmatrix} \\ &= \begin{pmatrix} 250 - 750\varepsilon \\ 200 + 300\varepsilon \\ 1000 + 15000\varepsilon \\ 10400 - 2400\varepsilon \end{pmatrix}. \end{aligned}$$

Also ist $J = \{1, 2, 5\}$ nach wie vor zulässig, wenn gilt

$$250 - 750\varepsilon \geq 0 \quad \iff \quad \varepsilon \leq \frac{1}{3}.$$

Für die Optimalität müssen wir den Vektor $\bar{c}_K$ überprüfen, d.h. das dritte und vierte Element in der letzten Tableauzeile. Da die dritte und vierte Spalte von $\left(\begin{smallmatrix} A & 0 & b \\ c^t & 1 & 0 \end{smallmatrix}\right)$ der erste bzw. zweite Einheitsvektor ist, interessieren uns also die ersten beiden Elemente in der letzten Zeile

von B_ε^{-1}. Aus obiger Darstellung von B_ε^{-1} erhalten wir

$$B_\varepsilon^{-1} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ \frac{8}{5} - \varepsilon\frac{48}{90} & \frac{8}{5} & * & * \end{pmatrix}.$$

Also ist $J = \{1, 2, 5\}$ nach wie vor optimal, wenn gilt

$$\frac{8}{5} - \varepsilon\frac{48}{90} \geq 0 \quad \Longleftrightarrow \quad \varepsilon \leq 3.$$

Für $0 \leq \varepsilon \leq \frac{1}{3}$ bleibt also die Basis optimal. Die Produktion von Damenschuhen wird reduziert, die von Herrenschuhen wird gesteigert, der Gewinn geht zurück.

I.7. Innere-Punkt-Methoden

Wie wir in Satz I.4.12 (S. 38) und Bemerkung I.4.13 (S. 38) gesehen haben, benötigt der Simplexalgorithmus im ungünstigsten Fall einen exponentiellen Aufwand. In diesem Abschnitt beschreiben wir Methoden, die einen algebraischen Aufwand benötigen. Sie erzeugen eine Folge von Näherungen, die im *Innern* der zulässigen Menge liegen. Nach $O(\sqrt{n}|\ln \varepsilon|)$ Schritten liefern sie eine Näherung, die in einer ε-Umgebung des Optimums liegt. Bei geeigneter Wahl von ε liegt in dieser Umgebung nur eine Ecke des Simplex. Man projiziert daher die gefundene Näherung auf den Rand des Simplexes und erhält dann mit wenigen zusätzlichen Simplexschritten das Optimum. Die Grundidee des neuen Verfahrens ist, das Optimierungsproblem in ein einfaches äquivalentes *nichtlineares* Gleichungssystem umzuformen und hierauf das Newton-Verfahren anzuwenden.

Im Folgenden bezeichnen wir für Vektoren $x, s \in \mathbb{R}^n$ mit $X, S \in \mathbb{R}^{n\times n}$ die Diagonalmatrizen, deren Diagonalelemente die Komponenten der Vektoren x, s sind. Weiter ist stets

$$e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

Wir betrachten das LP (I.1.2) (S. 14) in Standardform

$$\min\{c^t x : Ax = b\ ,\ x \geq 0\}. \tag{I.7.1}$$

Gemäß Definition I.5.1 (S. 39) lautet das zughörige duale Programm (I.5.1) (S. 39)

$$\max\{b^t y : A^t y \leq c\}.$$

Für dieses LP führen wir Schlupfvariable $s \in \mathbb{R}^n$ ein und erhalten die äquivalente Form

$$\max\{b^t y : A^t y + s = c\ ,\ s \geq 0\}. \tag{I.7.2}$$

Gemäß Satz I.5.2 (S. 39) sind x^*, y^* Optima von (I.7.1) bzw. (I.7.2) genau dann, wenn gilt

$$c^t x^* = b^t y^* \quad \Longleftrightarrow \quad c^t x^* - b^t y^* = 0.$$

Für die zugehörigen Schlupfvariablen s^* bedeutet dies

$$x^{*t} s^* = x^{*t}(c - A^t y^*) = x^{*t} c - x^{*t} A^t y^* = c^t x^* - b^t y^* = 0.$$

Wegen $x^* \geq 0$ und $s^* \geq 0$ ist dies äquivalent zu

$$x_i^* s_i^* = 0$$

für alle i oder äquivalent

$$X^* s^* = S^* x^* = 0.$$

Dies beweist:

SATZ I.7.1. *Definiere die Funktion*

$$\Psi_0 : \mathbb{R}^n_+ \times \mathbb{R}^m \times \mathbb{R}^n_+ \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$$

durch

$$\Psi_0(x, y, s) = \begin{pmatrix} Ax - b \\ A^t y + s - c \\ Xs \end{pmatrix}.$$

Dann lösen x^* und (y^*, s^*) die LPs (I.7.1) und (I.7.2) genau dann, wenn gilt

$$\Psi_0(x^*, y^*, s^*) = 0.$$

Wir wollen das Newton-Verfahren auf das System

$$\Psi_0(x, y, s) = 0$$

anwenden. Dazu müssen wir sicherstellen, dass $D\Psi_0$ regulär ist.

LEMMA I.7.2. *Es ist*

$$D\Psi_0(x, y, s) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^t & I \\ S & 0 & X \end{pmatrix}.$$

Falls gilt

$$\operatorname{rang} A = m, \quad x > 0, \quad s > 0,$$

ist $D\Psi_0$ regulär.

BEWEIS. Die Darstellung von $D\Psi_0$ folgt sofort aus

$$Xs = Sx.$$

Zum Nachweis der Regularität betrachten wir (u, v, w) mit

$$D\Psi_0(x, y, s) \begin{pmatrix} u \\ v \\ w \end{pmatrix} = 0.$$

Dann gilt

$$Au = 0, \quad A^t v + w = 0, \quad Su + Xw = 0.$$

Wegen $s > 0$ ist S invertierbar, und wir erhalten

$$u = -S^{-1}Xw.$$

Aus den ersten beiden Gleichungen folgt

$$u^t w = -u^t A^t v = -(Au)^t v = 0$$

und somit

$$0 = w^t X S^{-1} w.$$

Wegen $x > 0$ ist auch X regulär und daher

$$w = 0$$

und somit

$$u = 0.$$

Außerdem folgt

$$A^t v = 0.$$

Da A maximalen Rang hat, folgt

$$v = 0.$$

Dies beweist die Regularität von $D\Psi_0$. □

Wegen Lemma I.7.2 machen wir im Folgenden die Voraussetzung

$$\begin{aligned} \overset{\circ}{\mathcal{P}} &= \{x : Ax = b\ ,\ x > 0\} \neq \emptyset, \\ \overset{\circ}{\mathcal{D}} &= \{(y,s) : A^t y + s = c\ ,\ s > 0\} \neq \emptyset. \end{aligned} \tag{I.7.3}$$

Wir werden später zeigen, wie wir diese Bedingung umgehen können.

Wegen Lemma I.7.2 können wir das Newton-Verfahren auf der Menge $\overset{\circ}{\mathcal{P}} \times \overset{\circ}{\mathcal{D}}$ auf Ψ_0 anwenden. Falls es konvergiert, liefert es eine Nullstelle (x^*, y^*, s^*) von Ψ_0. Wegen

$$X^* s^* = S^* x^* = 0$$

muss diese aber auf dem Rand von $\overset{\circ}{\mathcal{P}} \times \overset{\circ}{\mathcal{D}}$ liegen. Dies führt zu Konvergenzproblemen beim Newton-Verfahren. Um diese zu umgehen, „regularisieren“ wir Ψ_0. Dazu definieren wir für $\mu > 0$ die Funktion Ψ_μ durch

$$\Psi_\mu(x,y,s) = \begin{pmatrix} Ax - b \\ A^t y + s - c \\ Xs - \mu e \end{pmatrix}.$$

Wir wenden nun das Newton-Verfahren auf Ψ_μ an und passen in seinem Verlauf den Parameter μ so an, dass er gegen Null strebt.

LEMMA I.7.3. *Die Matrix $A \in \mathbb{R}^{m\times n}$, $m < n$, habe den maximalen Rang m; die Matrix D sei symmetrisch und regulär. Dann ist AD^2A^t invertierbar, und*

$$\Pi_R = DA^t(AD^2A^t)^{-1}AD$$

beschreibt die Orthogonalprojektion auf den Bildraum $\mathrm{R}(DA^t)$ *von* DA^t.

BEWEIS. Wegen der Symmetrie von D ist

$$AD^2A^t = AD^tDA^t = (DA^t)^tDA^t$$

symmetrisch.
Für $y \in \mathbb{R}^m$ folgt

$$\begin{aligned} & y^tAD^2A^ty = 0 \\ \iff \quad & DA^ty = 0 \\ \iff \quad & A^ty = 0 \\ \iff \quad & y = 0, \end{aligned}$$

da $\operatorname{rang}(A) = m$ ist.
Also ist AD^2A^t invertierbar und Π_R wohl definiert. Sei

$$y \in \mathrm{R}(DA^t), \quad \text{d.h.} \quad y = DA^tw$$

für ein $w \in \mathbb{R}^m$. Dann folgt

$$\Pi_R y = DA^t(AD^2A^t)^{-1}ADDA^tw = DA^tw = y.$$

Sei $z \in \mathrm{R}(DA^t)^\perp$. Dann ist

$$z \in \ker((DA^t)^t) = \ker(AD^t) = \ker(AD)$$

und es folgt

$$\Pi_R z = DA^t(AD^2A^t)^{-1}\underbrace{ADz}_{=0} = 0.$$

Also ist Π_R die Orthogonalprojektion auf $\mathrm{R}(DA^t)$. □

LEMMA I.7.4. *Folgende Voraussetzungen seien erfüllt:*

$$\mu > 0, \quad x > 0, \quad s > 0, \quad Ax = b, \quad A^ty + s = c.$$

Definiere die positive Diagonalmatrix D durch

$$D^2 = XS^{-1} = S^{-1}X$$

und die Projektionsmatrix Π_R durch

$$\Pi_R = DA^t(AD^2A^t)^{-1}AD.$$

Dann gelten für die Lösung $(\Delta x, \Delta y, \Delta s)$ des LGS

$$D\Psi_\mu(x,y,s)\begin{pmatrix}\Delta x\\ \Delta y\\ \Delta s\end{pmatrix} = -\Psi_\mu(x,y,s)$$

die Beziehungen

$$\begin{aligned} DA^t\Delta y &= \Pi_R q, \\ \Delta x &= -D(I - \Pi_R)q, \\ \Delta s &= -D^{-1}\Pi_R q \end{aligned}$$

mit

$$q = DX^{-1}r$$

und

$$r = Xs - \mu e.$$

BEWEIS. Aus den Voraussetzungen und der Definition von Ψ_μ folgt

$$\Psi_\mu(x, y, s) = \begin{pmatrix} 0 \\ 0 \\ r \end{pmatrix}$$

und

$$D\Psi_\mu(x, y, s) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^t & I \\ S & 0 & X \end{pmatrix}.$$

Damit lautet das LGS

$$\begin{aligned} A\Delta x &= 0, \\ A^t\Delta y + \Delta s &= 0, \\ S\Delta x + X\Delta s &= -r. \end{aligned}$$

Wir zeigen, dass die im Lemma angegebenen Größen Δx, Δy und Δs dieses LGS lösen.
Da $I - \Pi_R$ die Orthogonalprojektion auf

$$\ker((DA^t)^t) = \ker(AD)$$

ist, folgt sofort

$$A\Delta x = -AD(I - \Pi_R)q = 0.$$

Ebenso erhalten wir

$$\begin{aligned} 0 = \Pi_R q - \Pi_R q &= \Pi_R q - DD^{-1}\Pi_R q = DA^t\Delta y + D\Delta s \\ &= D(A^t\Delta y + \Delta s) \end{aligned}$$

und wegen der Regularität von D

$$A^t\Delta y + \Delta s = 0$$

Schließlich folgt

$$\begin{aligned} S\Delta x + X\Delta s &= -SD(I - \Pi_R)q - XD^{-1}\Pi_R q \\ &= -SDq + \underbrace{SD}_{\substack{=SD^2D^{-1} \\ =SS^{-1}XD^{-1} \\ =XD^{-1}}} \Pi_R q - XD^{-1}\Pi_R q \\ &= -SDq \\ &= -SDDX^{-1}r \\ &= -SS^{-1}XX^{-1}r \\ &= -r. \end{aligned}$$

Dies beweist die Behauptung. □

LEMMA I.7.5. *Die Voraussetzungen und Bezeichnungen seien wie in Lemma I.7.4. Zusätzlich gelte*

$$\|r\|_2 \leq \beta\mu$$

für ein $\beta \in [0, \frac{1}{2}]$. *Dann gilt*

$$\begin{aligned} A(x + \Delta x) &= b, \\ A^t(y + \Delta y) + (s + \Delta s) &= c, \\ x + \Delta x &\geq 0, \\ s + \Delta s &\geq 0. \end{aligned}$$

Für das neue Residuum

$$\widetilde{r} = (X + \Delta X)(s + \Delta s) - \mu e$$

gilt zudem

$$\|\widetilde{r}\|_2 \leq \beta^2\mu.$$

BEWEIS. Die Gültigkeit der Gleichungen

$$\begin{aligned} A(x + \Delta x) &= b, \\ A^t(y + \Delta y) + (s + \Delta s) &= c \end{aligned}$$

folgt aus dem Newton-Schritt.
Da Diagonalmatrizen kommutieren, erhalten wir für das neue Residuum mit Lemma I.7.4

$$\begin{aligned} \widetilde{r} &= \underbrace{Xs - \mu e}_{=r} + \underbrace{X\Delta s + \Delta X s}_{=-r} + \Delta X \Delta s \\ &= \Delta X \Delta s \\ &= D^{-1}\Delta X D \Delta s \\ &= \operatorname{diag}((I - \Pi_R)q)\Pi_R q. \end{aligned}$$

Bezeichnen wir mit θ den Winkel zwischen q und $(I - \Pi_R)q$, so folgt hieraus

$$\begin{aligned} \|\widetilde{r}\|_2 &\leq \|q\|_2^2 |\cos\theta \sin\theta| = \frac{1}{2}\|q\|_2^2 |\sin(2\theta)| \\ &\leq \frac{1}{2}\|q\|_2^2. \end{aligned}$$

Weiter ist

$$DX^{-1} = \sqrt{S^{-1}X}X^{-1} = \sqrt{S^{-1}X^{-1}} = \sqrt{SX}^{-1} = \sqrt{R + \mu I}^{-1}.$$

Damit folgt

$$\|DX^{-1}\|_2^2 = \|(R + \mu I)^{-1}\|_2 = \max_i \frac{1}{|r_i + \mu|}.$$

Nach Voraussetzung gilt für jedes i

$$|r_i| \leq \|r\|_2 \leq \beta\mu$$

und somit

$$|r_i + \mu| \geq \mu - |r_i| \geq \mu - \beta\mu = (1-\beta)\mu.$$

Damit erhalten wir

$$\|DX^{-1}\|_2^2 \leq \frac{1}{(1-\beta)\mu}$$

und somit

$$\|q\|_2^2 \leq \|DX^{-1}\|_2^2 \|r\|_2^2 \leq \frac{1}{(1-\beta)\mu}\|r\|_2^2 \leq \frac{\beta^2}{(1-\beta)}\mu.$$

Insgesamt liefert dies wegen $\beta \in [0, \frac{1}{2}]$

$$\|\widetilde{r}\|_2 \leq \frac{1}{2}\|q\|_2^2 \leq \frac{1}{2(1-\beta)}\beta^2\mu \leq \beta^2\mu.$$

Dies beweist die Behauptung für $\|\widetilde{r}\|_2$.
Weiter folgt aus unseren Abschätzungen und Lemma I.7.4

$$\begin{aligned}
\|X^{-1}\Delta x\|_2 &= \|X^{-1}D(I-\Pi_R)q\|_2 \\
&\leq \|X^{-1}D\|_2 \underbrace{\|(I-\Pi_R)q\|_2}_{\leq \|q\|_2} \\
&\leq \frac{1}{\sqrt{(1-\beta)\mu}} \frac{\beta}{\sqrt{1-\beta}}\sqrt{\mu} \\
&= \frac{\beta}{1-\beta} \\
&\leq 1
\end{aligned}$$

wegen $\beta \leq \frac{1}{2}$. Hieraus folgt $x + \Delta x \geq 0$.
Analog erhalten wir

$$\begin{aligned}
\|S^{-1}\Delta s\|_2 &= \| \underbrace{S^{-1}D^{-1}}_{\substack{=S^{-1}XX^{-1}D^{-1} \\ =D^2X^{-1}D^{-1} \\ =DX^{-1}}} \Pi_R q\|_2 \\
&\leq \|DX^{-1}\|_2 \|q\|_2 \\
&\leq 1.
\end{aligned}$$

Dies beweist $s + \Delta s \geq 0$. □

Lemma I.7.6. *Die Voraussetzungen und Bezeichnungen seien wie in Lemma I.7.5. Weiter sei*

$$\widetilde{\mu} = (1-\delta)\mu$$

und

$$\widetilde{\widetilde{r}} = (X + \Delta X)(s + \Delta s) - \widetilde{\mu}e.$$

Falls

$$\delta \leq \frac{2\beta(1-\beta)}{3\sqrt{n}}$$

ist, gilt

$$\|\widetilde{\widetilde{r}}\|_2 \leq \beta\widetilde{\mu}.$$

BEWEIS. Offensichtlich ist

$$\widetilde{\widetilde{r}} = (X + \Delta X)(s + \Delta s) - \mu e + (\mu - \widetilde{\mu})e = \widetilde{r} + \delta\mu e.$$

Damit folgt aus Lemma I.7.5

$$\|\widetilde{\widetilde{r}}\|_2 \leq \|\widetilde{r}\|_2 + \delta\mu\|e\|_2 \leq \beta^2\mu + \delta\sqrt{n}\mu.$$

Es gilt

$$\begin{aligned} & \beta^2\mu + \delta\sqrt{n}\mu \leq \beta\widetilde{\mu} \\ & \qquad = \beta(1-\delta)\mu \\ \iff \quad & \beta^2 + \delta\sqrt{n} \leq \beta(1-\delta) \\ \iff \quad & \delta(\sqrt{n} + \beta) \leq \beta(1-\beta). \end{aligned}$$

Wegen

$$\sqrt{n} + \beta \leq \sqrt{n} + \frac{1}{2} \leq \frac{3}{2}\sqrt{n}$$

ist die letzte Bedingung sicher erfüllt, wenn gilt

$$\frac{3}{2}\sqrt{n}\delta \leq \beta(1-\beta) \quad \iff \quad \delta \leq \frac{2\beta(1-\beta)}{3\sqrt{n}}.$$

Dies beweist die Aussage des Lemmas. □

Die Lemmata I.7.4, I.7.5 und I.7.6 führen auf Algorithmus I.7.1, der die eingangs beschriebene Idee realisiert.

Algorithmus I.7.1 Innere-Punkt-Verfahren

Gegeben: Toleranz $\varepsilon > 0$, Vektoren $x > 0$, y, $s > 0$, Zahl $\mu > 0$ mit $Ax = b$, $A^t y + s = c$, $\frac{1}{\mu}\|Xs - \mu e\|_2 \leq \frac{1}{2}$

Gesucht: Näherungslösung des LP (I.7.1)

1: **while** $\mu > \frac{\varepsilon}{n}$ **do**
2: $\quad r \leftarrow Xs - \mu e$, $D \leftarrow \sqrt{XS^{-1}}$, $q \leftarrow DX^{-1}r$
3: $\quad z \leftarrow$ Lösung von $(AD^2A^t)z = ADq$
4: $\quad \Delta y \leftarrow z$, $\Delta x \leftarrow -Dq + D^2A^tz$, $\Delta s \leftarrow -A^tz$
5: $\quad x \leftarrow x + \Delta x$, $y \leftarrow y + \Delta y$, $s \leftarrow s + \Delta s$
6: $\quad \mu \leftarrow \left(1 - \dfrac{1}{6\sqrt{n}}\right)\mu$
7: **end while**

BEMERKUNG I.7.7. Zeilen 2 – 5 von Algorithmus I.7.1 sind der Newton-Schritt

$$D\Psi_{\mu_k}(x_k, y_k, s_k)\begin{pmatrix}\Delta x_k \\ \Delta y_k \\ \Delta s_k\end{pmatrix} = -\Psi_{\mu_k}(x_k, y_k, s_k) = \begin{pmatrix}0 \\ 0 \\ -r_k\end{pmatrix}.$$

Die Matrix $AD_k^2A^t$ ist wegen der Voraussetzung $\operatorname{rang} A = m$ symmetrisch positiv definit. Daher wird das LGS in Zeile 3 am besten mit dem Cholesky-Verfahren gelöst. Der Aufwand zur Berechnung von z_k

ist daher $O(n^3)$. Die Berechnung der restlichen Größen erfordert nur Matrix-Vektor-Multiplikationen und damit $O(n^2)$ Operationen.

SATZ I.7.8. *Algorithmus I.7.1 bricht nach höchstens* $6\sqrt{n}\ln(\frac{n\mu_0}{\varepsilon})$ *Schritten ab. Er liefert dann zulässige Näherungslösungen* x *für das LP* (I.7.1) *und* y, s *für das duale Problem* (I.7.2) *mit*

$$x > 0, \quad s > 0, \quad c^t x - b^t y \leq 2\varepsilon.$$

BEWEIS. Wegen Zeile 6 gilt nach k Iterationen von Algorithmus I.7.1

$$\mu_k = \left(1 - \frac{1}{6\sqrt{n}}\right)^k \mu_0.$$

Also bricht Algorithmus I.7.1 nach k Iterationen ab, wenn gilt

$$\begin{aligned}
&\left(1 - \frac{1}{6\sqrt{n}}\right)^k \mu_0 \leq \frac{\varepsilon}{n} \\
\iff \quad &\left(1 - \frac{1}{6\sqrt{n}}\right)^k \leq \frac{\varepsilon}{n\mu_0} \\
\iff \quad &k \ln\left(1 - \frac{1}{6\sqrt{n}}\right) \leq \ln\left(\frac{\varepsilon}{n\mu_0}\right) \\
\iff \quad &k \geq \frac{\ln\left(\frac{\varepsilon}{n\mu_0}\right)}{\ln\left(1 - \frac{1}{6\sqrt{n}}\right)} \\
\iff \quad &k \geq \frac{\ln\left(\frac{n\mu_0}{\varepsilon}\right)}{\ln\left(\frac{1}{1-\frac{1}{6\sqrt{n}}}\right)}.
\end{aligned}$$

Aus dem Mittelwertsatz folgt

$$\begin{aligned}
\ln\left(\frac{1}{1-\frac{1}{6\sqrt{n}}}\right) &= -\ln\left(1 - \frac{1}{6\sqrt{n}}\right) \\
&= \frac{1}{6\sqrt{n}} \frac{1}{1-\eta} \quad \text{mit } 0 < \eta < \frac{1}{6\sqrt{n}} \\
&\geq \frac{1}{6\sqrt{n}}.
\end{aligned}$$

Also ist

$$\frac{\ln\left(\frac{n\mu_0}{\varepsilon}\right)}{\ln\left(\frac{1}{1-\frac{1}{6\sqrt{n}}}\right)} \leq 6\sqrt{n}\ln\left(\frac{n\mu_0}{\varepsilon}\right).$$

Dies beweist die Behauptung über die maximale Zahl der Iterationen von Algorithmus I.7.1.
Die Aussage über die Zulässigkeit von x bzw. y, s folgt aus der Konstruktion des Verfahrens.

Die Aussage $x > 0$, $s > 0$ folgt aus Lemma I.7.4.
Bei Abbruch des Verfahrens gilt

$$c^t x - b^t y = x^t s = e^t X s = e^t(r + \mu e) = e^t r + \mu e^t e = e^t r + \mu n$$
$$\leq \sqrt{n}\|r\|_2 + n\mu.$$

Wegen Lemma I.7.5 ist[1]

$$\|r\|_2 \leq \frac{1}{2}\mu.$$

Aus dem Abbruchkriterium folgt daher

$$c^t x - b^t y \leq \mu\left(\frac{1}{2}\sqrt{n} + n\right) \leq \varepsilon\left(\frac{1}{2\sqrt{n}} + 1\right) \leq 2\varepsilon.$$

Dies beweist die Behauptung. □

Zum Abschluss wollen wir zeigen, wie man die Annahme (I.7.3) verifiziert bzw. vermeidet und wie man Startvektoren x_0, y_0, s_0 für Algorithmus I.7.1 findet. Hierzu benötigen wir sog. *selbst-duale LPs*.

Gegeben sei eine *schief-symmetrische Matrix* C, d.h. eine Matrix mit

$$C^t = -C$$

und ein Vektor $a \geq 0$. Betrachte das LP

$$\min\{a^t x : Cx \geq -a\ ,\ x \geq 0\}.$$

Wegen $a \geq 0$ ist offensichtlich $x = 0$ zulässig und zugleich optimal. Durch Einführen von Schlupfvariablen s erhalten wir die Standardform

$$\min\{a^t x + 0^t s : Cx - s = -a\ ,\ x \geq 0\ ,\ s \geq 0\}.$$

Das zugehörige duale LP lautet wegen $C^t = -C$

$$\begin{aligned}
&\max\left\{-a^t y : \begin{pmatrix} C^t \\ -I \end{pmatrix} y \leq \begin{pmatrix} a \\ 0 \end{pmatrix}\right\} \\
&\quad = \max\{-a^t y : C^t y \leq a\ ,\ -y \leq 0\} \\
&\quad = \max\{-a^t y : -Cy \leq a\ ,\ y \geq 0\} \\
&\quad = \max\{-a^t y : Cy \geq -a\ ,\ y \geq 0\} \\
&\quad = -\min\{a^t y : Cy \geq -a\ ,\ y \geq 0\}.
\end{aligned}$$

Bis auf das Vorzeichen der Zielfunktion ist also das duale Programm identisch mit dem primalen. Daher nennt man derartige Programme *selbst-dual*.

Wie wir gesehen haben, ist $x^* = 0$ eine Optimallösung. Die zugehörige Schlupfvariable ist $s^* = a$. Offensichtlich gilt

$$x^* + s^* = a \geq 0.$$

Man kann nun folgende Verschärfung beweisen [2, Satz 4.8.2]:

[1] Beachte: Algorithmus I.7.1 entspricht dem Fall $\beta = \frac{1}{2}$ in den Lemmata I.7.5 und I.7.6.

LEMMA I.7.9. *Es sei C schief-symmetrisch und $a \geq 0$. Dann gibt es Vektoren x und s, die* strikt komplementär *sind, d.h. die die Bedingungen*

$$Cx - s = -a, \quad x \geq 0, \quad s \geq 0, \quad x + s > 0, \quad x^t s = 0$$

erfüllen.

Wir wollen nun einem LP

$$\min\{c^t x : \widetilde{A}x \geq \widetilde{b}\,,\ x \geq 0\} \tag{I.7.4}$$

und seinem zugehörigen dualen LP

$$\max\{\widetilde{b}^t y : \widetilde{A}^t y \leq c\,,\ y \geq 0\} \tag{I.7.5}$$

ein selbst-duales LP zuordnen, dessen Optimallösungen gemäss Lemma I.7.9 eine Auskunft über die Optimallösungen von (I.7.4) und (I.7.5) liefern. Dazu beachte man, dass unser ursprüngliches LP (I.7.1) die Form (I.7.4) hat mit

$$\widetilde{A} = \begin{pmatrix} A \\ -A \end{pmatrix}, \quad \widetilde{b} = \begin{pmatrix} b \\ -b \end{pmatrix}.$$

Wir betrachten beliebige positive Vektoren $x_0, u_0 \in \mathbb{R}^n$ und $y_0, s_0 \in \mathbb{R}^m$ und setzen

$$\begin{aligned} \overline{b} &= \widetilde{b} - \widetilde{A}x_0 + s_0, \\ \overline{c} &= c - \widetilde{A}^t y_0 - u_0, \\ \alpha &= c^t x_0 - \widetilde{b}^t y_0 + 1, \\ \beta &= \alpha + \overline{b}^t y_0 - \overline{c}^t x_0 + 1. \end{aligned}$$

Mit diesen Definitionen betrachten wir das folgende LP

$$\begin{aligned} \min\Big\{ \beta\theta : \ & \widetilde{A}x + \overline{b}\theta - \widetilde{b}\tau \geq 0, \\ & -\widetilde{A}^t y - \overline{c}\theta + c\tau \geq 0, \\ & -\overline{b}^t y + \overline{c}^t x - \alpha\tau \geq -\beta, \\ & \widetilde{b}^t y - c^t x + \alpha\theta \geq 0, \\ & y \geq 0\,,\ x \geq 0\,,\ \theta \geq 0\,,\ \tau \geq 0 \Big\}. \end{aligned} \tag{I.7.6}$$

Wie man leicht nachprüft, ist (I.7.6) selbst-dual. Gemäß Lemma I.7.9 besitzt (I.7.6) eine strikt komplementäre Optimallösung x^*, y^*, θ^*, τ^*. Da $\beta > 0$ und der Optimalwert 0 ist, muss

$$\theta^* = 0$$

gelten. Daher folgt aus der strikten Komplementarität

$$\begin{aligned} y^* + \widetilde{A}x^* - \widetilde{b}\tau^* &> 0, \\ x^* - \widetilde{A}^t y^* + c\tau^* &> 0, \\ -\overline{b}^t y^* + \overline{c}^t x^* - \alpha\tau^* + \beta &> 0, \\ \tau^* + \widetilde{b}^t y^* - c^t x^* &> 0. \end{aligned} \tag{I.7.7}$$

Wir müssen nun zwei Fälle unterscheiden:
Fall $\tau^* > 0$: Betrachte

$$x = \frac{1}{\tau^*}x^*, \quad y = \frac{1}{\tau^*}y^*$$

Wegen $\theta^* = 0$ folgt aus den ersten beiden Nebenbedingungen von (I.7.6), dass x und y für (I.7.4) bzw. (I.7.5) strikt zulässig sind. Die Optimalität folgt aus der letzten Nebenbedingung von (I.7.6). Die strikte Komplementarität folgt aus (I.7.7).
Fall $\tau^* = 0$: Jetzt ist

$$\widetilde{A}x^* \geq 0, \quad \widetilde{A}^t y^* \leq 0,$$

und wegen der letzten Ungleichung von (I.7.7)

$$\widetilde{b}^t y^* - c^t x^* > 0.$$

Dies bedeutet aber, dass (I.7.4) unzulässig ist, falls $\widetilde{b}^t y^* > 0$ ist, oder dass (I.7.5) unzulässig ist, falls $c^t x^* > 0$ ist.

Aus einer strikt komplementären Lösung von (I.7.6) kann man also entweder Optimallösungen von (I.7.4) und (I.7.5) konstruieren oder die Information gewinnen, dass mindestens eines der LP (I.7.4) oder (I.7.5) nicht zulässig ist. Da wir aber mit x_0, y_0 und $\theta_0 = \tau_0 = 1$ strikt zulässige Vektoren für (I.7.6) kennen, können wir Algorithmus I.7.1 mit diesen Startwerten auf (I.7.6) anwenden.

KAPITEL II

Diskrete Optimierung

II.1. Ganzzahlige Optimierung

Wir greifen die linearen Programme aus Kapitel I auf. Allerdings machen wir jetzt die wesentliche Einschränkung, dass die Komponenten des Lösungsvektors ganzzahlig sein sollen. Sie ist dadurch motiviert, dass in vielen Anwendungen nur ganzzahlige Einheiten einen Sinn machen. So ist es in unserem allerersten Beispiel 1 (S. 5) sicher nicht sinnvoll, ein $\frac{1}{2}$ oder $\frac{2}{3}$ Paar Schuhe zu produzieren.

Zur Verdeutlichung der Konsequenzen unserer neuen Einschränkung greifen wir Beispiel 1 erneut auf.

Beispiel II.1.1. Beispiel 1 lautete

$$\max\Big\{16x + 32y : 6x + 15y \leq 4500,\ 4x + 5y \leq 2000,\\ 20x + 10y \leq 8000,\ x \geq 0, y \geq 0\Big\}.$$

Hierbei bezeichnet x, y die Menge der produzierten Damen- bzw. Herrenschuhe. Wir fordern nun zusätzlich $x \in \mathbb{Z}$, $y \in \mathbb{Z}$. Bezeichne mit

$$\mathcal{P}_{\mathbb{R}} = \Big\{(x, y) \in \mathbb{R}^2 : 6x + 15y \leq 4500,\ 4x + 5y \leq 2000,\\ 20x + 10y \leq 8000,\ x \geq 0,\ y \geq 0\Big\}$$

und

$$\mathcal{P}_{\mathbb{Z}} = \Big\{(x, y) \in \mathbb{Z}^2 : 6x + 15y \leq 4500,\ 4x + 5y \leq 2000,\\ 20x + 10y \leq 8000,\ x \geq 0,\ y \geq 0\Big\}$$

die zulässigen Mengen des alten und neuen Problems. Dann ist natürlich $\mathcal{P}_{\mathbb{Z}} \subset \mathcal{P}_{\mathbb{R}}$ und daher

$$\max\{16x + 32y : (x, y) \in \mathcal{P}_{\mathbb{R}}\} \geq \max\{16x + 32y : (x, y) \in \mathcal{P}_{\mathbb{Z}}\}.$$

Wie wir in Beispiel 1 (S. 5) gesehen haben, wird der Optimalwert für $\mathcal{P}_{\mathbb{R}}$ in $(250, 200)$ (Punkt B in Abbildungen 1 (S. 6) und II.1.1) angenommen. Da dieser Punkt in $\mathcal{P}_{\mathbb{Z}}$ liegt, ist dies auch das Optimum für $\mathcal{P}_{\mathbb{Z}}$ und nichts ändert sich.

Jetzt ändern wir die Gewinne für Damen- und Herrenschuhe auf 27 € bzw. 21 € , so dass die Zielfunktion $27x + 21y$ ist. Der Wert der neuen Zielfunktion im Punkt $(250, 200)$ ist 10950 € . Im Punkt $A = (\frac{1000}{3}, \frac{400}{3})$

beträgt er 11800 € . In allen anderen Ecken von $\mathcal{P}_{\mathbb{R}}$ ist er niedriger. Also ist das Optimum für $\mathcal{P}_{\mathbb{R}}$ jetzt $(\frac{1000}{3}, \frac{400}{3})$. Dieser Punkt liegt aber nicht in $\mathcal{P}_{\mathbb{Z}}$.
Was tun? Runden wir, erhalten wir $\widetilde{A} = (333, 133)$. Dieser Punkt ist für $\mathcal{P}_{\mathbb{Z}}$ zulässig (Das ist ein glücklicher Zufall!) und liefert den Gewinn 11784 € . Aber ist dies wirklich das Optimum?
Reduzieren wir y um 1 und erhöhen x um 1, steigern wir den Gewinn. Dies liefert den Punkt $\widehat{A} = (334, 132)$. Dieser Punkt liegt in $\mathcal{P}_{\mathbb{Z}}$ und liefert einen besseren Wert!
Wiederholten wir diese Operation, verließen wir den zulässigen Bereich. Das sichert aber nicht die Optimalität von $\widehat{A}$. Damit bleibt die Frage nach dem Optimum für $\mathcal{P}_{\mathbb{Z}}$ vorerst offen.

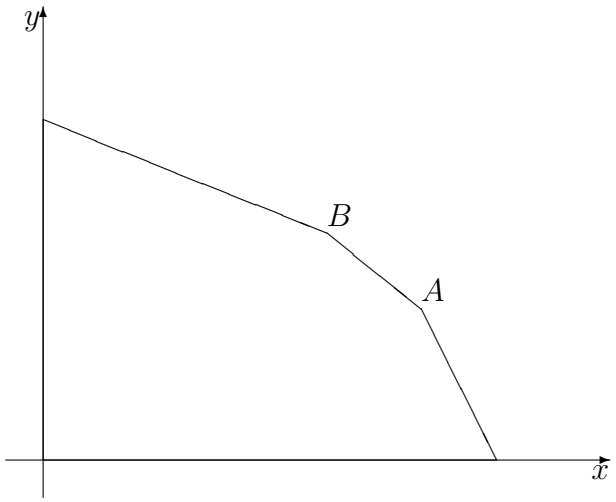

ABBILDUNG II.1.1. Zulässiger Bereich $\mathcal{P}_{\mathbb{R}}$ für das Optimierungsproblem aus Beispiel II.1.1

In Anlehnung an §I.1 definieren wir (vgl. Definitionen I.1.2 (S. 13) und I.1.5 (S. 14)):

DEFINITION II.1.2. Gegeben seien ein Vektor $c \in \mathbb{Z}^n$, eine Matrix $A \in \mathbb{Z}^{m\times n}$, Vektoren $\underline{b}, \overline{b} \in [\mathbb{Z} \cup \{-\infty, \infty\}]^m$ und Vektoren $\ell, u \in [\mathbb{Z} \cup \{-\infty, \infty\}]^n$. Dann nennt man die Aufgabe

$$\min\{c^t x : \underline{b} \le Ax \le \overline{b}\ ,\ \ell \le x \le u\ ,\ x \in \mathbb{Z}^n\} \tag{II.1.1}$$

ein *ganzzahliges lineares Optimierungsproblem* kurz *GLP*. Die Menge

$$\mathcal{P}_{\mathbb{Z}} = \{x \in \mathbb{Z}^n : \underline{b} \le Ax \le \overline{b}\ ,\ \ell \le x \le u\}$$

heißt der *Zulässigkeitsbereich* des GLP. Ist speziell

$$\underline{b} = \overline{b} \in \mathbb{Z}^m, \quad \ell = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \quad u = \begin{pmatrix} \infty \\ \vdots \\ \infty \end{pmatrix},$$

sprechen wir von einem *ganzzahligen linearen Optimierungsproblem in Standardform* oder kurz *GP*. Es hat die Form

$$\min\{c^t x : Ax = b\ ,\ x \geq 0\ ,\ x \in \mathbb{Z}^n\}. \tag{II.1.2}$$

Bemerkung II.1.3. (1) Durch Einführen von Schlupfvariablen und geeigneter Zerlegung von x kann wie in Bemerkung I.1.6 (S. 14) jedes GLP in ein äquivalentes GP überführt werden. Gleiches gilt für die Transformation von Maximierungsproblemen.
(2) Die Voraussetzung, dass die Daten c, A, $\underline{b}$, $\overline{b}$, ℓ, u bzw. c, A, b ganzzahlig sein sollen, ist für die Praxis keine Einschränkung. Denn dort sind die Daten stets mindestens rational und können durch Multiplikation mit dem kgV aller Nenner auf ganzzahlige Daten transformiert werden.

Wir wollen die Lösung ganzahliger Optimierungsprobleme auf die Lösung linearer Optimierungsprobleme, wie wir sie in Kapitel I betrachtet haben, reduzieren. Dazu benötigen wir einige zusätzliche Notationen.

Definition II.1.4. Gegeben sei ein GP in Standardform (II.1.2) mit Zulässigkeitsbereich

$$\mathcal{P}_{\mathbb{Z}} = \{x \in \mathbb{Z}^n : Ax = b\ ,\ x \geq 0\}.$$

Dann heißt das LP in Standardform

$$\min\{c^t x : x \in \mathbb{R}^n\ ,\ Ax = b\ ,\ x \geq 0\} \tag{II.1.3}$$

mit Zulässigkeitsbereich

$$\mathcal{P}_{\mathbb{R}} = \{x \in \mathbb{R}^n : Ax = b\ ,\ x \geq 0\}.$$

das zugehörige *relaxierte LP*.

Bemerkung II.1.5. Betrachte ein GP (II.1.2) und das zugehörige relaxierte LP (II.1.3). Dann gilt offensichtlich:

(1) $\mathcal{P}_{\mathbb{Z}} \subset \mathcal{P}_{\mathbb{R}}$.
(2) $\min\{c^t x : x \in \mathcal{P}_{\mathbb{R}}\} \leq \min\{c^t x : x \in \mathcal{P}_{\mathbb{Z}}\}$.
(3) Löst x^* Problem (II.1.3) und ist $x^* \in \mathbb{Z}^n$, so löst x^* auch das Problem (II.1.2).

Definition II.1.6. (1) Die Menge $M \subset \mathbb{Z}^n$ werde durch lineare Ungleichungen beschrieben. Dann ist $\overline{M} \subset \mathbb{R}^n$ die Menge, die durch dieselben Ungleichungen ohne die Ganzzahligkeitsbedingung beschrieben wird.
(2) Für $x \in \mathbb{R}$ sei

$$\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}, \quad \lceil x \rceil = \min\{z \in \mathbb{Z} : z \geq x\}.$$

Bemerkung II.1.7. Sind $\mathcal{P}_{\mathbb{Z}}$ und $\mathcal{P}_{\mathbb{R}}$ der Zulässigkeitsbereich eines GP bzw. des zugehörigen relaxierten LP, so ist $\mathcal{P}_{\mathbb{R}} = \overline{\mathcal{P}_{\mathbb{Z}}}$.

Algorithmus II.1.1 leistet das Gewünschte.

Algorithmus II.1.1 Branch and Bound Algorithmus von Dakin

Gegeben: $c \in \mathbb{Z}^n$, $A \in \mathbb{Z}^{m\times n}$, $b \in \mathbb{Z}^m$
Gesucht: Lösung des GP $\min\{c^t x : x \in \mathcal{P}_\mathbb{Z}\}$ mit $\mathcal{P}_\mathbb{Z} = \{x \in \mathbb{Z}^n : Ax = b\ ,\ x \geq 0\}$

1: $\ell \leftarrow \infty$, $\mathcal{K} \leftarrow \{\mathcal{P}_\mathbb{Z}\}$
2: **while** $\mathcal{K} \neq \emptyset$ **do**
3: $\quad M \leftarrow$ Element in $\mathcal{K}$ ▷ Branching
4: $\quad x \leftarrow$ Lösung des relaxierten LP zu $\overline{M}$
5: $\quad$ **if** relaxierte LP nicht lösbar **then**
6: $\quad\quad \mathcal{K} \leftarrow \mathcal{K} \setminus M$
7: $\quad$ **else**
8: $\quad\quad d \leftarrow c^t x$ ▷ Bounding
9: $\quad\quad$ **if** $d \geq \ell$ **then**
10: $\quad\quad\quad \mathcal{K} \leftarrow \mathcal{K} \setminus M$
11: $\quad\quad$ **else**
12: $\quad\quad\quad$ **if** $x \in \mathbb{Z}^n$ **then**
13: $\quad\quad\quad\quad \ell \leftarrow d$, $\mathcal{K} \leftarrow \mathcal{K} \setminus M$
14: $\quad\quad\quad$ **else**
15: $\quad\quad\quad\quad i \leftarrow$ Index mit $x_i \notin \mathbb{Z}$
16: $\quad\quad\quad\quad M^- \leftarrow M \cap \{x \in \mathbb{Z}^n : x_i \leq \lfloor x_i^* \rfloor\}$
17: $\quad\quad\quad\quad M^+ \leftarrow M \cap \{x \in \mathbb{Z}^n : x_i \geq \lceil x_i^* \rceil\}$
18: $\quad\quad\quad\quad \mathcal{K} \leftarrow (\mathcal{K} \setminus M) \cup M^- \cup M^+$
19: $\quad\quad\quad$ **end if**
20: $\quad\quad$ **end if**
21: $\quad$ **end if**
22: **end while**

▷ $\ell = \infty$: GP nicht lösbar
▷ $\ell < \infty$: GP lösbar, ℓ Optimalwert, x Optimallösung

Beispiel II.1.8. Wir wenden Algorithmus II.1.1 auf Beispiel II.1.1 an. Man beachte, dass wir hierzu die Zielfunktion mit -1 multiplizieren müssen, um aus dem Maximierungsproblem ein Minimierungsproblem zu machen.
Beim Start ist $\ell = \infty$ und $\mathcal{K} = \{\mathcal{P}_\mathbb{Z}\}$. Wir wählen in Zeile 3 $M = \mathcal{P}_\mathbb{Z}$ und kommen zu Zeilen 8 – 18 mit

$$x^* = \begin{pmatrix} \frac{1000}{3} \\ \frac{400}{3} \end{pmatrix}, \quad c^* = -11800.$$

Damit greifen Zeilen 15 – 18. Wir wählen $i = 1$[1] und erhalten

$$M^- = \{x \in \mathcal{P}_\mathbb{Z} : x_1 \leq 333\} \quad M^+ = \{x \in \mathcal{P}_\mathbb{Z} : x_1 \geq 334\}.$$

Damit ist der erste Durchgang abgeschlossen und $\mathcal{K} = \{M^-, M^+\}$ (vgl. Abbildung II.1.2).

[1] Die Wahl $i = 2$ wäre genauso gut möglich.

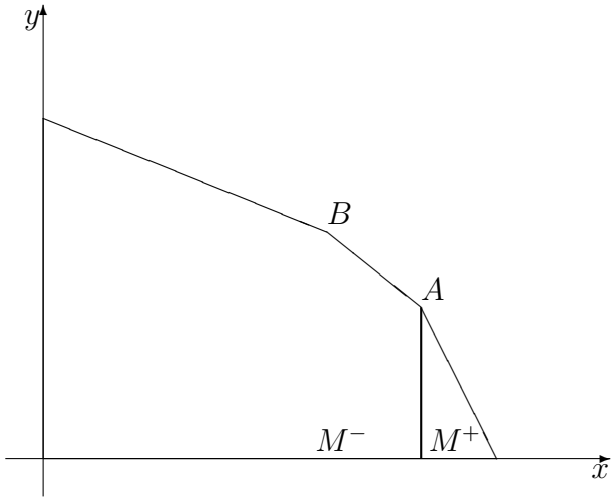

ABBILDUNG II.1.2. Aufspaltung des Zulässigkeitsbereiches in Beispiel II.1.8 nach dem ersten Durchlauf von Algorithmus II.1.1

Wir kommen wieder zu Zeile 3 und wählen $M = M^+$ aus.[2] M^+ hat die Ecken

$$\begin{pmatrix}334\\0\end{pmatrix}, \quad \begin{pmatrix}400\\0\end{pmatrix}, \quad \begin{pmatrix}334\\132\end{pmatrix}$$

mit den Funktionswerten

$$-9018\ ,\ -10800\ ,\ -11790.$$

Also ist

$$c^* = -11790, \quad x^* = \begin{pmatrix}334\\132\end{pmatrix} \in \mathcal{P}_{\mathbb{Z}}.$$

Damit kommen wir zu Zeile 13, setzen

$$\ell = -11790, \quad x = \begin{pmatrix}334\\132\end{pmatrix} \in \mathcal{P}_{\mathbb{Z}}$$

und entfernen M^+ aus $\mathcal{K}$.
Wir kommen wieder zu Zeile 3. Jetzt ist $\mathcal{K} = \{M^-\}$ und wir wählen $M = M^-$ aus. M^- hat die Ecken

$$\begin{pmatrix}250\\200\end{pmatrix}, \quad \begin{pmatrix}0\\300\end{pmatrix}, \quad \begin{pmatrix}0\\0\end{pmatrix}, \quad \begin{pmatrix}332\\0\end{pmatrix}, \quad \begin{pmatrix}332\\134.4\end{pmatrix}$$

mit den Funktionswerten

$$-10950\ ,\ -6100\ ,\ -0\ ,\ -8964\ ,\ -11786.4.$$

Also ist

$$x^* = \begin{pmatrix}332\\134.4\end{pmatrix} \notin \mathcal{P}_{\mathbb{Z}}, \quad c^* = -11768.4 > -11790 = \ell.$$

[2] Genauso gut hätten wir erst $M = M^-$ auswählen können.

Damit greift Zeile 10 und M^- wird aus $\mathcal{K}$ entfernt.
Da nun $\mathcal{K} = \emptyset$ ist, sind wir fertig und erhalten den Optimalwert -11790 an der Stelle $(334, 132)$.

Wir wollen eine Alternative zu Algorithmus II.1.1 angeben. Dazu brauchen wir folgendes technisches Hilfsergebnis.

LEMMA II.1.9. *Betrachte ein GP (II.1.2) und das zugehörige relaxierte LP (II.1.3). Für die Optimallösung x^* des relaxierten Problems gelte $x^* \notin \mathcal{P}_\mathbb{Z}$. Bezeichne mit J die zu x^* gehörige Basismenge und mit K das entsprechende Komplement. Setze*

$$\overline{A} = A_J^{-1} A_K$$

und wähle ein $i \in J$ mit $x_i^* \notin \mathbb{Z}$. Definiere

$$r = \lfloor x_i^* \rfloor - x_i^*$$

und den Vektor $d \in \mathbb{R}^n$ durch

$$d_j = \begin{cases} \lfloor \overline{A}_{ij} \rfloor - \overline{A}_{ij} & \text{für } j \in K, \\ 0 & \text{für } j \notin K. \end{cases}$$

Dann ist

$$d^t x^* > r$$

und für alle $x \in \mathcal{P}_\mathbb{Z}$ gilt

$$d^t x \leq r.$$

D.h., die Hyperebene $\{d^t x = r\}$ trennt $\mathcal{P}_\mathbb{Z}$ und x^*.

BEWEIS. Konstruktionsgemäß ist

$$x_J^* = A_J^{-1} b, \quad x_K^* = 0.$$

Für alle $x \in \mathcal{P}_\mathbb{Z}$ gilt

$$A_J x_J + A_K x_K = Ax = b$$

und somit

$$x_J + \overline{A} x_K = x_J + A_J^{-1} A_K x_K = A_J^{-1} b = x_J^*. \tag{II.1.4}$$

Insbesondere ist für alle $x \in \mathcal{P}_\mathbb{Z}$

$$x_i + \sum_{j \in K} \overline{A}_{ij} x_j = x_i^*. \tag{II.1.5}$$

Hieraus folgt durch Abrunden

$$x_i + \sum_{j \in K} \lfloor \overline{A}_{ij} \rfloor x_j \leq x_i^*.$$

Da die linke Seite ganzzahlig und $x_i^* \notin \mathbb{Z}$ ist, gilt so gar

$$x_i + \sum_{j \in K} \lfloor \overline{A}_{ij} \rfloor x_j \leq \lfloor x_i^* \rfloor.$$

Subtrahieren wir hiervon die Gleichung (II.1.5), erhalten wir für alle $x \in \mathcal{P}_{\mathbb{Z}}$

$$\sum_{j \in K} \left(\lfloor \overline{A}_{ij} \rfloor - \overline{A}_{ij} \right) x_j \leq \lfloor x_i^* \rfloor - x_i^*$$

bzw.

$$d^t x \leq r.$$

Andererseits ist

$$d^t x^* = 0$$

wegen $x_j^* = 0$ für $j \in K$ und

$$r < 0$$

wegen $x_i^* \notin \mathbb{Z}$. Dies beweist die Behauptung. □

Bemerkung II.1.10. Die Bezeichnungen seien wie in Lemma II.1.9. Zusätzlich gelte, dass x^* nicht in der konvexen Hülle von $\mathcal{P}_{\mathbb{Z}}$ liegt. Dann folgt aus Satz III.2.16 (S. 124), dass x^* und die konvexe Hülle von $\mathcal{P}_{\mathbb{Z}}$, also insbesondere x^* und $\mathcal{P}_{\mathbb{Z}}$, eigentlich getrennt werden können. Lemma II.1.9 geht aber wesentlich weiter:

- Es zeigt, dass die Trennung so gar strikt ist.
- Es ist konstruktiv.
- Aus $x^* \notin \mathcal{P}_{\mathbb{Z}}$ folgt nicht, dass x^* nicht in der konvexen Hülle von $\mathcal{P}_{\mathbb{Z}}$ liegt.

Diese Überlegungen führen auf Algorithmus II.1.2.

Algorithmus II.1.2 Schnittebenen-Verfahren

Gegeben: $c \in \mathbb{Z}^n$, $A \in \mathbb{Z}^{m \times n}$, $b \in \mathbb{Z}^m$
Gesucht: Lösung des GP $\min\{c^t x : x \in \mathcal{P}_{\mathbb{Z}}\}$ mit $\mathcal{P}_{\mathbb{Z}} = \{x \in \mathbb{Z}^n : Ax = b\ ,\ x \geq 0\}$

1: $M \leftarrow \mathcal{P}_{\mathbb{Z}}$
2: **loop** ▷ wird in Zeilen 5 oder 8 beendet
3: $x \leftarrow$ Lösung des relaxierten LP zu $\overline{M}$, $J \leftarrow$ zugehörige Basis
4: **if** relaxierte LP nicht lösbar **then**
5: **stop** ▷ GP nicht lösbar
6: **end if**
7: **if** $x \in \mathbb{Z}^n$ **then**
8: **stop** ▷ x ist Optimallösung
9: **end if**
10: $i \leftarrow$ Index in J mit $x_i \notin \mathbb{Z}$, $r \leftarrow \lfloor x_i \rfloor - x_i$
11: **for** $j \notin J$ **do** $d_j \leftarrow \lfloor \overline{A}_{ij} \rfloor - \overline{A}_{ij}$
12: **end for**
13: **for** $j \in J$ **do** $d_j \leftarrow 0$
14: **end for**
15: $M \leftarrow M \cap \{y \in \mathbb{Z}^n : d^t y \leq r\}$
16: **end loop**

BEMERKUNG II.1.11. In Zeile 15 von Algorithmus II.1.2 wird eine Ungleichung hinzugefügt, die das aktuelle x ausschließt, aber die Menge $\mathcal{P}_\mathbb{Z}$ nicht beeinflusst. Daher wird Zeile 3 für das neue M ein anderes Optimum liefern. Da Zeile 15 die Menge $\mathcal{P}_\mathbb{Z}$ nicht beeinflusst, liefert Algorithmus II.1.2 entweder die Information, dass das GP nicht lösbar ist, oder gibt das exakte Optimum des GP zurück.

BEISPIEL II.1.12. Wir wenden Algorithmus II.1.2 auf Beispiel II.1.1 an. Zu Beginn ist $M = \mathcal{P}_\mathbb{Z}$. Zeile 3 liefert

$$x = \begin{pmatrix} \frac{1000}{3} \\ \frac{400}{3} \end{pmatrix}.$$

Es ist[3]

$$J = \{1,2,3\}, \quad K = \{4,5\}, \quad A = \begin{pmatrix} 6 & 15 & 1 & 0 & 0 \\ 4 & 5 & 0 & 1 & 0 \\ 20 & 10 & 0 & 0 & 1 \end{pmatrix}$$

und

$$\overline{A}_K = \begin{pmatrix} -\frac{1}{6} & \frac{1}{12} \\ \frac{1}{3} & -\frac{1}{15} \\ -4 & \frac{1}{2} \end{pmatrix}.$$

Wir wählen $i = 1$[4] und erhalten in Lemma II.1.9

$$r = -\frac{1}{3}, \quad d = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\frac{5}{6} \\ -\frac{1}{12} \end{pmatrix}.$$

Damit lautet die neue Ungleichung

$$d^t \begin{pmatrix} x \\ y \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} = -\frac{5}{6}s_2 - \frac{1}{12}s_3 \leq -\frac{1}{3}.$$

Wir wollen diese Zusatzbedingung für das ursprüngliche System ohne Schlupfvariablen übernehmen. Dazu beachten wir, dass

$$s_2 = 2000 - 4x - 5y, \quad s_3 = 8000 - 20x - 10y$$

[3]Beachte, dass wir Schlupfvariablen für die drei Ungleichungen einführen müssen und dass die erste Ungleichung für x^* strikt ist.

[4]Die Wahl $i = 2$ wäre genau so gut möglich. Die Wahl $i = 3$ ist nicht möglich, da die Schlupfvariable für die erste Ungleichung ganzzahlig ist.

ist und setzen diese Gleichungen in die neue Ungleichung ein:

$$\begin{aligned} -\frac{1}{3} &\geq -\frac{5}{6}[2000 - 4x - 5y] - \frac{1}{12}[8000 - 20x - 10y] \\ &= -\frac{7000}{3} + 5x + 5y \\ \iff \quad 5x + 5y &\leq \frac{7000 - 1}{3} \\ &= 2333. \end{aligned}$$

Wir erhalten also die zusätzliche Nebenbedingung

$$5x + 5y \leq 2333.$$

Die neue Menge $\mathcal{P}_{\mathbb{R}}$ ist in Abbildung II.1.3 skizziert. Die neue Menge $\mathcal{P}_{\mathbb{Z}}$ hat die Ecken

$$\begin{pmatrix} 250 \\ 200 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 300 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 400 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 333.4 \\ 133.2 \end{pmatrix}, \quad \begin{pmatrix} 333 \\ 133.6 \end{pmatrix}.$$

Der Optimalwert ist -11799 und wird in $(333.4, 133.2)$ angenommen. Mit diesem Wert müssen wir dann wieder in Zeilen 10 – 15 von Algorithmus II.1.2 einsteigen.

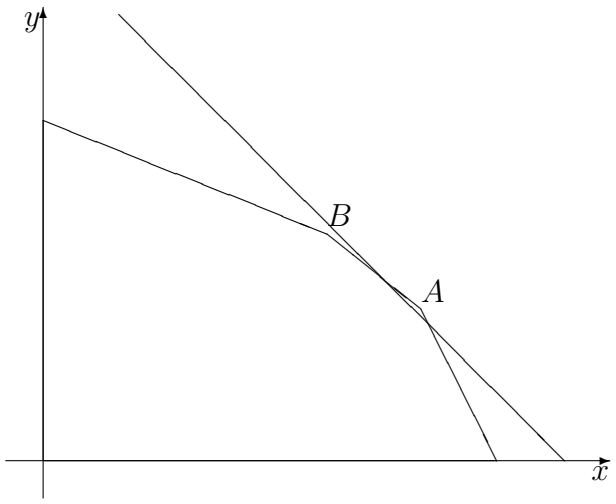

ABBILDUNG II.1.3. Zulässiger Bereich $\mathcal{P}_{\mathbb{R}}$ für das Optimierungsproblem aus Beispiel II.1.1 und II.1.12 mit zusätzlicher Bedingung $5x + 5y \leq 2333$

II.2. Grundzüge der Graphentheorie

DEFINITION II.2.1. Ein *(gerichteter) Graph* oder auch *Digraph* ist ein Tupel $G = (V, E)$ bestehend aus einer endlichen *Knotenmenge* V und einer endlichen *Kantenmenge* $E \subset V \times V$. Die Elemente von V heißen *Knoten*, diejenigen von E *Kanten*. Für jede Kante $e = (u, v)$

wird $u \neq v$ verlangt; u heißt *Anfangsknoten* von e, v der *Endknoten*; u heißt *Vorgänger* von v, v heißt *Nachfolger* von u.

BEMERKUNG II.2.2. Es werden gelegentlich *ungerichtete Graphen* betrachtet. Bei diesen wird zwischen den Kanten (u, v) und (v, u) nicht unterschieden. Aus jedem ungerichteten Graphen kann man einen gerichteten Graphen konstruieren, indem man für jedes ungeordnete Tupel $\{u, v\}$ von Knoten die Kanten (u, v) und (v, u) in die Kantenmenge aufnimmt. Wir beschränken uns daher im Folgenden auf gerichtete Graphen.

DEFINITION II.2.3. Wir ordnen jedem Graphen $G = (V, E)$ zwei Abbildungen $\alpha : E \to V$, $\omega : E \to V$ und zwei Abbildungen $K^+ : V \to \mathcal{P}(E)$, $K^- : V \to \mathcal{P}(E)$ wie folgt zu:

$\alpha(e)$**:** beschreibt den Anfangsknoten der Kante e, d.h.

$$\alpha((u, v)) = u.$$

$\omega(e)$**:** beschreibt den Endpunkt der Kante e, d.h.

$$\omega((u, v)) = v.$$

$K^-(v)$**:** beschreibt die Menge aller Kanten mit Anfangsknoten v, d.h.

$$K^-(v) = \{e \in E : \alpha(e) = v\}.$$

$K^+(v)$**:** beschreibt die Menge aller Kanten mit Endknoten v, d.h.

$$K^+(v) = \{e \in E : \omega(e) = v\}.$$

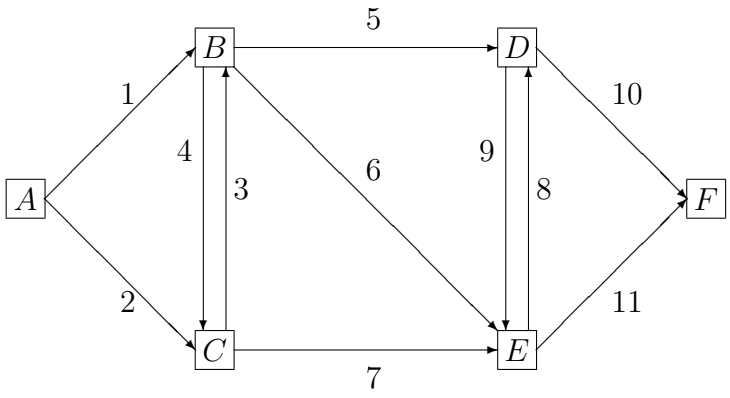

ABBILDUNG II.2.1. Graph der Beispiele II.2.4, II.2.6, II.2.10, II.2.12 und II.2.15 mit Nummerierung der Kanten

Beispiel II.2.4. Für den Graphen aus Abbildung II.2.1 erhalten wir:

$$\begin{aligned} V &= \{A, B, C, D, E, F\}, \\ E &= \{(A,B),(A,C),(B,C),(C,B),(B,D),(B,E), \\ &\qquad (C,E),(D,E),(E,D),(D,F),(E,F)\} \\ K^-(A) &= \{(A,B),(A,C)\}, \\ K^+(A) &= \emptyset, \\ K^-(F) &= \emptyset, \\ K^+(F) &= \{(D,F),(E,F)\}. \end{aligned}$$

Definition II.2.5. Wir ordnen jedem Graphen $G = (V, E)$ zwei Zahlen $n = \sharp V$ und $m = \sharp E$ sowie zwei Matrizen $A \in \mathbb{R}^{n\times m}$ und $B \in \mathbb{R}^{n\times n}$ wie folgt zu:

$$A_{ve} = \begin{cases} 1 & \text{falls } \omega(e) = v, \\ -1 & \text{falls } \alpha(e) = v, , \\ 0 & \text{sonst,} \end{cases}$$

$$B_{uv} = \begin{cases} 1 & \text{falls } (u,v) \in E, \\ 0 & \text{sonst.} \end{cases}$$

A heißt die *Inzidenzmatrix* des Graphen, B heißt die *Adjazenzmatrix* des Graphen.

Beispiel II.2.6. Wir setzen Beispiel II.2.4 fort und nummerieren die Kanten des Graphen wie in Abbildung II.2.1 (S. 72) angegeben. Die Knoten werden lexikographisch nummeriert. Dann ist

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

$$B = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Bemerkung II.2.7. Seien A und B die Inzidenz- bzw. Adjazenzmatrix eines Graphen. Dann gilt:

(1) Die Spaltensumme jeder Spalte von A ist 0, d.h.

$$e^t A = 0$$

für $e = (1, \ldots, 1)^t \in \mathbb{R}^n$ mit $n = \sharp V$.

(2) Die Diagonalelemente von B sind alle gleich 0.

(3) Eine Zeile von B ist genau dann gleich 0, wenn für den zugehörigen Knoten u gilt

$$K^-(u) = \emptyset.$$

(4) Eine Spalte von B ist genau dann gleich 0, wenn für den zugehörigen Knoten v gilt

$$K^+(v) = \emptyset.$$

(5) Die Zeilensumme der zu dem Knoten u gehörenden Zeile von B ist $\sharp K^-(u)$.

DEFINITION II.2.8. Sei $G = (V, E)$ ein Graph mit zugehöriger Inzidenzmatrix A und zugehöriger Adjazenzmatrix B.

(1) Die Zahl

$$\deg(v) = \sharp K^-(v) + \sharp K^+(v) = \sum_{e \in E} |A_{ve}|$$

heißt der *Grad* des Knotens v.

(2) Zwei Knoten u und v heißen *benachbart*, wenn (u, v) oder (v, u) in E enthalten ist, d.h. wenn

$$B_{uv} + B_{vu} > 0$$

ist.

(3) Die Menge

$$\Gamma(u) = \{v \in V : u, v \text{ sind benachbart}\}$$

ist die Menge der *Nachbarknoten* des Knotens u.

(4) Ist $W \subset V$, so ist

$$\Gamma(W) = \bigcup_{u \in W} \Gamma(u)$$

die Menge der *Nachbarknoten* von W.

(5) Ein Knoten u mit $\Gamma(u) = \emptyset$ bzw. äquivalent

$$\sum_{v \in V} (B_{uv} + B_{vu}) = 0$$

heißt *isoliert*.

(6) Ist $W \subset V$, so heißt

$$\delta(W) = \{e \in E : \alpha(e) \in W \ , \ \omega(e) \notin W\}$$

der zugehörige *Schnitt*. Ist $u \in W$ und $v \notin W$, so nennt man $\delta(W)$ einen u und v *trennenden Schnitt*.

BEMERKUNG II.2.9. Ist $u \in V$, so ist $\delta(\{u\}) = K^-(u)$.

Beispiel II.2.10. In Beispiel II.2.4 (vgl. Abbildung II.2.1 (S. 72)) ist

$$\begin{aligned} \deg(A) &= 2, & \Gamma(A) &= \{B, C\}, \\ \deg(B) &= 5, & \Gamma(B) &= \{A, C, D, E\}, \\ \delta(\{B, C\}) &= \{(B, D), (B, E), (C, E)\}. \end{aligned}$$

Definition II.2.11. Sei $G = (V, E)$ ein Graph.

(1) Ein *Weg* von $u \in V$, dem *Anfangspunkt des Weges*, nach $v \in V$, dem *Endpunkt des Weges*, ist eine Folge $(e_1, \ldots, e_k)$, $k \geq 1$, von Kanten mit folgenden Eigenschaften:
   (a) $\alpha(e_1) = u$,
   (b) $\omega(e_k) = v$,
   (c) $\omega(e_i) = \alpha(e_{i+1})$, $1 \leq i \leq k-1$.
(2) Ein Weg heißt *geschlossen*, wenn er den gleichen Anfangs- und Endpunkt hat.
(3) Ein Weg heißt *einfach*, wenn alle beteiligten Kanten verschieden sind.
(4) Ein einfacher, geschlossener Weg heißt *Kreis*.
(5) G heißt *zusammenhängend*, wenn es zu jedem $u \in V$ und jedem $v \in V$ mit $u \neq v$ einen Weg von u nach v gibt.
(6) G heißt *einfach zusammenhängend*, wenn es zu jedem $u \in V$ und jedem $v \in V$ mit $u \neq v$ einen einfachen Weg von u nach v gibt.
(7) Ein *ungerichteter Weg* von $u \in V$ nach $v \in V$ ist eine Folge $((v_0, v_1), (v_1, v_2), \ldots, (v_{k-1}, v_k))$, $k \geq 1$, von Tupeln mit folgenden Eigenschaften:
   (a) $v_i \in V$ für alle $0 \leq i \leq k$,
   (b) $v_0 = u$, $v_k = v$,
   (c) $(v_{i-1}, v_i) \in E$ oder $(v_i, v_{i-1}) \in E$ für alle $1 \leq i \leq k$.
(8) G heißt *schwach zusammenhängend*, wenn es zu jedem $u \in V$ und jedem $v \in V$ mit $u \neq v$ einen ungerichteten Weg von u nach v gibt.
(9) Ein *Zyklus* ist ein ungerichteter Weg mit gleichem Anfangs- und Endpunkt, in dem – unabhängig von der Orientierung – keine Kante zweimal durchlaufen wird.

Beispiel II.2.12. Der Graph aus Beispiel II.2.4 (vgl. Abbildung II.2.1 (S. 72)) ist einfach zusammenhängend. Er enthält keine Kreise.

$$\big((A, B), (B, D), (D, F)\big)$$

und

$$\big((A, B), (B, C), (C, E), (E, D), (D, F)\big)$$

sind Wege von A nach F.

$$\big((B, D), (D, E), (E, C), (C, B)\big)$$

ist ein Zyklus, die Kante (C, E) wird in entgegengesetzter Richtung durchlaufen.

DEFINITION II.2.13. Sei $G = (V, E)$ ein Graph und $E' \subset E$. Dann heißt $G' = (V, E')$ der von E' *induzierte Graph*.

DEFINITION II.2.14. (1) Ein schwach zusammenhängender, zyklenfreier Graph heißt *Baum*.
(2) Sei $G = (V, E)$ ein Graph und $E' \subset E$. Der von E' induzierte Graph G' heißt *aufspannender Baum* für G, wenn G' ein Baum ist und für jedes $v \in V$ ein $e' \in E'$ existiert mit $v = \alpha(e')$ oder $v = \omega(e')$.

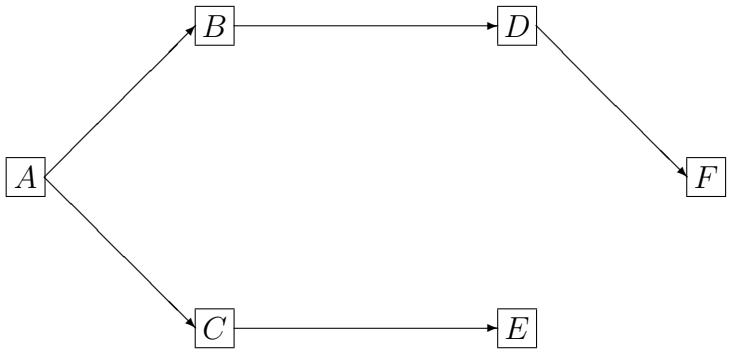

ABBILDUNG II.2.2. Aufspannender Baum für den Graphen aus Beispiel II.2.4 (vgl. Abbildung II.2.1)

BEISPIEL II.2.15. Die Graphen der Abbildungen II.2.2 und II.2.3 sind aufspannende Bäume für den Graphen aus Beispiel II.2.4 (vgl. Abbildung II.2.1 (S. 72)).

SATZ II.2.16. *Es sei $G = (V, E)$ ein schwach zusammenhängender Graph mit $n = \sharp V$ und $m = \sharp E$. Dann gelten folgende Aussagen:*

(1) *Das sukzessive Hinzufügen von Kanten $e \in E$ zu einer wachsenden Kantenmenge T ausgehend von $T = \emptyset$ führt nach $n-1$ Schritten zu einem aufspannenden Baum von G.*
(2) *Gilt*
$$\deg(v) \geq 2$$
für alle $v \in V$, so enthält G einen Zyklus.
(3) *Der sukzessive Abbau von G durch Entfernen eines Knotens v mit*
$$\deg(v) = 1$$
und der dazu inzidenten Kante löscht genau dann alle Kanten, wenn G keine Zyklen enthält.

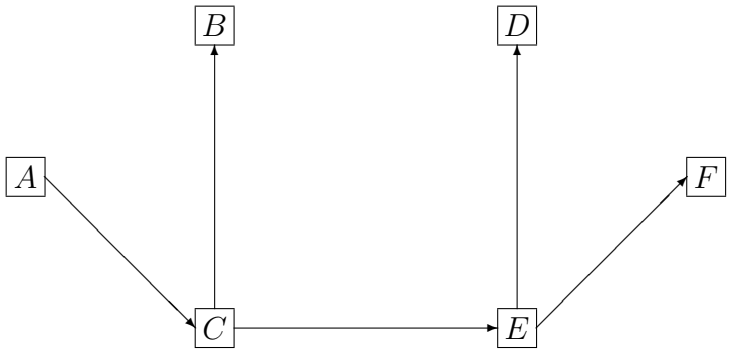

ABBILDUNG II.2.3. Aufspannender Baum für den Graphen aus Beispiel II.2.4 (vgl. Abbildung II.2.1)

(4) *Jeder aufspannende Baum von G hat genau $n-1$ Kanten und jede Kantenmenge $E' \subset E$ mit mehr als $n-1$ Elementen enthält eine Zyklus.*

BEWEIS. *ad (1):* Wir wählen zwei benachbarte Knoten $v_1, v_2 \in V$. Wegen des Zusammenhanges gibt es eine Kante e_{12}, auf der v_1 und v_2 liegen. Wir fügen diese zu T hinzu. Dies liefert einen Baum mit zwei Knoten und einer Kante.
Haben wir bereits einen Baum mit k Knoten und $k-1$ Kanten konstruiert und ist $k < n$, gibt es wegen des Zusammenhanges einen Knoten

$$v_{k+1} \notin \{v_1, \ldots, v_k\}$$

und eine Kante e, auf der v_{k+1} und einer der Knoten aus $\{v_1, \ldots, v_k\}$ liegen. Wir fügen e zu T hinzu. So erhalten wir einen Baum mit $k+1$ Knoten und k Kanten.
Wenn wir so n Knoten abgearbeitet haben, ist V erschöpft und wir sind fertig.
ad (2): Wir wählen zwei benachbarte Knoten v_1 und v_2 und eine verbindende Kante e_{12}. Wegen $\deg(v_2) \geq 2$, gibt es eine weitere von e_{12} verschiedene Kante, auf der v_2 liegt. Sei v_3 der von v_2 verschiedene Knoten auf dieser Kante. Falls $v_3 = v_1$ ist, haben wir einen Zyklus konstruiert und sind fertig.
Andernfalls setzen wir den Prozess mit v_3 an Stelle von v_2 fort.
Dies geht so lange, bis wir einen bereits erreichten Knoten ein zweites Mal erreichen. Jedesmal, wenn wir auf einen neuen Knoten treffen, können wir ihn auf einer anderen Kante verlassen, da sein Grad mindestens 2 ist.
Löschen wir jede bereits besuchte Kante, sinkt der Grad der inzidenten Knoten jeweils um 1. Der Prozess muss nach $n-1$ Schritten enden, da dann keine unbesuchten Knoten mehr existieren.

ad (3)*:* Baut man einen Graphen in der beschriebenen Weise ab, sind zwei Endzustände möglich:

- Es werden nicht alle Kanten gelöscht und die dazu inzidenten Knoten haben mindestens den Grad 2.
- Es sind keine Kanten mehr vorhanden.

Im ersten Fall enthält G wegen Teil (2) einen Zyklus.
Im zweiten Fall wird mit jeder Kante auch ein Knoten gelöscht. Sind daher keine Kanten mehr vorhanden, so auch keine Knoten und der gesamte Graph ist gelöscht.
ad (4)*:* Man braucht mindestens $n-1$ Kanten. Denn betrachtet man zunächst n Knoten als Einzelkomponenten, dann reduziert das Einfügen einer Kante die Komponentenzahl höchstens um 1. Also ist bei weniger als $n-1$ eingefügten Kanten die Komponentenzahl noch ≥ 1. Man braucht auch höchstens $n-1$ Kanten. Denn hätte man mehr als $n-1$ Kanten, könnte man wie in Teil (3) sukzessive alle Knoten vom Grad 1 und die dazu inzidenten Kanten abbauen. Wegen des Zusammenhanges und der Kantenzahl $\geq n$ kann dieser Prozess nicht zur vollständigen Löschung führen, und der Graph enthält wegen Teil (3) einen Zyklus. □

Satz II.2.17. *Sei $G = (V, E)$ ein schwach zusammenhängender Graph und A die zugehörige Inzidenzmatrix. Dann gilt:*

(1) *Die Zeilen von A sind linear abhängig;* $\operatorname{rang}(A) \leq n-1$.
(2) *Die Spalten von A sind genau dann linear unabhängig, wenn G ein Baum ist .*

Beweis. *ad* (1)*:* Folgt aus Bemerkung II.2.7 (1).
ad (2)*:* Wir zeigen die Verneinung der Aussage.
G enthalte also einen Zyklus. Definiere den Vektor $y \in \mathbb{R}^{\sharp E}$ wie folgt

$y_e = \ \ 1$**:** falls e im Zyklus positiv durchlaufen wird,
$y_e = -1$**:** falls e im Zyklus negativ durchlaufen wird,
$y_e = \ \ 0$**:** falls e im Zyklus nicht durchlaufen wird.

Da jeder Knoten im Zyklus den Grad 2 hat, folgt aus der Definition von A und y die Beziehung $Ay = 0$.
Sei nun umgekehrt $y \in \mathbb{R}^{\sharp E} \setminus \{0\}$ ein Vektor mit $Ay = 0$. Definiere $E' = \{e \in E : y_e \neq 0\}$ und bezeichne mit G' den durch E' induzierten Graphen, mit A' die zugehörige Inzidenzmatrix und mit $y' \in \mathbb{R}^{\sharp E'}$ den Vektor, der durch Streichen aller Nullkomponenten von y entsteht. Dann sind alle Komponenten von y' ungleich Null und es gilt $A'y' = 0$. Angenommen G' enthielte einen Knoten v mit Grad 1 in G'. Dann enthält die entsprechende Zeile von A' genau einen von Null verschiedenen Eintrag. Dies widerspricht den Eigenschaften von y'. Also sind die Knoten von G' isoliert oder haben in G' mindestens den Grad 2. Wegen $y \neq 0$ sind nicht alle Knoten von G' isoliert. Aus Satz II.2.16 (2)

angewandt auf eine nicht triviale Zusammenhangskomponenten von G' folgt, dass G' und damit G einen Zyklus enthält. □

II.3. Kürzeste Wege

Gegeben sei ein Graph $G = (V, E)$ und eine Abbildung $d : E \to \mathbb{R}$, die jeder Kante eine „Länge“ zuordnet. Für je zwei verschiedene Knoten $u, v \in V$ bezeichne $\Gamma(u, v)$ die Menge aller Wege mit Anfangspunkt u und Endpunkt v. Das *Problem des kürzesten Weges* besteht nun darin, zu gegebenem $u, v \in V$ ein $\gamma \in \Gamma(u, v)$ zu finden, so dass die „Länge“

$$d(\gamma) = \sum_{e \in \gamma} d(e)$$

minimal ist.

Bemerkung II.3.1. Die Interpretation „Länge“ legt die Voraussetzung $d(e) \geq 0$ für alle $e \in E$ nahe. Für einige Anwendungen ist es aber notwendig, auch negative Werte für d zuzulassen. Wie wir sehen werden, führt dies allerdings zu einigen Erschwernissen.

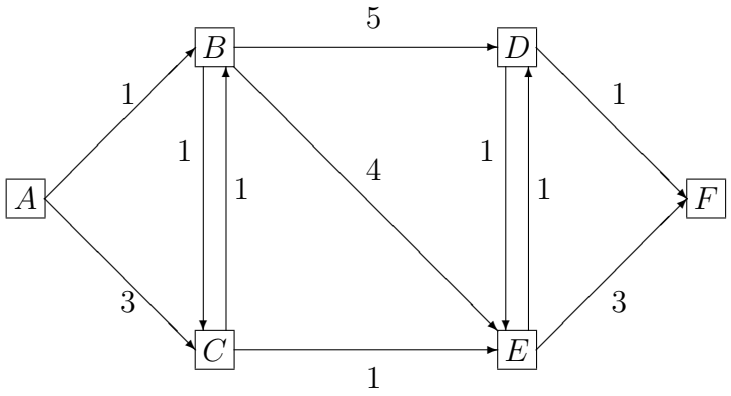

Abbildung II.3.1. Graph aus Beispiel II.3.2 mit Länge der Kanten

Beispiel II.3.2. Wir betrachten den in Abbildung II.3.1 skizzierten Graphen (vgl. Beispiel II.2.4 (S. 73) und Abbildung II.2.1 (S. 72)), wobei die Zahlen an den Kanten die Längen der jeweiligen Kante angeben. Dann besteht $\Gamma(A, F)$ aus folgenden Wegen mit folgender Länge:

A, B, D, F	7,	A, B, D, E, F	10,
A, B, E, F	8,	A, B, E, D, F	7,
A, B, C, E, F	6,	A, B, C, E, D, F	5,
A, C, E, F	7,	A, C, E, D, F	6,
A, C, B, D, F	10,	A, C, B, D, E, F	13,

$$A, C, B, E, F \quad 11, \qquad\qquad A, C, B, E, D, F \quad 10.$$

Der kürzeste Weg ist offensichtlich A, B, C, E, D, F mit der Länge 5.

Algorithmus II.3.1 löst das Problem des kürzesten Weges unter der Voraussetzung $d(e) \geq 0$ für alle $e \in E$.

Algorithmus II.3.1 Algorithmus von Dijkstra

Gegeben: Graph $G = (V, E)$, Längenfunktion $d : E \to \mathbb{R}_+$, Anfangsknoten $s \in V$, Endknoten $t \in V \setminus \{s\}$
Gesucht: Kürzester Weg von s nach t

▷ $D(v)$ Länge des bisher kürzesten Weges von s nach v
▷ $V(v)$ Vorgänger von v auf dem bisher kürzesten (s, v)-Weg
▷ $M(v)$ Länge des kürzesten Weges von s nach v
▷ $\mathcal{M}$ markierte Knoten, $\mathcal{U}$ unmarkierte Knoten

1: $D(s) \leftarrow 0$, $M(s) \leftarrow 0$, $\mathcal{M} \leftarrow \{s\}$, $\mathcal{U} \leftarrow V \setminus \{s\}$ ▷ Initialisierung
2: **for** $v \in V$ mit $(s, v) \in E$ **do** $D(v) \leftarrow d((s, v))$
3: **end for**
4: **for** $v \in V$ mit $(s, v) \notin E$ **do** $D(v) \leftarrow \infty$
5: **end for**
6: **for** $v \in V \setminus \{s\}$ **do** $V(v) \leftarrow s$
7: **end for**
8: **while** $\mathcal{U} \neq \emptyset$ **do**
9: $u \leftarrow \operatorname{argmin}\{D(v) : v \in \mathcal{U}\}$
10: **if** $D(u) = \infty$ **then**
11: **stop** ▷ es gibt keinen (s, t)-Weg
12: **end if**
13: **while** $u \neq t$ **do**
14: $\mathcal{M} \leftarrow \mathcal{M} \cup \{u\}$, $\mathcal{U} \leftarrow \mathcal{U} \setminus \{u\}$, $M(u) \leftarrow D(u)$
15: **for** $v \in \mathcal{U}$ mit $(u, v) \in E$ **do**
16: **if** $D(V) > M(u) + d((u, v))$ **then**
17: $D(V) \leftarrow M(u) + d((u, v))$, $V(v) \leftarrow u$
18: **end if**
19: **end for**
20: **end while**
21: **end while**

▷ $M(t)$ ist die gesuchte Länge.
▷ Der kürzeste Weg ist $t, V(t), V(V(t)), V(V(V(t))), \ldots, s$.

Beispiel II.3.3. Wir wenden Algorithmus II.3.1 auf den Graphen aus Beispiel II.3.2 mit $s = A$ und $t = F$ an.
Nach der Initialisierung ist

$$\mathcal{M} = \{A\}, \qquad \mathcal{U} = \{B, C, D, E, F\},$$
$$D(A) = 0,$$
$$D(B) = 1, \qquad V(B) = A,$$

$$\begin{aligned} D(C) &= 3, & V(C) &= A, \\ D(D) &= \infty, & V(D) &= A, \\ D(E) &= \infty, & V(E) &= A, \\ D(F) &= \infty, & V(F) &= A, \\ M(s) &= 0. \end{aligned}$$

Wir kommen zu Zeile 9 und erhalten

$$u = B.$$

Damit ergibt Zeile 14

$$\mathcal{M} = \{A, B\}, \quad \mathcal{U} = \{C, D, E, F\}, \quad M(B) = 1.$$

Zeile 17 liefert

$$\begin{aligned} D(C) &= 2, & V(C) &= B, \\ D(D) &= 6, & V(D) &= B, \\ D(E) &= 5, & V(E) &= B, \\ D(F) &= \infty, & V(F) &= A. \end{aligned}$$

Damit kommen wir erneut zu Zeile 9 und erhalten jetzt

$$u = C.$$

Zeile 14 ergibt

$$\mathcal{M} = \{A, B, C\}, \quad \mathcal{U} = \{D, E, F\}, \quad M(C) = 2.$$

Zeile 17 liefert

$$D(E) = 3, \ V(E) = C.$$

Alle anderen Einträge bleiben unverändert.
Wir kommen erneut zu Zeile 9 und erhalten nun

$$u = E.$$

Zeile 14 ergibt

$$\mathcal{M} = \{A, B, C, E\}, \quad \mathcal{U} = \{D, F\}, \quad M(E) = 3.$$

Zeile 17 liefert

$$\begin{aligned} D(D) &= 4, & V(D) &= E, \\ D(F) &= 6, & V(F) &= E. \end{aligned}$$

Wir kommen erneut zu Zeile 9 und erhalten nun

$$u = D.$$

Zeile 14 ergibt

$$\mathcal{M} = \{A, B, C, E, D\}, \quad \mathcal{U} = \{F\}, \quad M(D) = 4.$$

Zeile 17 liefert

$$D(F) = 5, \ V(F) = D.$$

Damit kommen wir erneut zu Zeile 9 und erhalten jetzt

$$u = F.$$

Wegen $u = t$ endet Algorithmus II.3.1 und liefert als Endergebnis

$$M(F) = 5.$$

Mit vollständiger Induktion kann man zeigen, dass Algorithmus II.3.1 tatsächlich den kürzesten Weg liefert [1, Satz 24.6]. Bei negativen Weglängen kann Algorithmus II.3.1 jedoch versagen. Dies zeigt das folgende Beispiel.

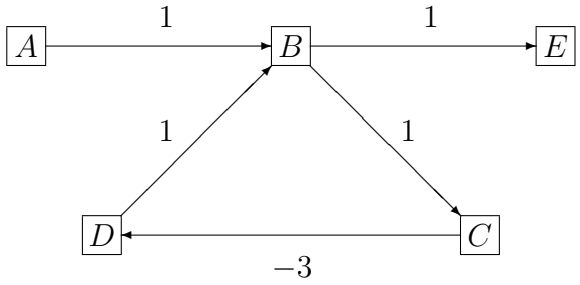

ABBILDUNG II.3.2. Graph aus Beispiel II.3.4 mit „Länge“ der jeweiligen Kanten

BEISPIEL II.3.4. Betrachte den in Abbildung II.3.2 skizzierten Graphen, wobei die Zahlen an den Kanten wieder die „Länge“ der jeweiligen Kante angeben.
Offensichtlich hat der Weg

$$A, B, E \text{ die Länge } 2,$$

der Weg

$$A, B, C, D, B, E \text{ die Länge } 1$$

und der Weg

$$A, B, C, D, B, C, D, B, E \text{ die Länge } 0.$$

Indem wir beliebig oft den Kreis B, C, D durchlaufen, können wir die Länge des Gesamtweges beliebig klein machen. Das Problem des kürzesten Weges hat also keine Lösung.
Anders sieht es aus, wenn wir den Kreis B, C, D entfernen, indem wir die Kante (D, B) streichen.

Obiges Beispiel zeigt, dass man bei negativen Werten der Funktion d mindestens die Kreisfreiheit des Graphen fordern muss. Dies ergibt das Problem, Kreise aufzufinden. Algorithmus II.3.2 leistet dies. Er ist durch Satz II.2.16(3) (S. 76) motiviert.

Enthält ein Graph einen Kreis, kann man diesen entfernen, indem man eine Kante entfernt. Dabei sollte der Zusammenhang des Graphen erhalten bleiben. So kann man z. B. in dem Graphen aus Beispiel II.3.4 und Abbildung II.3.2 die Kante (D, B) entfernen, nicht aber die Kanten (B, C) oder (C, D). Wir brauchen also einen Algorithmus, der prüft,

Algorithmus II.3.2 Auffinden eines Kreises in einem Graphen

Gegeben: Graph $G = (V, E)$
Gesucht: Kreis in G oder Information der Kreisfreiheit

Entferne alle Knoten mit Grad 1 aus V und alle zugehörigen Kanten aus E. Verringere die Grade der restlichen Knoten entsprechend.
if $V = \emptyset$ **then**
    stop ▷ Graph ist kreisfrei
end if
$n \leftarrow \sharp V$, $v_0 \leftarrow$ Knoten in V
for $i = 1, \ldots, n-1$ **do**
    $e \leftarrow$ Kante in $K^-(v_{i-1})$, $v_i \leftarrow \omega(e)$
    if $v_i \in \{v_0, \ldots, v_{i-1}\}$ **then**
        stop ▷ Kreis gefunden
    else
        $E \leftarrow E \setminus e$, $\deg(v_{i-1}) \leftarrow \deg(v_{i-1}) - 1$, $\deg(v_i) \leftarrow \deg(v_i) - 1$
        if $\deg(v_{i-1}) = 0$ **then**
            $\{v_0, \ldots, v_{i-1}\} \leftarrow \{v_0, \ldots, v_{i-1}\} \setminus \{v_{i-1}\}$, $i \leftarrow i - 1$
        end if
    end if
end for

ob der Zusammenhang erhalten bleibt, wenn eine Kante entfernt wird. Dies leistet Algorithmus II.3.3.

Algorithmus II.3.3 Test auf Zusammenhang nach Entfernen einer Kante

Gegeben: $G = (V, E)$, Kante $e = (u, v)$
Gesucht: Information, ob $(V, E \setminus e)$ zusammenhängend ist

$T \leftarrow \{u\}$, $E \leftarrow E \setminus e$
while existiert $e \in E$ mit $\alpha(e) \in T$ und $\omega(e) \notin T$ **do**
    $T \leftarrow T \cup \omega(e)$
end while
if $v \in T$ **then**
    stop ▷ Kante (u, v) darf entfernt werden.
else
    stop ▷ Kante (u, v) darf nicht entfernt werden.
end if

LEMMA II.3.5. *Ist der Graph $G = (V, E)$ kreisfrei, können die Knoten so nummeriert werden, dass für alle Kanten $(v_i, v_j) \in E$ gilt $i < j$. Umgekehrt folgt aus der Existenz einer solchen Nummerierung die Kreisfreiheit.*

BEWEIS. Die Umkehrung ist offensichtlich.
Zum Nachweis der Existenz der Nummerierung zeigen wir zunächst,

dass es ein $v \in V$ gibt mit $\sharp K^+(v) = 0$.
Denn andernfalls hat jeder Knoten einen Vorgänger und bei Zurückverfolgung stoßen wir wegen der endlichen Knotenzahl irgendwann auf einen schon behandelten Knoten und haben so einen Kreis gefunden. Dies ist ein Widerspruch.
Wir setzen $i = 1$ und suchen ein $v \in V$ mit $\sharp K^+(v) = 0$. Wir setzen $v_i = v$ und entfernen v aus V und $K^-(v)$ aus E. Danach setzen das Verfahren mit $i + 1$ an Stelle von i fort. □

Algorithmus II.3.4 liefert die gewünschte Nummerierung und setzt das im Beweis von Lemma II.3.5 benutzte Verfahren um.

Algorithmus II.3.4 Sortieralgorithmus

Gegeben: kreisfreier Graph $G = (V, E)$
Gesucht: aufsteigende Nummerierung von V ▷ $(v_i, v_j) \in E \Rightarrow i < j$

1: $\mathcal{U} \leftarrow V$, $\mathcal{M} \leftarrow \emptyset$
2: **for** $i = 1, \ldots, \sharp V$ **do**
3: $v \leftarrow$ Knoten in $\mathcal{U}$ mit $\sharp K^+(v) = 0$
4: $v_i \leftarrow v$, $\mathcal{U} \leftarrow \mathcal{U} \setminus \{v\}$, $\mathcal{M} \leftarrow \mathcal{M} \cup \{v\}$
5: Lösche alle Kanten aus $K^-(v)$.
6: **end for**

Algorithmus II.3.5 baut auf diesen Vorbereitungen auf und löst das Problem des kürzesten Weges auch bei negativen „Längen“. Dabei können die Algorithmen II.3.2, II.3.3 und II.3.4 eingesetzt werden, um die erforderliche Voraussetzungen herzustellen.

Manchmal benötigt man für einen Graphen sämtliche kürzeste Wege zwischen allen Knotenpaaren. Dies kann man erreichen, indem man Algorithmus II.3.1 bei nicht negativen Längen bzw. Algorithmus II.3.5 für alle Knotenpaare sukzessive aufruft. Algorithmus II.3.6 ist für diese Aufgabenstellung effizienter. Im Gegensatz zu den Algorithmen II.3.1 und II.3.5 benötigt er weder nicht negative Längen noch Kreisfreiheit. Für einzelne Knotenpaare sind aber die Algorithmen II.3.1 und II.3.5 effizienter.

Bemerkung II.3.6. Algorithmus II.3.6 liefert die Wege rückwärts durch rekursive Auswertung der Matrix P: Für je zwei Indizes i und j erhält man die Indizes der Knoten auf einem kürzesten Weg von v_i nach v_j rückwärts durch

$$j\,,\ j_q = p_{ij}\,,\ j_{q-1} = p_{ij_q}\,,\ \ldots\,,\ j_1 = p_{ij_2}\,,\ i = p_{ij_1}.$$

II.4. Flüsse in Netzwerken

Definition II.4.1. Ein *Netzwerk* ist ein Graph $G = (V, E)$ zusammen mit einer *Kapazitätsfunktion* $c : E \to \mathbb{R}_+$. $c(e)$ heißt die *Kapazität* der Kante $e \in E$.

Algorithmus II.3.5 Algorithmus von Moore-Bellman

Gegeben: kreisfreier, zusammenhängender Graph $G = (V, E)$ mit aufsteigender Nummerierung, Längenfunktion $d : E \to \mathbb{R}$, Anfangsknoten v_i, Endknoten v_j

▷ $(v_m, v_n) \in E \Rightarrow m < n$

Gesucht: kürzester Weg von v_i nach v_j und seine Länge ▷ $i < j$

1: $D(v_i) \leftarrow 0$
2: **for** $(v_i, v) \in E$ **do** $D(v) \leftarrow d((v_i, v))$
3: **end for**
4: **for** $(v_i, v) \notin E$ **do** $D(v) \leftarrow \infty$
5: **end for**
6: **for** $v \in V \setminus \{v_i\}$ **do** $V(v) \leftarrow v_i$
7: **end for**
8: **for** $k = i + 2, \ldots, j$ **do**
9: **for** $\ell = i + 1, \ldots, k - 1$ **do**
10: **if** $(v_\ell, v_k) \in E$ und $D(v_\ell) + d((v_\ell, v_k)) < D(v_k)$ **then**
11: $D(v_k) \leftarrow D(v_\ell) + d((v_\ell, v_k))$, $V(v_k) \leftarrow v_\ell$
12: **end if**
13: **end for**
14: **end for**

▷ kürzeste Weg: $v_j, V(v_j), V(V(v_j)), \ldots, v_i$; Länge: $D(v_j)$.

Definition II.4.2. Seien $G = (V, E)$, c ein Netzwerk und $s, t \in V$ zwei verschiedene Knoten. Eine Abbildung $x : E \to \mathbb{R}$ heißt ein *zulässiger* (s, t)*-Fluss*, wenn folgende Bedingungen erfüllt sind:

(II.4.1) $0 \le x(e) \le c(e)$ für alle $e \in E$,

Kapazitätsbeschränkungen

(II.4.2) $\sum_{e \in K^-(v)} x(e) = \sum_{e \in K^+(v)} x(e)$ für alle $v \in V \setminus \{s, t\}$.

Flusserhaltungsgleichungen

Ist x ein zulässiger (s, t)-Fluss, heißt

$$\varphi(x) = \sum_{e \in K^-(s)} x(e) - \sum_{e \in K^+(s)} x(e)$$

der *Wert* des Flusses.

Wir betrachten nun folgende Aufgabe:

> Finde zu gegebenem Netzwerk und gegebenen Knoten s, t einen zulässigen Fluss mit maximalem Wert.

Physikalisch können wir uns diese Aufgabe so vorstellen, dass wir durch ein Rohrnetz mit gegebener Kapazität möglichst viel Material von s nach t transportieren wollen, wobei zwischendurch nichts verloren gehen soll.

Algorithmus II.3.6 Algorithmus von Floyd-Warshall

Gegeben: Graph $G = (V, E)$, $n = \sharp V$, Längenfunktion $d : E \to \mathbb{R}$
Gesucht: Matrix $W = (w_{ij})_{1 \le i,j \le n}$ mit Länge des kürzesten Weges von Knoten v_i zu Knoten v_j, Matrix $P = (p_{ij})_{1 \le i,j \le n}$ mit Nummer des vorletzten Knotens auf dem kürzesten Weg von Knoten v_i zu Knoten v_j

```
1:  for i = 1, ..., n do                                   ▷ Initialisierung
2:      for j = 1, ..., n do
3:          if (v_i, v_j) ∈ E then
4:              w_ij ← d((v_i, v_j)), p_ij ← 1
5:          else
6:              w_ij ← ∞, p_ij ← ∞
7:          end if
8:      end for
9:  end for
10: for ℓ = 1, ..., n do
11:     for i = 1, ..., n do
12:         for j = 1, ..., n do          ▷ kann in Zeile 17 abbrechen
13:             if w_ij > w_iℓ + w_ℓj then
14:                 w_ij ← w_iℓ + w_ℓj, p_ij ← p_ℓj
15:             end if
16:             if i = j und w_ii < 0 then
17:                 stop          ▷ Graph enthält Kreis negativer Länge
18:             end if
19:         end for
20:     end for
21: end for
```

Definition II.4.3. Die Bezeichnungen seien wie in Definition II.4.2. Weiter sei P ein ungerichteter Weg von s nach t. Eine Kante $e \in P$ heißt ein *Vorwärtsbogen*, wenn sie von s aus in Richtung t verläuft. Andernfalls heißt sie ein *Rückwärtsbogen*. P heißt ein *augmentierender* oder *ergänzender Weg*, wenn für alle Kanten $e \in P$ gilt:

$$x(e) < c(e) \quad \text{falls } e \text{ Vorwärtsbogen ist,}$$
$$x(e) > 0 \quad \text{falls } e \text{ Rückwärtsbogen ist.}$$

Beispiel II.4.4. Betrachte den Graphen aus Abbildung II.4.1, wobei die Zahlen an den Kanten die jeweilige Kapazität angeben. Abbildung II.4.2 gibt zu diesem Netzwerk einen zulässigen Fluss von A nach F an. Der Wert des Flusses ist 3. Abbildung II.4.3 zeigt dann einen ergänzenden Weg, wobei die Zahlen an den Kanten Werte angeben, um die der Fluss aus Abbildung II.4.2 erhöht bzw. erniedrigt werden kann, ohne die Zulässigkeit zu zerstören. Die Kante (E, D) ist ein Rückwärtsbogen, die anderen drei Kanten sind Vorwärtsbögen. Der

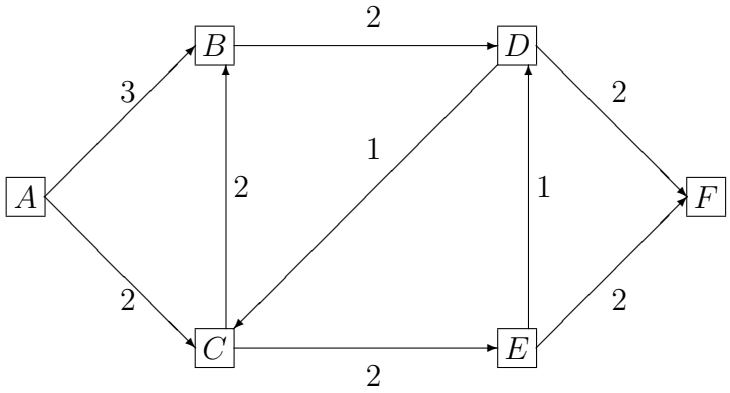

ABBILDUNG II.4.1. Graph aus Beispiel II.4.4 mit Kapazitäten der einzelnen Kanten

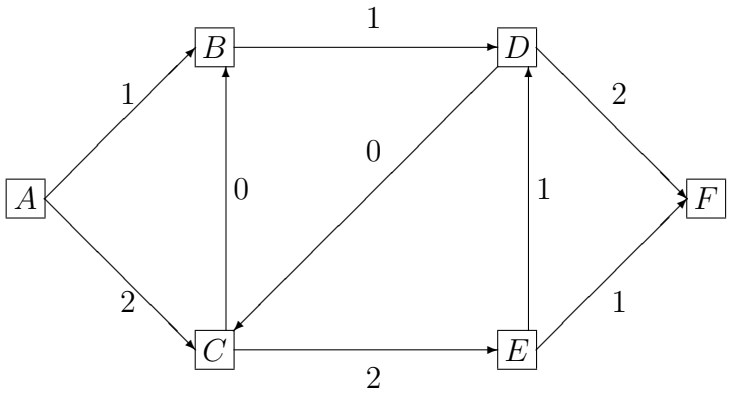

ABBILDUNG II.4.2. Zulässiger Fluss von A nach F für das Netzwerk aus Abbildung II.4.1; die Zahlen geben die jeweiligen Flusswerte an

resultierende Fluss ist in Abbildung II.4.4 dargestellt. Er hat den Wert 4.

SATZ II.4.5. *Ein zulässiger (s,t)-Fluss ist genau dann maximal, wenn es keinen ergänzenden Weg von s nach t gibt.*

BEWEIS. Wir zeigen die Äquivalenz:

> Ein zulässiger (s,t)-Fluss ist genau dann nicht maximal, wenn es einen ergänzenden Weg gibt.

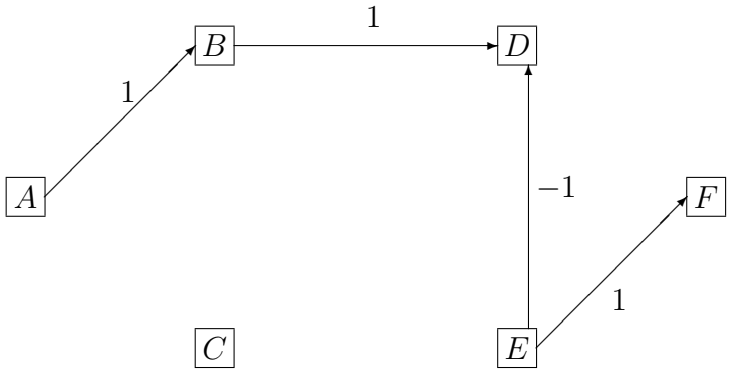

ABBILDUNG II.4.3. Ergänzender Weg zu dem Fluss aus Abbildung II.4.2; die Zahlen geben die jeweiligen zusätzlichen Flusswerte an

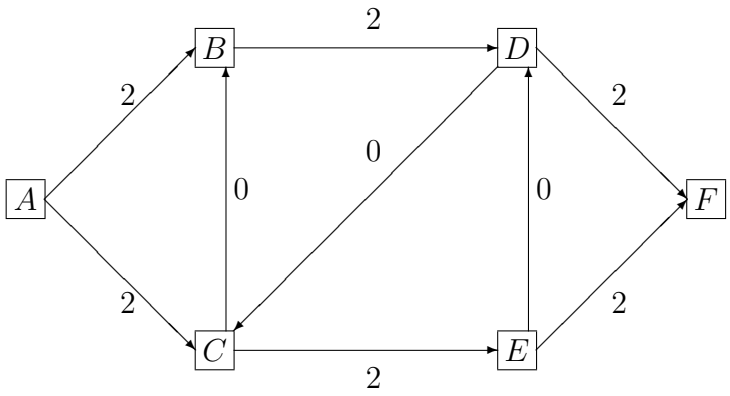

ABBILDUNG II.4.4. Resultierender Fluss aus dem Fluss aus Abbildung II.4.2 und dem ergänzenden Weg aus Abbildung II.4.3; die Zahlen geben die jeweiligen Flusswerte an

„$\Longleftarrow$“: Sei P ein ergänzender Weg. Dann ist

$$\varepsilon = \min\Big\{\min\{c(e) - x(e) : e \in P \text{ ist Vorwärtsbogen}\},\\ \min\{x(e) : e \in P \text{ ist Rückwärtsbogen}\}\Big\}$$

positiv. Definiere $y : E \to \mathbb{R}$ durch

$$y(e) = \begin{cases} x(e) & \text{falls } e \notin P,\\ x(e) + \varepsilon & \text{falls } e \in P \text{ ein Vorwärtsbogen},\\ x(e) - \varepsilon & \text{falls } e \in P \text{ ein Rückwärtsbogen}. \end{cases}$$

Wegen der Wahl von ε erfüllt y die Kapazitätsbeschränkungen (II.4.1). Ist $v \in V\backslash\{s,t\}$ ein Knoten, der nicht von P durchlaufen wird, unterscheidet sich y auf allen zu v inzidenten Kanten nicht von x. Daher erfüllt y in v die Flusserhaltungsgleichung (II.4.2).
Ist dagegen $v \in V\backslash\{s,t\}$ ein Knoten, der von P durchlaufen wird, so gibt es eine Kante e^+, auf der P in v hineinläuft, und eine Kante e^-, auf der P aus v herausläuft. Sind e^+ und e^- beides Vorwärts- oder Rückwärtsbögen, können nicht beide in der gleichen Menge $K^{\pm}(v)$ liegen. Daher gilt in diesem Knoten die Flusserhaltungsgleichung (II.4.2) für y. Ist aber eine der Kanten ein Vorwärts- und die andere ein Rückwärtsbogen, so müssen beide Kanten in der gleichen Menge $K^{\pm}(v)$ liegen. Daher ändert sich auch in diesem Fall nichts an der Gültigkeit der Flusserhaltungsgleichung (II.4.2).
Also ist y ein zulässiger Fluss.
Da es mindestens eine Kante in $K^+(s) \cup K^-(s)$ gibt, die zu P gehört, ist

$$\varphi(y) > \varphi(x).$$

Also war x nicht maximal.
„$\Longrightarrow$“: Sei nun x ein nicht maximaler zulässiger (s,t)-Fluss und y ein zulässiger (s,t)-Fluss mit größerem Wert. Setze

$$z = y - x.$$

Dann erfüllt z die Flusserhaltungsgleichungen (II.4.2) und es ist

$$\sum_{e \in K^-(s)} z(e) - \sum_{e \in K^+(s)} z(e) > 0.$$

Daher gibt es mindestens eine Kante $e_1 \in K^-(s) \cup K^+(s)$, auf der z nicht verschwindet. Sei v_1 der andere Knoten auf e_1. Wegen der Flusserhaltungsgleichung (II.4.2) in v_1 gibt es eine Kante $e_2 \in K^-(v_1) \cup K^+(v_1)\backslash\{e_1\}$, auf der z nicht verschwindet. Sei v_2 der von v_1 verschiedene Knoten auf e_2.
Wenn wir diesen Prozess fortsetzen, erhalten wir einen Weg von s nach t. Aus der Konstruktion von z folgt, dass dieser Weg ergänzend ist. □

Satz II.4.5 führt auf Algorithmus II.4.1.

Beispiel II.4.6. Wir wenden Algorithmus II.4.1 auf Beispiel 4 (S. 7) an. Abbildung II.4.5 gibt das Netzwerk mit den jeweiligen maximalen Flusswerten und der Nummerierung der Kanten durch kleine Buchstaben an. Wir ordnen die Knoten in der Reihenfolge Q, A, B, C, D, R. Es ist $s = Q$, $t = R$.
Im ersten Durchlauf erhalten wir:

Zeile 6:	$\mathcal{M} = \{Q\}$,	$\mathcal{U} = \{Q\}$,
Zeile 11:	$u = Q$,	$\mathcal{U} = \emptyset$,
Zeilen 12 – 18:	$\mathcal{V}(D) = Q$,	$\mathcal{S}(D) = 1$,

Algorithmus II.4.1 Algorithmus von Ford-Fulkerson

Gegeben: Graph $G = (V, E)$, Kapazität c, Knoten $s, t \in V$
Gesucht: Maximaler zulässiger (s,t)-Fluss x

▷ $\mathcal{M}$ markierte Knoten, $\mathcal{U}$ noch zu bearbeitende Knoten
▷ $\mathcal{V}$ Vorgänger auf einem ergänzenden Weg, $\mathcal{S}$ Vorzeichen
▷ $\mathcal{E}$ Steigerungsmöglichkeiten auf einem ergänzenden Weg

1: **for** $e \in E$ **do** $x(e) \leftarrow 0$ ▷ Initialisierung
2: **end for**
3: **for** $v \in V$ **do** $\mathcal{E}(v) \leftarrow \infty$
4: **end for**
5: **loop** ▷ wird in Zeile 9 verlassen
6: $\mathcal{M} \leftarrow \{s\}, \mathcal{U} \leftarrow \{s\}$ ▷ Markieren und Überprüfen
7: **loop** ▷ wird in Zeile 31 verlassen
8: **if** $\mathcal{U} = \emptyset$ **then**
9: **stop** ▷ Fluss x ist maximal.
10: **end if**
11: $u \leftarrow$ Element in $\mathcal{U}, \mathcal{U} \leftarrow \mathcal{U} \setminus \{u\}$
12: **for** $e \in K^-(u)$ mit $\omega(e) \notin \mathcal{M}$ **do**
13: $v \leftarrow \omega(e)$
14: **if** $x(e) < c(e)$ **then**
15: $\mathcal{V}(v) \leftarrow u$, $\mathcal{S}(v) \leftarrow 1$, $\mathcal{E}(v) \leftarrow \min\{c(e) - x(e), \mathcal{E}(u)\}$
16: $\mathcal{M} \leftarrow \mathcal{M} \cup \{v\}, \mathcal{U} \leftarrow \mathcal{U} \cup \{v\}$
17: **end if**
18: **end for**
19: **for** $e \in K^+(u)$ mit $\alpha(e) \notin \mathcal{M}$ **do**
20: $v \leftarrow \alpha(e)$
21: **if** $x(e) > 0$ **then**
22: $\mathcal{V}(v) \leftarrow u$, $\mathcal{S}(v) \leftarrow -1$, $\mathcal{E}(v) \leftarrow \min\{x(e), \mathcal{E}(u)\}$
23: $\mathcal{M} \leftarrow \mathcal{M} \cup \{v\}, \mathcal{U} \leftarrow \mathcal{U} \cup \{v\}$
24: **end if**
25: **end for**
26: **if** $t \in \mathcal{M}$ **then** ▷ Augmentierung
27: $u \leftarrow t$, $v \leftarrow t$
28: **while** $v \neq s$ **do**
29: $v \leftarrow \mathcal{V}(u)$, $x((v,u)) \leftarrow x((v,u)) + \mathcal{S}(u)\mathcal{E}(u)$, $u \leftarrow v$
30: **end while**
31: **break** ▷ springe in Zeile 6
32: **end if**
33: **end loop**
34: **end loop**

$$\begin{aligned} \mathcal{E}(D) &= 6, & \mathcal{V}(A) &= Q, \\ \mathcal{S}(A) &= 1, & \mathcal{E}(A) &= 4, \\ \mathcal{M} &= \{Q, D, A\}, & \mathcal{U} &= \{D, A\}, \end{aligned}$$

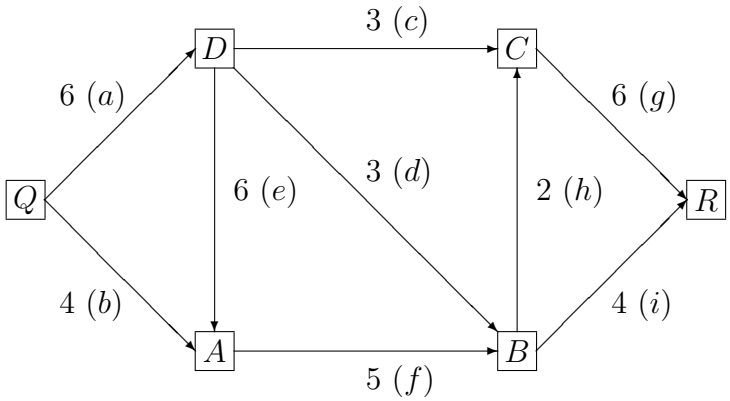

ABBILDUNG II.4.5. Netzwerk aus Beispiel II.4.6 mit Kapazitäten der einzelnen Kanten und Nummerierung der Kanten (kleine Buchstaben)

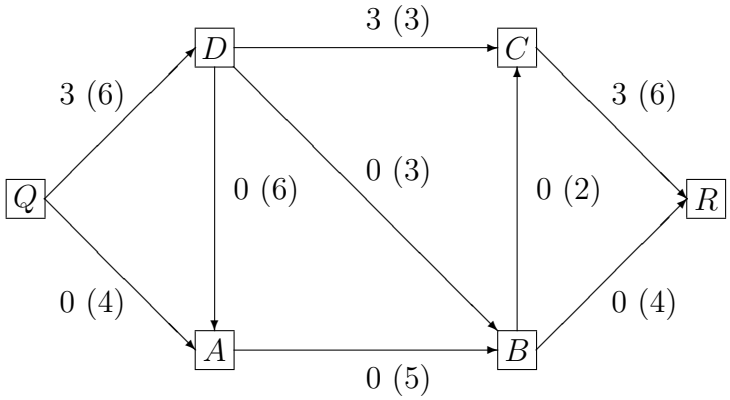

ABBILDUNG II.4.6. Fluss nach dem ersten Erreichen von Zeilen 26 – 32 von Algorithmus II.4.1 mit jeweiligem Flusswert bzw. in Klammern Kapazität

Zeile 11:	$u = D,$	$\mathcal{U} = \{A\},$
Zeilen 12 – 18:	$\mathcal{V}(C) = D,$	$\mathcal{S}(C) = 1,$
	$\mathcal{E}(C) = 3,$	$\mathcal{V}(B) = D,$
	$\mathcal{S}(B) = 1,$	$\mathcal{E}(B) = 3,$
	$\mathcal{M} = \{Q, D, A, C, B\},$	$\mathcal{U} = \{A, C, B\},$
Zeilen 11:	$u = A,$	$\mathcal{U} = \{C, B\},$
Zeilen 11:	$u = C,$	$\mathcal{U} = \{B\},$
Zeilen 12 – 18:	$\mathcal{V}(R) = C,$	$\mathcal{S}(R) = 1,$

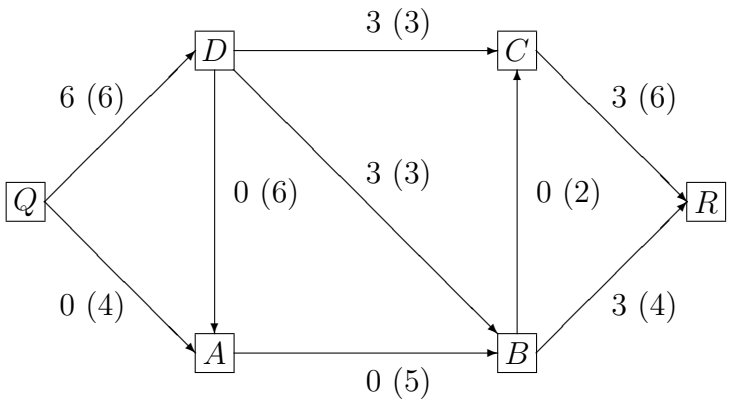

ABBILDUNG II.4.7. Fluss nach dem zweiten Erreichen von Zeilen 26 – 32 von Algorithmus II.4.1 mit jeweiligen Flusswerten bzw. in Klammern Kapazität

$$\begin{aligned}\mathcal{E}(R) &= 3,\\ \mathcal{M} &= \{Q, D, A, C, B, R\},\\ \mathcal{U} &= \{B, R\}.\end{aligned}$$

Zeilen 26 – 32 liefern jetzt den in Abbildung II.4.6 angegebenen Fluss. Die Zahlen an den Kanten geben die jeweiligen Flusswerte bzw. in Klammern die Kapazität an.
Im nächsten Durchlauf erhalten wir:

Zeile 6:	$\mathcal{M} = \{Q\}$,	$\mathcal{U} = \{Q\}$,
Zeile 11:	$u = Q$,	$\mathcal{U} = \emptyset$,
Zeilen 12 – 18:	$\mathcal{V}(D) = Q$,	$\mathcal{S}(D) = 1$,
	$\mathcal{E}(D) = 3$,	$\mathcal{V}(A) = Q$,
	$\mathcal{S}(A) = 1$,	$\mathcal{E}(A) = 4$,
	$\mathcal{M} = \{Q, D, A\}$,	$\mathcal{U} = \{D, A\}$,
Zeile 11:	$u = D$,	$\mathcal{U} = \{A\}$,
Zeilen 12 – 18:	$\mathcal{V}(B) = D$,	$\mathcal{S}(B) = 1$,
	$\mathcal{E}(B) = 3$,	$\mathcal{M} = \{Q, D, A, B\}$,
	$\mathcal{U} = \{A, B\}$,	
Zeile 11:	$u = A$,	$\mathcal{U} = \{B\}$,
Zeile 11:	$u = B$,	$\mathcal{U} = \emptyset$,
Zeilen 12 – 18:	$\mathcal{V}(C) = B$,	$\mathcal{S}(C) = 1$,
	$\mathcal{E}(C) = 2$,	
	$\mathcal{M} = \{Q, D, A, B, C\}$,	

$$\begin{aligned} \mathcal{U} &= \{C\}, & \mathcal{V}(R) &= B, \\ \mathcal{S}(R) &= 1, & \mathcal{E}(R) &= 3, \\ \mathcal{M} &= \{Q, D, A, B, C, R\}, \\ \mathcal{U} &= \{C, R\}. \end{aligned}$$

Zeilen 26 – 32 liefern jetzt den in Abbildung II.4.7 angegebenen Fluss. Die Zahlen an den Kanten geben die jeweiligen Flusswerte bzw. in Klammern die Kapazität an.
Im nächsten Durchlauf erhalten wir:

$$\begin{aligned} \text{Zeile 6:} \quad & \mathcal{M} = \{Q\}, & \mathcal{U} &= \{Q\}, \\ \text{Zeile 11:} \quad & u = Q, & \mathcal{U} &= \emptyset, \\ \text{Zeilen 12 – 18:} \quad & \mathcal{V}(A) = Q, & \mathcal{S}(A) &= 1, \\ & \mathcal{E}(A) = 4, & \mathcal{M} &= \{Q, A\}, \\ & \mathcal{U} = \{A\}, \\ \text{Zeile 11:} \quad & u = A, & \mathcal{U} &= \emptyset, \\ \text{Zeilen 12 – 18:} \quad & \mathcal{V}(B) = A, & \mathcal{S}(B) &= 1, \\ & \mathcal{E}(B) = 4, & \mathcal{M} &= \{Q, A, B\}, \\ & \mathcal{U} = \{B\}, \\ \text{Zeile 11:} \quad & u = B, & \mathcal{U} &= \emptyset, \\ \text{Zeilen 12 – 18:} \quad & \mathcal{V}(C) = B, & \mathcal{S}(C) &= 1, \\ & \mathcal{E}(C) = 2, & \mathcal{M} &= \{Q, A, B, C\}, \\ & \mathcal{U} = \{C\}, & \mathcal{V}(R) &= B, \\ & \mathcal{S}(R) = 1, & \mathcal{E}(R) &= 1, \\ & \mathcal{M} = \{Q, A, B, C, R\}, & \mathcal{U} &= \{C, R\}, \\ \text{Zeilen 19 – 25:} \quad & \mathcal{V}(D) = B, & \mathcal{S}(D) &= -1, \\ & \mathcal{E}(D) = 3, \\ & \mathcal{M} = \{Q, A, B, C, R, D\}, \\ & \mathcal{U} = \{C, R, D\}. \end{aligned}$$

Zeilen 26 – 32 liefern jetzt den in Abbildung II.4.8 angegebenen Fluss. Die Zahlen an den Kanten geben die jeweiligen Flusswerte bzw. in Klammern die Kapazität an.
Im nächsten Durchlauf erhalten wir:

$$\begin{aligned} \text{Zeile 6:} \quad & \mathcal{M} = \{Q\}, & \mathcal{U} &= \{Q\}, \\ \text{Zeile 11:} \quad & u = Q, & \mathcal{U} &= \emptyset, \\ \text{Zeilen 12 – 18:} \quad & \mathcal{V}(A) = Q, & \mathcal{S}(A) &= 1, \\ & \mathcal{E}(A) = 3, & \mathcal{M} &= \{Q, A\}, \\ & \mathcal{U} = \{A\}, \end{aligned}$$

Zeile 11:	$u = A,$	$\mathcal{U} = \emptyset,$
Zeilen 12 – 18:	$\mathcal{V}(B) = A,$	$\mathcal{S}(B) = 1,$
	$\mathcal{E}(B) = 3,$	$\mathcal{M} = \{Q, A, B\},$
	$\mathcal{U} = \{B\},$	
Zeile 11:	$u = B,$	$\mathcal{U} = \emptyset,$
Zeilen 12 – 18:	$\mathcal{V}(C) = B,$	$\mathcal{S}(C) = 1,$
	$\mathcal{E}(C) = 2,$	$\mathcal{M} = \{Q, A, B, C\},$
	$\mathcal{U} = \{C\},$	$\mathcal{V}(R) = B,$
	$\mathcal{M} = \{Q, A, B, C\},$	$\mathcal{U} = \{C\},$
Zeilen 19 – 25:	$\mathcal{V}(D) = B,$	$\mathcal{S}(D) = -1,$
	$\mathcal{E}(D) = 3,$	
	$\mathcal{M} = \{Q, A, B, C, D\},$	
	$\mathcal{U} = \{C, D\},$	
Zeile 11:	$u = C,$	$\mathcal{U} = \{D\},$
Zeilen 12 – 18:	$\mathcal{V}(R) = C,$	$\mathcal{S}(R) = 1,$
	$\mathcal{E}(R) = 2,$	
	$\mathcal{M} = \{Q, A, B, C, D, R\},$	
	$\mathcal{U} = \{D, R\}.$	

Zeilen 26 – 32 liefern jetzt den in Abbildung II.4.9 angegebenen Fluss. Die Zahlen an den Kanten geben die jeweiligen Flusswerte bzw. in Klammern die Kapazität an.
Im nächsten Durchlauf erhalten wir:

Zeile 6:	$\mathcal{M} = \{Q\},$	$\mathcal{U} = \{Q\},$
Zeile 11:	$u = Q, \& \mathcal{U}$	$= \emptyset,$
Zeilen 12 – 18:	$\mathcal{V}(A) = Q,$	$\mathcal{S}(A) = 1,$
	$\mathcal{E}(A) = 1,$	$\mathcal{M} = \{Q, A\},$
	$\mathcal{U} = \{A\},$	
Zeile 11:	$u = A,$	$\mathcal{U} = \emptyset,$
Zeilen 12 – 18:	$\mathcal{V}(B) = A,$	$\mathcal{S}(B) = 1,$
	$\mathcal{E}(B) = 1,$	$\mathcal{M} = \{Q, A, B\},$
	$\mathcal{U} = \{B\},$	
Zeile 11:	$u = B,$	$\mathcal{U} = \emptyset,$
Zeilen 19 – 25:	$\mathcal{V}(D) = B,$	$\mathcal{S}(D) = -1,$
	$\mathcal{E}(D) = 3,$	$\mathcal{M} = \{Q, A, B, D\},$
	$\mathcal{U} = \{D\},$	

Zeile 11: $u = D$, $\mathcal{U} = \emptyset$,

Zeile 9: Fluss ist optimal.

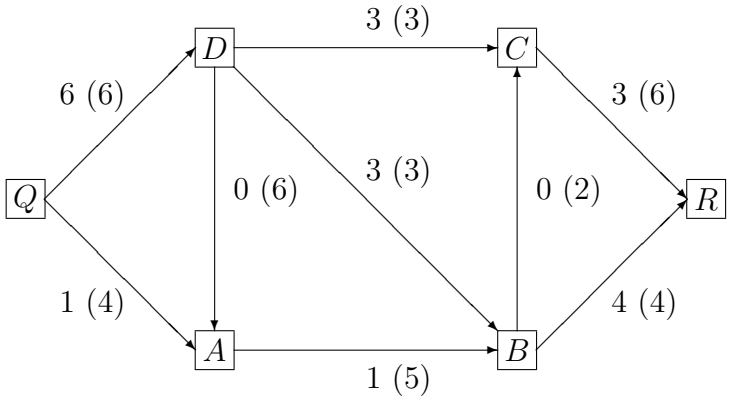

Abbildung II.4.8. Fluss nach dem dritten Erreichen von Zeilen 26 – 32 von Algorithmus II.4.1 mit jeweiligen Flusswerten bzw. in Klammern Kapazität

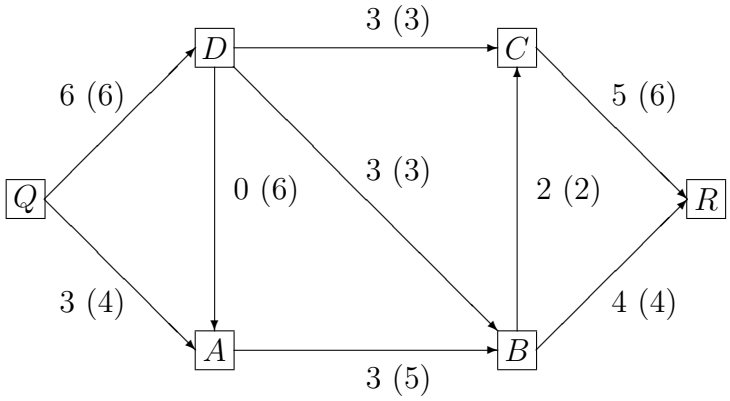

Abbildung II.4.9. Fluss nach dem vierten Erreichen von Zeilen 26 – 32 von Algorithmus II.4.1 mit jeweiligen Flusswerten bzw. in Klammern Kapazität; der Fluss ist optimal

Wir betrachten nun das Problem, einen zulässigen Fluss mit vorgegebenem Wert und minimalen Kosten zu finden. Dazu seien $G = (V, E)$, c ein Netzwerk, $s, t \in V$ ausgezeichnete Knoten, $w : E \to \mathbb{R}$

eine Kostenfunktion und $\Phi \in \mathbb{R}$ ein Flusswert. Dann lautet unser Problem

$$\min \left\{ \sum_{e \in E} w(e)x(e) : \quad x : E \to \mathbb{R} \text{ ist ein zulässiger } (s,t)\text{-Fluss mit Wert } \varphi(x) = \Phi \right\}.$$

Zur Lösung diese Problems benötigen wir einige Notationen und Hilfsmittel.

DEFINITION II.4.7. Gegeben seien ein Netzwerk $G = (V, E)$, c, ausgezeichnete Knoten $s, t \in V$ und ein zulässiger (s,t)-Fluss $x : E \to \mathbb{R}$. Weiter sei C ein Zyklus in G[5]. Sei $\mathcal{O}$ eine der beiden möglichen Orientierungen von C. Eine Kante von C in Richtung von $\mathcal{O}$ heißt dann *Vorwärtsbogen* bzgl. $\mathcal{O}$ und eine Kante gegen die Richtung von $\mathcal{O}$ *Rückwärtsbogen* bzgl. $\mathcal{O}$. C heißt *augmentierend* bzgl. x, wenn es eine Orientierung $\mathcal{O}$ von C gibt, so dass für alle Vorwärtsbögen e bzgl. $\mathcal{O}$ gilt

$$x(e) < c(e)$$

und für alle Rückwärtsbögen e' bzgl. $\mathcal{O}$

$$x(e') > 0.$$

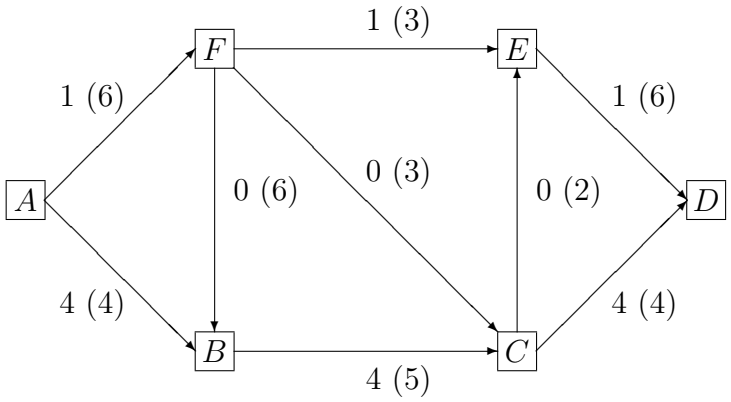

ABBILDUNG II.4.10. Netzwerk aus Beispiel II.4.8 mit jeweiligen Flusswerten bzw. in Klammern Kapazität

BEISPIEL II.4.8. Betrachte das Netzwerk aus Abbildung II.4.10, wobei die Zahlen die jeweiligen Flusswerte bzw. in Klammern die Kapazitäten angeben.
B, C, E, F, B ist ein Zyklus in G. Bezüglich dieser Orientierung ist

[5]Die Kantenorientierung von G wird also nicht notwendig berücksichtigt.

(F, E) ein Rückwärtsbogen und (B, C), (C, E), (F, B) sind Vorwärtsbögen. Dieser Zyklus ist augmentierend.
F, B, C, F ist auch ein Zyklus. Bezüglich dieser Orientierung ist (F, C) ein Rückwärtsbogen mit Flusswert 0. Bezüglich der entgegengesetzten Orientierung F, C, B, F ist (F, B) ein Rückwärtsbogen mit Flusswert 0. Also ist dieser Zyklus nicht augmentierend.

DEFINITION II.4.9. Seien $G = (V, E)$, c ein Netzwerk, $s, t \in V$ ausgezeichnete Knoten, $w : E \to \mathbb{R}$ eine Kostenfunktion und x ein zulässiger (s, t)-Fluss.

(1) $w(e)$ heißt *Kostenfaktor* der Kante $e \in E$.
(2) Sei P ein augmentierender Weg in G. Die *Kostenfunktion* von P ist definiert als die Summe der Kostenfaktoren aller Vorwärtsbögen abzüglich der Kostenfaktoren aller Rückwärtsbögen. Die *Kosten einer Augmentierung* sind definiert als das Produkt aus Kostenfaktor und Augmentierungswert.

SATZ II.4.10. *Ein Fluss x mit Wert Φ hat genau dann minimale Kosten, wenn es keinen augmentierenden Zyklus mit negativen Kosten gibt.*

BEWEIS. „$\Longleftarrow$“: Wir nehmen an, dass es einen Fluss y mit geringeren Kosten gibt. Setze

$$z = y - x.$$

Durch Induktion über die Kanten und Fallunterscheidung [1, Satz 25.5] kann man zeigen, dass z als Summe augmentierender Zyklen dargestellt werden kann. Da z negative Werte hat, muss ein augmentierender Zyklus mit negativen Kosten existieren. Dies ist ein Widerspruch.
„$\Longrightarrow$“: Wir nehmen an, C sei ein augmentierender Zyklus mit negativen Kosten. Dann setzen wir

$$\varepsilon = \min\Big\{\min\{c(e) - x(e) : e \in C \text{ ist Vorwärtsbogen}\},\\ \min\{x(e) : e \in C \text{ ist Rückwärtsbogen}\}\Big\}$$

und

$$y(e) = \begin{cases} x(e) & \text{falls } e \notin C, \\ x(e) + \varepsilon & \text{falls } e \in C \text{ ein Vorwärtsbogen}, \\ x(e) - \varepsilon & \text{falls } e \in C \text{ ein Rückwärtsbogen}. \end{cases}$$

Dann hat y geringere Kosten als x. Dies ist ein Widerspruch. □

SATZ II.4.11. *Sei x ein Fluss mit Wert Φ und minimalen Kosten. P sei ein (s, t)-augmentierender Weg mit minimalen Kosten und Augmentierungswert $\delta > 0$. Dann liefert die Augmentierung durch P einen Fluss y mit Wert $\Phi + \delta$ und minimalen Kosten.*

BEWEIS. Wir müssen zeigen, dass y minimale Kosten hat. Angenommen, dies sei nicht der Fall. Wegen Satz II.4.10 gibt es dann zu

y einen augmentierenden Zyklus C mit negativen Kosten. Mittels Induktion und Fallunterscheidung [1, Satz 25.6] zeigt man dann, dass C auch augmentierend zu x ist. Dies ist ein Widerspruch. □

DEFINITION II.4.12. Seien $G = (V, E)$ ein Graph, $c : E \to \mathbb{R}_+$ eine Kapazitätsfunktion, $w : E \to \mathbb{R}$ eine Kostenfunktion, $s, t \in V$ ausgezeichnete Knoten und $x : E \to \mathbb{R}$ ein zulässiger (s,t)-Fluss. Wir ordnen diesen Größen einen Graphen $G_x = (V_x, E_x)$ und Funktionen $c_x : E_x \to \mathbb{R}_+$ und $w_x : E_x \to \mathbb{R}$ wie folgt zu:

$$
\begin{aligned}
V_x &= V,\\
E_x &= \{e \in E : x(e) < c(e)\} \cup \{(v,u) : (u,v) \in E\ ,\ x((u,v)) > 0\},\\
c_x(e) &= \begin{cases} c(e) - x(e) & \text{falls } e \in E_x \text{ mit } c(e) > x(e),\\ x(e) & \text{falls } e \in E_x \text{ mit } x(e) > 0, \end{cases}\\
w_x &= \begin{cases} w(e) & \text{falls } e \in E_x \text{ mit } c(e) > x(e),\\ -w(e) & \text{falls } e \in E_x \text{ mit } x(e) > 0. \end{cases},
\end{aligned}
$$

(V_x, E_x, c_x, w_x) heißt das zu G und x gehörende *augmentierende Netzwerk*.

Aus Definition II.4.12 und Satz II.4.10 folgt:

LEMMA II.4.13. *(1) Jedem augmentierenden Zyklus in G entspricht ein (gerichteter) Kreis im augmentierenden Netzwerk.*
(2) Ein zulässiger Fluss ist genau dann kostenminimal, wenn es im zugehörigen augmentierenden Netzwerk keinen Kreis mit negativen Kosten gibt.

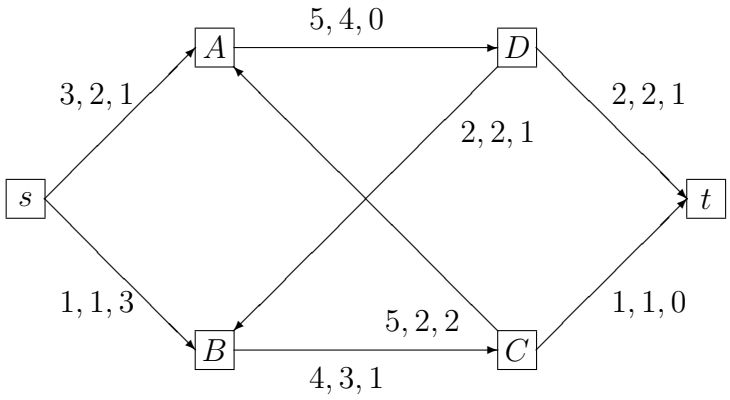

ABBILDUNG II.4.11. Graph aus Beispiel II.4.14 mit Kapazität, Flusswert und Kosten der jeweiligen Kanten

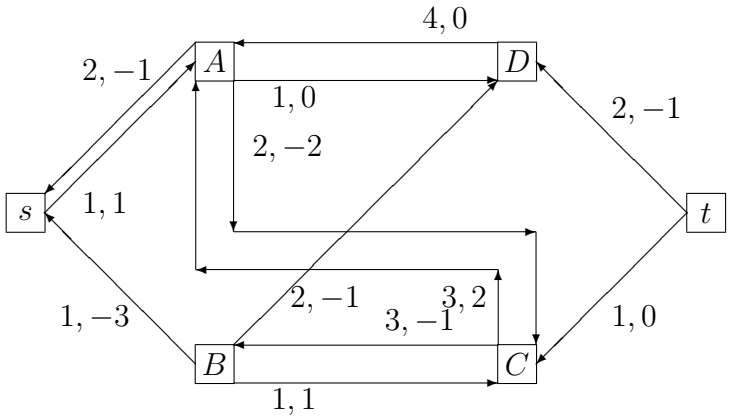

ABBILDUNG II.4.12. Augmentierendes Netzwerk zum Graphen aus Beispiel II.4.14 und Abbildung II.4.11 mit Kapazität und Kosten der jeweiligen Kanten

BEISPIEL II.4.14. Wir betrachten den Graphen aus Abbildung II.4.11, wobei die Zahlentripel an den Kanten jeweils die Kapazität, die Flusswerte und die Kosten angeben. Der Wert des Flusses ist 3, die Kosten betragen

$$\begin{aligned} &2 \cdot 1 + 4 \cdot 0 + 2 \cdot 1 + 1 \cdot 3 + 3 \cdot 1 + 1 \cdot 0 + 2 \cdot 1 + 2 \cdot 2 \\ &= 16. \end{aligned}$$

Abbildung II.4.12 zeigt das zugehörige augmentierende Netzwerk, wobei die Zahlentupel an den Kanten jeweils die Kapazität und die Kosten angeben. D, A, C, B, D ist ein Kreis mit minimaler Kapazität 2 und Kostenfaktor

$$0 - 2 - 1 - 1 = -4.$$

Er bringt eine Ersparnis von

$$2 \cdot (-4) = -8.$$

Der entsprechende augmentierte Fluss ist in Abbildung II.4.13 angegeben.

Algorithmus II.4.2 setzt diese Ergebnisse um und konstruiert zunächst augmentierende Zyklen mit negativen Kosten, so lange bis keine derartigen Zyklen mehr existieren. In dieser Phase werden jeweils Flusswerte und Kosten geändert. Falls am Ende der gewünschte Wert erreicht ist, ist der kostenminimale Fluss gefunden. Andernfalls werden in der zweiten Phase durch augmentierende Pfade die Flusswerte bei gleichbleibenden Kosten angepasst. In beiden Phasen müssen kürzeste-Wege-Probleme gelöst werden. In der zweiten Phase wird der Dijkstra-Algorithmus II.3.1 (S. 80) angewandt, obwohl negative Weglängen auftreten können. Dies ist erlaubt, da nach Abschluss der ersten Phase

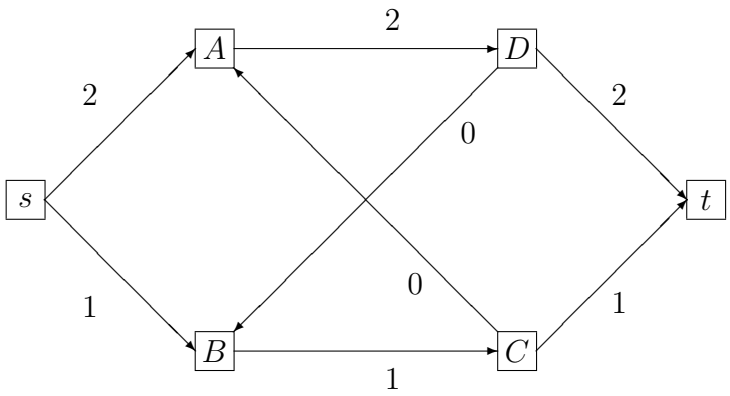

ABBILDUNG II.4.13. Augmentierter Fluss zu dem Graphen aus Beispiel II.4.14 und Abbildung II.4.11 mit dem augmentierenden Netzwerk aus Abbildung II.4.12

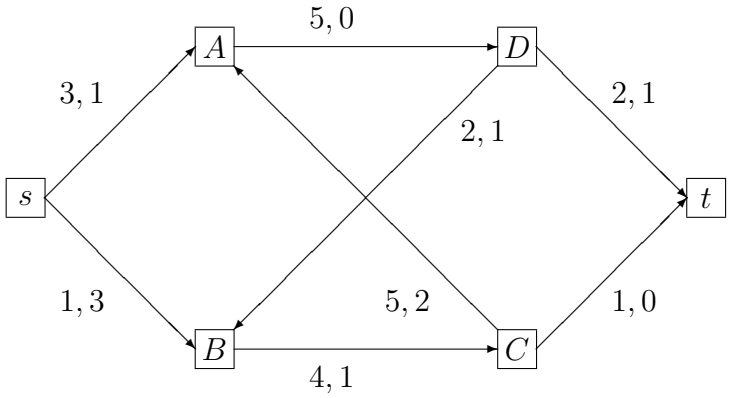

ABBILDUNG II.4.14. Netzwerk aus Beispiel II.4.15 mit Kapazitäten und Kosten der jeweiligen Kante

Zyklen negativer Länge wie in Beispiel II.3.4 (S. 82) und Abbildung II.3.2 (S. 82) ausgeschlossen sind.

BEISPIEL II.4.15. Wir betrachten das in Abbildung II.4.14 gezeigte Netzwerk, wobei die Tupel an den Kanten die Kapazitäten und Kosten angeben. Gesucht ist ein kostenminimaler (s, t)-Fluss mit Wert 3. Abbildung II.4.15 zeigt das zu $x = 0$ gehörende augmentierende Netzwerk, wobei die Zahlen an den Kanten die jeweiligen Kapazitäten und Kosten angeben. Der einzige Kreis in diesem Netzwerk ist A, D, B, C, A und hat bzgl. w_x die Länge 4. Damit gelangen wir zu Zeile 18 und erhalten den Weg s, A, D, t mit der Länge 2 bzgl. w_x. Es ist $\alpha' = 2$, $\alpha = 2$. Damit erhalten wir das in Abbildung II.4.16 gezeigte Netzwerk,

Algorithmus II.4.2 Bestimmung eines kostenminimalen Flusses mit vorgegebenem Wert

Gegeben: Graph $G = (V, E)$, Kapazität $c : E \to \mathbb{R}_+$, Kosten $w : E \to \mathbb{R}$, Knoten $s, t \in V$, Wert $\Phi \in \mathbb{R}$
Gesucht: Kostenminimaler zulässiger (s,t)-Fluss x mit Wert Φ

1: **for** $e \in E$ **do** $x(e) \leftarrow 0$
2: **end for**
3: **loop** ▷ wird in Zeilen 16 oder 20 verlassen
4: $(V_x, E_x, c_x, w_x) \leftarrow$ augmentierendes Netzwerk
5: Wende Algorithmus II.3.6 auf (V_x, E_x) mit $d = w_x$ an. ▷ Floyd-Warshall
6: **if** (V_x, E_x) enthält einen Kreis K negativer Länge **then**
7: $\varepsilon \leftarrow \min\{c_x(e) : e \in K\}$
8: **for** $e \in K \cap E$ **do** $x(e) \leftarrow x(e) + \varepsilon$
9: **end for**
10: **for** $e \in K \setminus E$ **do** $x(e) \leftarrow x(e) - \varepsilon$
11: **end for**
12: **else**
13: **loop** ▷ wird in Zeilen 16 oder 20 verlassen
14: $\varphi \leftarrow$ Wert von x
15: **if** $\varphi = \Phi$ **then**
16: **stop** ▷ x ist gesuchte Fluss
17: **else**
18: $P \leftarrow$ kürzeste (s,t)-Weg ▷ Algorithmus II.3.1 von Dijkstra für (V_x, E_x) mit $d = w_x$
19: **if** P existiert nicht **then**
20: **stop** ▷ gesuchte Fluss existiert nicht
21: **else**
22: $\alpha' \leftarrow \min\{c_x(e) : e \in P\}$, $\alpha \leftarrow \min\{\alpha', \Phi - \varphi(x)\}$
23: **for** $e \in P \cap E$ **do** $x(e) \leftarrow x(e) + \alpha$
24: **end for**
25: **for** $e \in P \setminus E$ **do** $x(e) \leftarrow x(e) - \alpha$
26: **end for**
27: $(V_x, E_x, c_x, w_x) \leftarrow$ augmentierendes Netzwerk
28: **end if**
29: **end if**
30: **end loop**
31: **end if**
32: **end loop**

wobei die Zahlen an den Kanten jeweils die Kapazität, den Flusswert und die Kosten angeben. Es ist $\varphi(x) = 2$ mit Kosten 4.

Abbildung II.4.17 zeigt das neue zugehörige augmentierende Netzwerk. Zeile 18 liefert den Weg s, A, D, B, C, t mit Länge 3 bzgl. w_x. Es

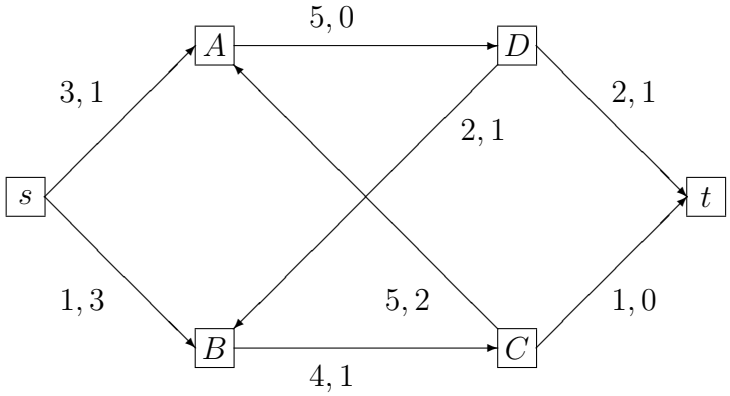

ABBILDUNG II.4.15. Augmentierendes Netzwerk zum Fluss $x = 0$ und dem Netzwerk aus Beispiel II.4.15 und Abbildung II.4.14 mit Kapazitäten und Kosten der jeweiligen Kante

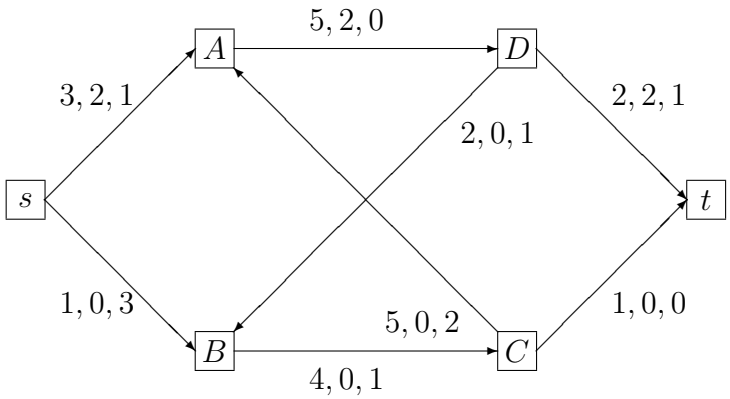

ABBILDUNG II.4.16. Netzwerk aus Beispiel II.4.15 mit Kapazität, Flusswert und Kosten der jeweiligen Kanten nach dem ersten Durchlauf von Algorithmus II.4.2

ist $\alpha' = 1$, $\alpha = 1$. Damit erhalten wir das in Abbildung II.4.18 gezeigte neue Netzwerk. Es ist $\varphi(x) = 3$ mit Kosten 7. Dies ist der gesuchte kostenminimale Fluss.

BEISPIEL II.4.16. Zur Verdeutlichung der Arbeitsweise von Algorithmus II.4.2 starten wir mit dem Netzwerk aus Beispiel II.4.14 und Abbildung II.4.11, d.h., wir überspringen die Initialisierungsphase von Algorithmus II.4.2 und starten mit dem „Gott gegebenen“ Fluss aus Beispiel II.4.14 und Abbildung II.4.11 (S. 98). Dann liefern die Zeilen

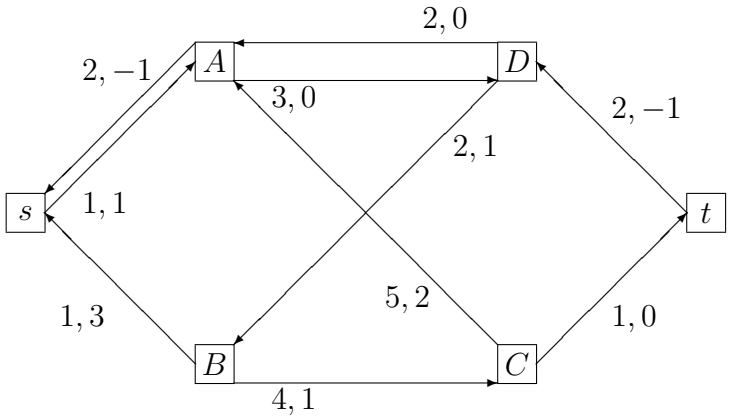

ABBILDUNG II.4.17. Augmentierendes Netzwerk beim zweiten Durchlauf von Algorithmus II.4.2 in Beispiel II.4.15 mit Kapazitäten und Kosten der jeweiligen Kante

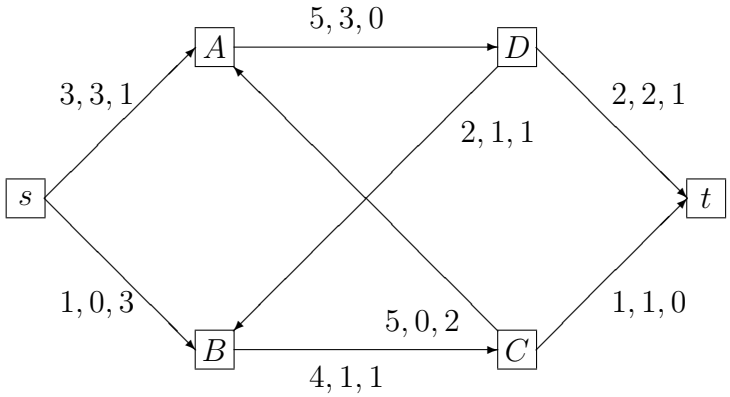

ABBILDUNG II.4.18. Netzwerk aus Beispiel II.4.15 mit Kapazität, Flusswert und Kosten der jeweiligen Kanten nach dem zweiten Durchlauf von Algorithmus II.4.2; der Fluss hat minimale Kosten und den vorgegebenen Wert 3

13 – 30 von Algorithmus II.4.2 den Fluss aus Beispiel II.4.14 und Abbildung II.4.13 (S. 100). Der Fluss hat den Wert 3 und die Kosten 8. Abbildung II.4.19 zeigt das zugehörige augmentierende Netzwerk. Es enthält folgende Kreise

s, A, D, B, s	Wert	-1,
s, A, D, B, C, A, s	Wert	4,
A, D, A	Wert	0,
B, C, B	Wert	0,

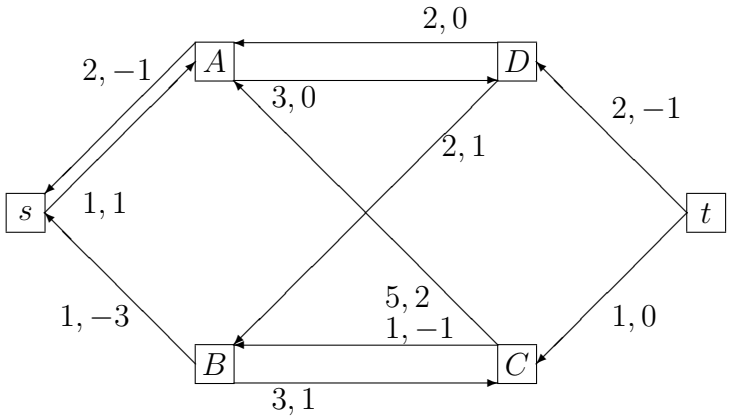

ABBILDUNG II.4.19. Augmentierendes Netzwerk beim ersten Durchlauf von Algorithmus II.4.2 in Beispiel II.4.16 mit Kapazitäten und Kosten der jeweiligen Kante

$$A, D, B, C, A \qquad \text{Wert} \qquad 4.$$

Also liefern Zeilen 7 – 11 den ersten dieser Kreise mit $\varepsilon = 1$. Das resultierende Netzwerk ist das kostenminimale Netzwerk aus Beispiel II.4.15 und Abbildung II.4.18.
Anders als in Beispiel II.4.15 wissen wir in der jetzigen Phase von Algorithmus II.4.2 noch nicht, dass dieses Netzwerk optimal ist. Daher bestimmt Algorithmus II.4.2 das zugehörige augmentierende Netzwerk, das in Abbildung II.4.20 gezeigt ist. Es enthält die Kreise

$$\begin{aligned} s, B, C, A, s \quad & \text{Wert} \quad 5, \\ B, C, B \quad & \text{Wert} \quad 0, \\ A, D, A \quad & \text{Wert} \quad 0, \\ B, D, B \quad & \text{Wert} \quad 0, \\ B, C, A, D, B \quad & \text{Wert} \quad 4. \end{aligned}$$

Keiner hat einen negativen Wert. Daher kommen wir zu Zeile 16 und der Algorithmus bricht wegen $\varphi(x) = 3$ mit der Information ab, dass der gesuchte Fluss gefunden ist.

BEMERKUNG II.4.17. Man kann in Zeile 5 von Algorithmus II.4.2 auch den Algorithmus II.4.1 nutzen. Dies hat den Nachteil höherer Kosten für diesen einzelnen Schritt. Andererseits hat es den Vorteil, dass die restlichen Schritte unter Umständen weniger häufig durchgeführt werden müssen. Dies ist z. B. der Fall, wenn der vorgegebene Wert Φ kleiner als der Wert des Maximalflusses ist.

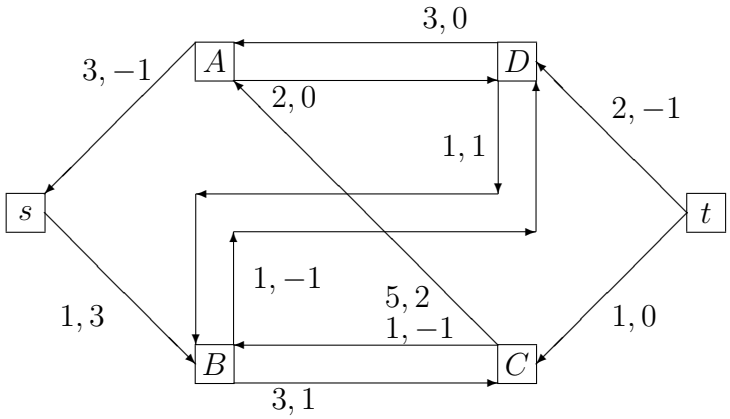

ABBILDUNG II.4.20. Augmentierendes Netzwerk beim zweiten Durchlauf von Algorithmus II.4.2 in Beispiel II.4.16 mit Kapazitäten und Kosten der jeweiligen Kante

KAPITEL III

Nichtlineare Optimierung

III.1. Minimierung ohne Nebenbedingungen

In diesem Abschnitt betrachten wir folgendes Problem:

> Gegeben seien eine nicht leere Menge $D \subset \mathbb{R}^n$ mit $n \geq 1$ und eine stetige Funktion $f : D \to \mathbb{R}$. Finde eine Minimum von f, d.h. löse

$$\min\{f(x) : x \in D\}. \tag{III.1.1}$$

Selbst für einfache Funktionen einer Variablen ist diese Aufgabe häufig nicht lösbar. Stattdessen müssen wir uns mit lokalen Minima zufrieden geben. Dabei heißt $x_0 \in D$ ein *lokales Minimum* von f, wenn es eine Umgebung U von x_0 gibt, so dass x_0 die Funktion f in $U \cap D$ minimiert, d.h. $f(y) \geq f(x_0)$ für alle $y \in U \cap D$.

Für differenzierbare Funktionen ist folgende Charakterisierung lokaler Minima wohl bekannt:

Satz III.1.1. *(1) Es sei D offen und $f \in C^1(D, \mathbb{R})$. Dann ist jedes lokale Minimum x^* von f auch ein* stationärer Punkt *von f, d.h. es gilt*

$$Df(x^*) = 0.$$

(2) Es sei D offen, $f \in C^2(D, \mathbb{R})$ und x^* ein stationärer Punkt von f. Dann ist x^* ein lokales Minimum von f, wenn die Hesse-Matrix $D^2 f(x^*)$ positiv definit ist.

Satz III.1.1 legt folgende Vorgehensweise zur Bestimmung lokaler Minima nahe:

> Bestimme mit dem Newton-Verfahren eine Nullstelle x_0 von Df und prüfe, ob $D^2 f(x_0)$ positiv definit ist.

Neben den bekannten Schwierigkeiten des Newton-Verfahrens ergeben sich mit diesem Ansatz weitere Probleme:

- Das Newton-Verfahren liefert allenfalls stationäre Punkte. Diese können auch Maxima oder Sattelpunkte sein.
- Der Nachweis der postiven Definitheit einer Matrix ist sehr aufwändig.
- Die Vorgehensweise erfordert zweimalige Differenzierbarkeit und die Kenntnis der zweiten Ableitungen.

Aus diesen Gründen interessieren wir uns auch für Verfahren, die ohne (höhere) Ableitungen auskommen. Außerdem möchten wir das Newton-Verfahren in eine größere Verfahrensklasse einbetten. Dies erlaubt dann die leichtere Analyse von Modifikationen des Newton-Verfahrens.

Wir beginnen mit dem einfachen Spezialfall von Funktionen einer Variablen, d.h. $n = 1$ und $D = [a, b]$. Dieser ist von Interesse, da fast alle Verfahren für den allgemeinen Fall $n \geq 2$ eindimensionale Minimierungsprobleme als Teilaufgaben enthalten.

LEMMA III.1.2. *Sei $a < x < b$, $f \in C([a, b], \mathbb{R})$ und*

$$f(x) \leq \min\{f(a), f(b)\}.$$

Dann besitzt f ein lokales Minimum $\eta \in (a, b)$. Falls f differenzierbar ist, ist $f'(\eta) = 0$.

BEWEIS. Da f stetig und $[a, b]$ kompakt ist, besitzt f ein globales Minimum $\mu \in [a, b]$.
Falls $\mu \in (a, b)$ ist, ist nichts zu zeigen.
Falls $\mu = a$ ist, folgt

$$f(\mu) \leq f(x) \quad \text{und} \quad f(x) \leq f(a) = f(\mu).$$

Also ist

$$f(x) = f(\mu)$$

und $x \in (a, b)$ ist ein lokales Minimum.
Analog argumentiert man im Fall $\mu = b$. □

Lemma III.1.2 führt auf folgenden Ansatz zur Bestimmung eines lokalen Minimums:

> Gegeben seien a, x und b wie in Lemma III.1.2. Bestimme den Mittelpunkt u des längeren der beiden Teilintervalle $[a, x]$ und $[x, b]$. Wähle aus den Punkten a, b, x, u drei Punkte $\widetilde{a}$, $\widetilde{b}$ und $\widetilde{x}$ so aus, dass $\widetilde{a} < \widetilde{x} < \widetilde{b}$ und
>
> $$f(\widetilde{x}) \leq \min\{f(\widetilde{a}), f(\widetilde{b})\}$$
>
> ist.

Dieser Ansatz führt auf Algorithmus III.1.1.

SATZ III.1.3. *Für die Folgen $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}}$ und $(x_k)_{k\in\mathbb{N}}$, die Algorithmus III.1.1 mit $\varepsilon = 0$ erzeugt, gilt:*

(1) *$a_k < x_k < b_k$ für alle k.*
(2) *$f(x_k) \leq \min\{f(a_k), f(b_k)\}$ für alle k.*
(3) *$b_k - a_k \leq (\frac{3}{4})^{k-1}(b_0 - a_0)$ für alle k.*
(4) *Falls f differenzierbar ist, ist*

$$\eta = \lim_{k\to\infty} a_k = \lim_{k\to\infty} x_k = \lim_{k\to\infty} b_k$$

ein stationärer Punkt.
Falls f zweimal stetig differenzierbar ist, ist $f''(\eta) \geq 0$.

Algorithmus III.1.1 Eindimensionale Minimierung

Gegeben: Punkte $a < x < b$ mit $f(x) \leq \min\{f(a), f(b)\}$, Toleranz $\varepsilon > 0$
Gesucht: Minimum x von f

```
while b − a > ε do
    if x ≤ ½(a + b) then
        u ← ½(x + b)
    else
        u ← ½(x + a)
    end if
    if f(x) ≤ f(u) then
        if x ≤ ½(a + b) then
            b ← u
        else
            a ← u
        end if
    else
        if x ≤ ½(a + b) then
            a ← x
        else
            b ← x
        end if
    end if
end while
```

Beweis. Die Punkte (1) und (2) folgen direkt aus der Konstruktion des Algorithmus.
Falls $f(x_0) \leq f(u_0)$ ist, folgt sofort

$$b_1 - a_1 \leq \frac{3}{4}(b_0 - a_0).$$

Falls $f(x_0) > f(u_0)$ ist, ist zwar $b_1 - a_1 < b_0 - a_0$, aber der Quotient $\frac{b_1 - a_1}{b_0 - a_0}$ kann beliebig nahe bei 1 sein. Allerdings ist jetzt $x_1 = \frac{1}{2}(a_1 + b_1)$ und im nächsten Schritt wird die Intervalllänge mindestens um den Faktor $\frac{3}{4}$ reduziert. Danach gilt diese Reduktion für jeden Schritt. Dies beweist die Behauptung (3).
Wegen (1) und (3) konvergieren die Folgen (a_k), (x_k) und (b_k) gegen denselben Grenzwert. Gemäß Lemma III.1.2 liegt in jedem Intervall (a_k, b_k) ein lokales Minimum μ_k. Dann gilt auch

$$\lim_{k \to \infty} \mu_k = \eta.$$

Falls f stetig differenzierbar ist, gilt $f'(\mu_k) = 0$ für alle k und daher $f'(\eta) = 0$. Falls f zweimal stetig differenzierbar ist, gilt $f''(\mu_k) \geq 0$ für alle k und daher $f''(\eta) \geq 0$. □

Wir wenden uns nun dem Fall mehrerer Veränderlicher zu, d.h. $D = \mathbb{R}^n$ mit $n \geq 2$. Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$, $x \in \mathbb{R}^n$ und $s \in \mathbb{R}^n$ mit $\|s\| = 1$, wobei $\|\cdot\|$ die euklidische Norm auf $\mathbb{R}^n$ bezeichnet. Wir betrachten die Funktion $\varphi : \mathbb{R}_+ \to \mathbb{R}$ mit

$$\varphi(t) = f(x + ts).$$

Offensichtlich gilt

$$\begin{aligned}
\varphi(0) &= f(x), \\
\varphi'(t) &= Df(x + ts)s, \\
\varphi'(0) &= Df(x)s, \\
\varphi''(t) &= s^t D^2 f(x + ts)s.
\end{aligned}$$

Also ist

$$\begin{aligned}
f(x + ts) &= \varphi(t) \\
&= \varphi(0) + t\varphi'(0) + \frac{1}{2}t^2\varphi''(\eta) \\
&= f(x) + tDf(x)s + \frac{1}{2}t^2 s^t D^2 f(x + \eta s)s
\end{aligned}$$

für ein $\eta \in (0, t)$. Falls $Df(x)s < 0$ ist, gilt daher $f(x + ts) < f(x)$ für hinreichend kleines $t > 0$. Andererseits ist $Df(x)s$ minimal für

$$s = -\frac{1}{\|Df(x)\|} Df(x).$$

Dies führt auf Algorithmus III.1.2, der eine ganze Verfahrensklasse beschreibt.

Algorithmus III.1.2 Abstiegsverfahren

Gegeben: Parameter $0 < c_1 \leq c_2 < 1$, $0 < \gamma \leq 1$, Vektor $x \in \mathbb{R}^n$, Toleranz $\varepsilon > 0$

Gesucht: stationärer Punkt x von f

1: **while** $\|Df(x)\| > \varepsilon$ **do**

2: $s \leftarrow$ Vektor in $\mathbb{R}^n$ mit $\|s\| = 1$ und $-Df(x)s \geq \gamma \|Df(x)\|$ ▷ Suchrichtung

3: $\lambda \leftarrow$ positive Zahl mit ▷ Schrittweite

(III.1.2)
$$\begin{aligned}
f(x + \lambda s) &\leq f(x) + \lambda c_1 Df(x)s \\
Df(x + \lambda s)s &\geq c_2 Df(x)s
\end{aligned}$$

4: $x \leftarrow x + \lambda s$

5: **end while**

Bemerkung III.1.4. (1) Es ist typischerweise $c_2 \leq \frac{1}{2}$.
(2) Algorithmus III.1.2 beschreibt eine ganze Verfahrensklasse, da die Wahl der Suchrichtungen s und die Bestimmung der Schrittweiten λ

nicht spezifiziert sind.
(3) Bei der Wahl

$$s = -\frac{1}{\|Df(x)\|} Df(x)$$

ist die Bedingung von Zeile 2 für jedes $\gamma \leq 1$ erfüllt.
(4) Je kleiner γ gewählt wird, desto größer ist der Bereich der zulässigen Suchrichtungen. Im Grenzfall $\gamma \to 0$ wird in Zeile 2 nur ausgeschlossen, dass s senkrecht ist zum negativen Gradienten $-Df(x)$.
(5) Die Wahl

$$s = -\frac{1}{\|Df(x)\|} Df(x)$$

liefert das *gedämpfte Newton-Verfahren*. Die Bedingung in Zeile 3 ist erfüllt, falls die Hesse-Matrizen $D^2 f(x)$ stets positiv definit sind und es eine gemeinsame untere Schranke für ihren kleinsten Eigenwert gibt.
(6) Für die oben eingeführte Hilfsfunktion φ bedeutet die Bedingung (III.1.2)

$$\begin{aligned} \varphi(\lambda) &\leq \varphi(0) + \lambda c_1 \varphi'(0) \\ \varphi'(\lambda) &\geq c_2 \varphi'(0). \end{aligned}$$

Es gibt verschiedene Strategien zur Realisierung von (III.1.2). Ebenso kann man (III.1.2) durch andere Bedingungen ersetzen. Zwei hiervon werden wir später kennen lernen.

Beispiel III.1.5. Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch positiv definit und $b \in \mathbb{R}^n$. Bekanntlich löst $x \in \mathbb{R}^n$ das LGS

$$Ax = b$$

genau dann, wenn es das eindeutige Minimum von

$$f(x) = \frac{1}{2} x^t A x - b^t x$$

ist.
Wählen wir in Algorithmus III.1.2

$$s = -\frac{1}{\|Df(x)\|} Df(x) = \frac{1}{\|b - Ax\|} (b - Ax)$$

und bestimmen λ so, dass es $t \mapsto f(x + ts)$ minimiert, erhalten wir das aus der Vorlesung „Einführung in die Numerik“ bekannte *Gradientenverfahren*. Zeile 2 gilt mit $\gamma = 1$. Eine leichte Rechnung zeigt, dass die Bedingung (III.1.2) mit $c_2 = \frac{1}{2}$ und beliebigem $c_1 \leq c_2$ erfüllt ist. Aus der Vorlesung „Einführung in die Numerik“ wissen wir, dass das Gradientenverfahren konvergiert und dass es gemessen in der Energienorm $(x^t A x)^{\frac{1}{2}}$ die Konvergenzrate

$$\frac{\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} - 1}{\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} + 1} = \frac{\lambda_{\max}(A) - \lambda_{\min}(A)}{\lambda_{\max}(A) + \lambda_{\min}(A)}$$

hat.
Ebenso kann man das *konjugierte Gradienten-Verfahren* als Algorithmus III.1.2 mit passender Wahl der Suchrichtungen interpretieren. Gleiches gilt für das *vorkonditionierte konjugierte Gradienten-Verfahren.*

Im Folgenden bezeichnet $\|\cdot\|$ stets die euklidische Norm auf $\mathbb{R}^n$ und $|||\cdot|||$ die zugehörige Matrixnorm

$$|||A||| = \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|}{\|x\|} = \left(\frac{\lambda_{\max}(A^tA)}{\lambda_{\min}(A^tA)}\right)^{\frac{1}{2}}.$$

LEMMA III.1.6. *Es sei* $f \in C^2(\mathbb{R}^n,\mathbb{R})$ *mit*

$$\inf_{x\in\mathbb{R}^n} f(x) > -\infty$$

und $0 < c_1 \le c_2 < 1$ *sowie* $0 < \gamma \le 1$. *Für ein* $x \in \mathbb{R}^n$ *und ein* $s \in \mathbb{R}^n$ *gelte*

$$Df(x) \neq 0, \quad \|s\| = 1, \quad -Df(x)s \ge \gamma\|Df(x)\|.$$

Dann gibt es ein kleinstes $\lambda^* > 0$, *das die zweite Bedingung von* (III.1.2) *erfüllt, d.h.*

$$Df(x+\lambda^* s)s = c_2 Df(x)s$$

und

$$Df(x+ts)s < c_2 Df(x)s$$

für alle $0 < t < \lambda^*$.
Für λ^* *gilt auch die erste Bedingung von* (III.1.2), *d.h.*

$$f(x+\lambda^* s) \le f(x) + \lambda^* c_1 Df(x)s.$$

Sei

$$L \ge \max_{0\le t\le\lambda^*} |||D^2 f(x+ts)|||.$$

Dann gilt für jedes λ, *das die Bedingung* (III.1.2) *erfüllt, die Abschätzung*

$$\begin{aligned} \inf_{t\ge 0} f(x+ts) &\le f(x+\lambda s) \\ &\le f(x) - \frac{c_1(1-c_2)\gamma^2}{L}\|Df(x)\|^2. \end{aligned} \tag{III.1.3}$$

BEWEIS. Definiere $\varphi : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi(t) = f(x+ts).$$

Wir zeigen zuerst, dass es ein $\lambda > 0$ gibt mit

$$\varphi'(\lambda) > c_2\varphi'(0).$$

Denn andernfalls würde für alle $t > 0$ gelten

$$\varphi'(t) \le c_2\varphi'(0).$$

Hieraus folgte

$$\varphi(\lambda) - \varphi(0) = \int_0^\lambda \varphi'(t)dt \leq \int_0^\lambda c_2\varphi'(0)dt = c_2\lambda\varphi'(0)$$
$$\underset{\lambda\to\infty}{\longrightarrow} -\infty$$

wegen $\varphi'(0) < 0$. Dies ist ein Widerspruch zu der Annahme

$$\inf_{x\in\mathbb{R}^n} f(x) > -\infty.$$

Da $\varphi'(0) < c_2\varphi'(0)$ ist, folgt aus dem soeben Gezeigten und der Stetigkeit von φ', dass es ein kleinstes $\lambda^* > 0$ gibt mit $\varphi'(\lambda^*) = c_2\varphi'(0)$. Dies beweist die Existenz von λ^*.
Weiter ist

$$\begin{aligned}\varphi(\lambda^*) &= \varphi(0) + \int_0^{\lambda^*} \varphi'(t)dt\\ &\leq \varphi(0) + \int_0^{\lambda^*} c_2\varphi'(0)dt\\ &= \varphi(0) + c_2\lambda^*\varphi'(0)\\ &\leq \varphi(0) + c_1\lambda^*\varphi'(0).\end{aligned}$$

Also erfüllt λ^* auch die zweite Bedingung von (III.1.2).
Wegen $\|s\| = 1$ gilt für alle $t \in [0, \lambda^*]$

$$\varphi''(t) = s^t D^2 f(x + ts)s \leq |||D^2 f(x + ts)||| \leq L.$$

Wegen

$$\varphi'(\lambda^*) = c_2\varphi'(0)$$

und

$$-\varphi'(0) = -Df(x)s \geq \gamma\|Df(x)\|$$

folgt hieraus

$$\begin{aligned}\lambda^* L &\geq \lambda^* \max_{0\leq t\leq\lambda^*} \varphi''(t)\\ &\geq \int_0^{\lambda^*} \varphi''(t)dt\\ &= \varphi'(\lambda^*) - \varphi'(0)\\ &= c_2\varphi'(0) - \varphi'(0)\\ &= (1 - c_2)(-\varphi'(0))\\ &\geq (1 - c_2)\gamma\|Df(x)\|.\end{aligned}$$

Da $\varphi'(t) = Df(x + ts)s$ nicht konstant ist, ist $L > 0$. Also können wir durch L dividieren und erhalten

$$\lambda^* \geq \frac{(1 - c_2)\gamma}{L}\|Df(x)\|.$$

Die erste Ungleichung in (III.1.3) ist offensichtlich erfüllt. Für jedes λ. das (III.1.2) erfüllt, gilt konstruktionsgemäß $\lambda \geq \lambda^*$. Damit erhalten wir für jedes derartige λ

$$\begin{aligned} f(x+\lambda s) &= \varphi(\lambda) \\ &\leq \varphi(0) + c_1\lambda \underbrace{\varphi'(0)}_{\leq -\gamma\|Df(x)\|} \qquad \text{(erste Bed. von (III.1.2))} \\ &\leq \varphi(0) - c_1\lambda\gamma\|Df(x)\| \\ &\leq \varphi(0) - c_1\lambda^*\gamma\|Df(x)\| \\ &\leq \varphi(0) - \frac{(1-c_2)c_1\gamma}{L}\|Df(x)\|^2 \\ &= f(x) - \frac{(1-c_2)c_1\gamma}{L}\|Df(x)\|^2. \end{aligned}$$

Dies beweist die zweite Ungleichung in (III.1.3). □

Satz III.1.7. *Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$, $x_0 \in \mathbb{R}^n$ und*

$$K = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$$

kompakt. Dann ist Algorithmus III.1.2 durchführbar. Er bricht entweder nach endlich vielen Schritten mit einem $x_k \in \mathbb{R}^n$ mit $Df(x_k) = 0$ ab, wobei gilt $f(x_k) < f(x_{k-1}) < \ldots < f(x_0)$, oder er erzeugt eine Folge $(x_k)_{k\in\mathbb{N}}$ mit folgenden Eigenschaften:

(1) *$(f(x_k))_{k\in\mathbb{N}}$ ist streng monoton fallend.*
(2) *$(x_k)_{k\in\mathbb{N}}$ besitzt mindestens einen Häufungspunkt x^*.*
(3) *Für jeden Häufungspunkt x^* von $(x_k)_{k\in\mathbb{N}}$ gilt $Df(x^*) = 0$.*

Beweis. Offensichtlich ist

$$\inf_{x\in\mathbb{R}^n} f(x) = \inf_{x\in K} f(x).$$

Da K kompakt ist, gilt

$$\inf_{x\in\mathbb{R}^n} f(x) > -\infty.$$

Wegen Lemma III.1.6 sind daher die Zeilen 2 und 3 von Algorithmus III.1.2 immer durchführbar, sofern Zeile 1 nicht zum Abbruch führt. Außerdem ist konstruktionsgemäß die Sequenz der Funktionswerte streng monoton fallend, so dass (1) gilt. Wir müssen also nur noch die Eigenschaften (2) und (3) beweisen.
Da $(x_k)_{k\in\mathbb{N}} \subset K$ und K kompakt ist, folgt (2).
Für den Nachweis von (3) sei

$$L = \max_{x\in K} |||D^2 f(x)|||.$$

Gemäß Lemma III.1.6 existiert für jedes k eine kleinste Zahl $\lambda_k^* > 0$, die die Bedingung (III.1.2) erfüllt. Daher gilt

$$\lambda_k \geq \lambda_k^* > 0$$

und

$$f(x_k + ts_k) \le f(x_k) \le f(x_0)$$

für alle k und alle $t \in [0, \lambda_k]$. Also ist

$$L \ge \sup_{k\in\mathbb{N}} \max_{0\le t\le\lambda_k} |||D^2 f(x_k + ts_k)|||.$$

Mit

$$\alpha = \frac{c_1(1-c_2)\gamma^2}{L}$$

folgt daher aus Lemma III.1.6

$$\begin{aligned} f(x_{k+1}) &\le f(x_k) - \alpha\|Df(x_k)\|^2 \\ &\le f(x_{k-1}) - \alpha\|Df(x_{k-1})\|^2 - \alpha\|Df(x_k)\|^2 \\ &\le f(x_0) - \alpha\sum_{j=0}^{k}\|Df(x_j)\|^2 \end{aligned}$$

und somit

$$\sum_{j=0}^{k}\|Df(x_j)\|^2 \le \frac{1}{\alpha}(f(x_0) - f(x_k)).$$

Da $(f(x_k))_{k\in\mathbb{N}}$ monoton fallend und durch $\inf_{x\in K} f(x) > -\infty$ nach unten beschränkt ist, ist die Reihe $\sum_j \|Df(x_j)\|^2$ beschränkt und damit $(\|Df(x_j)\|)_{j\in\mathbb{N}}$ eine Nullfolge. Dies beweist (3). □

KOROLLAR III.1.8. *Satz III.1.7 bleibt gültig, wenn in Algorithmus III.1.2 die Bedingung* (III.1.2) *zur Bestimmung der Schrittweite* λ_k *durch eine der beiden folgenden Bedingungen (B) oder (C) ersetzt wird:*

(B) (exakte Liniensuche)

$$\lambda_k = \operatorname{argmin}\{f(x_k + \lambda s_k) : \lambda > 0\}.$$

(C) (Armijo-Liniensuche) *Fixiere eine Konstante* $\sigma > 0$. *Im* k*-ten Schritt von Algorithmus III.1.2 bestimme eine* $\lambda^*_{k,0}$ *mit*

$$\lambda^*_{k,0} \ge \sigma\|Df(x_k)\|$$

und bestimme die kleinste natürliche Zahl j_k *mit*

$$f(x_k + 2^{-j_k}\lambda^*_{k,0}s_k) \le f(x_k) + 2^{-j_k}c_1 Df(x_k)s_k.$$

Setze

$$\lambda_k = 2^{-j_k}\lambda^*_{k,0}$$

oder

$$\lambda_k = \operatorname{argmin}\{f(x_k + 2^{-i}\lambda^*_{k,0}s_k) : 0 \le i \le j_k\}.$$

BEWEIS. *Strategie (B):* Die Existenz von λ_k folgt aus der Kompaktheit der Niveaumenge K. Wegen der ersten Ungleichung in (III.1.3) in Lemma III.1.6 bleibt der Beweis von Satz III.1.7 für diese Wahl von λ_k

gültig.
Strategie (C): Wir halten den Iterationsindex k fest und setzen wieder

$$\varphi(t) = f(x_k + ts_k).$$

Dann lautet die Bedingung für $j = j_k$

$$\varphi(2^{-j}\lambda_{k,0}^*) \leq \varphi(0) + c_1 2^{-j}\lambda_{k,0}^*\varphi'(0).$$

Angenommen, diese Bedingung wäre für alle $j \in \mathbb{N}$ verletzt. Dann würde gelten

$$\begin{aligned} & c_1 2^{-j}\lambda_{k,0}^*\varphi'(0) < \varphi(2^{-j}\lambda_{k,0}^*) - \varphi(0) \\ \iff \qquad & c_1\varphi'(0) < \frac{\varphi(2^{-j}\lambda_{k,0}^*) - \varphi(0)}{2^{-j}\lambda_{k,0}^*} \xrightarrow[j\to\infty]{} \varphi'(0). \end{aligned}$$

Dies ist aber ein Widerspruch zu $\varphi'(0) < 0$ und $0 < c_1 < 1$. Dies beweist die Existenz von j_k.
Um den Beweis von Satz III.1.7 zu übertragen, reicht es zu zeigen, dass es ein festes $\alpha > 0$ gibt mit

$$f(x_{k+1}) \leq f(x_k + 2^{-j_k}\lambda_{k,0}^* s_k) \leq f(x_k) - \alpha\|Df(x_k)\|^2. \tag{III.1.4}$$

Wir müssen hierzu zwei Fälle unterscheiden:
Fall $j_k = 0$: Wir haben dann

$$\begin{aligned} \varphi(\lambda_{k,0}^*) - \varphi(0) &\leq c_1\lambda_{k,0}^*\varphi'(0) \\ &\leq -c_1\sigma\|Df(x_k)\|\gamma\|Df(x_k)\| \\ &= -c_1\sigma\gamma\|Df(x_k)\|^2, \end{aligned}$$

was (III.1.4) beweist.
Fall $j_k > 0$: Gemäß Lemma III.1.6 existiert ein größtes $\lambda_k^* > 0$ mit $\varphi'(t) < c_2\varphi'(0)$ für alle $0 \leq t < \lambda_k^*$. Angenommen es wäre $2^{-(j_k-1)}\lambda_{k,0}^* \leq \lambda_k^*$. Dann würde gelten

$$\begin{aligned} \varphi(2^{-(j_k-1)}\lambda_{k,0}^*) - \varphi(0) &= \int_0^{2^{-(j_k-1)}\lambda_{k,0}^*} \varphi'(t)dt \\ &\leq \int_0^{2^{-(j_k-1)}\lambda_{k,0}^*} c_2\varphi'(0)dt \\ &= c_2 2^{-(j_k-1)}\lambda_{k,0}^*\varphi'(0) \\ &\leq c_1 2^{-(j_k-1)}\lambda_{k,0}^*\varphi'(0) \end{aligned}$$

Also würde die Auswahlbedingung schon für $j_k - 1$ erfüllt. Dies ist ein Widerspruch. Also ist $2^{-(j_k-1)}\lambda_{k,0}^* \geq \lambda_k^*$. Damit folgt aus Lemma III.1.6

$$2^{-j_k}\lambda_{k,0}^* = \frac{1}{2}2^{-(j_k-1)}\lambda_{k,0}^* \geq \frac{1}{2}\lambda_k^* \geq \frac{(1-c_2)\gamma}{2L}\|Df(x_k)\|.$$

Wie in Lemma III.1.6 liefert dies

$$\varphi(2^{-j_k}\lambda_{k,0}^*) - \varphi(0) \leq -\frac{c_1(1-c_2)\gamma}{2L}\|Df(x_k)\|^2,$$

was wiederum (III.1.4) beweist.
Damit folgt der Rest der Behauptung wie im Beweis von Satz III.1.7. □

III.2. Konvexität und Trennungssätze

Wir erinnern an die Definition I.2.3 (S. 16) der Konvexität von Mengen und Lemma I.2.5 (S. 16) der Durchschnittsstabilität der Konvexität.

Definition III.2.1. Sei $S \subset \mathbb{R}^n$ eine nicht leere Menge. Dann heißt

$$\operatorname{conv}(S) = \bigcap_{\substack{S \subset C \\ C \text{ konvex}}} C$$

die *konvexe Hülle* von S.

Satz III.2.2. *Es gilt folgende Charakterisierung der konvexen Hülle*

$$\operatorname{conv}(S) = \Big\{ x : \exists N \in \mathbb{N}\ ,\ \exists x_1, \dots, x_N \in S\ ,\ \exists \lambda_1 \geq 0, \dots, \lambda_N \geq 0 \\ \textit{mit } \sum_{i=1}^{N} \lambda_i = 1 \textit{ und } x = \sum_{i=1}^{N} \lambda_i x_i \Big\}.$$

Beweis. Bezeichne mit M die Menge auf der rechten Seite obiger Gleichung. Offensichtlich ist $S \subset M$. Wie man leicht nachrechnet, ist M konvex. Also gilt $\operatorname{conv}(S) \subset M$.
Seien nun andererseits $N \in \mathbb{N}$, $x_1, \dots, x_N \in S$ und $\lambda_1, \dots, \lambda_N$ mit $\lambda_i \geq 0$ und $\sum \lambda_i = 1$ beliebig. Dann gilt natürlich $x_1, \dots, x_N \in \operatorname{conv}(S)$ und daher

$$\sum_{i=1}^{N} \lambda_i x_i \in \operatorname{conv}(S).$$

Dies beweist $M \subset \operatorname{conv}(S)$. □

Satz III.2.3. *Die Menge $S \subset \mathbb{R}^n$ sei endlich. Dann ist ihre konvexe Hülle* $\operatorname{conv}(S)$ *kompakt.*

Beweis. Sei

$$S = \{x_1, \dots, x_m\}$$

und

$$T_m = \Big\{ \lambda \in \mathbb{R}^m : \lambda_1 \geq 0, \dots, \lambda_m \geq 0\ ,\ \sum_{i=1}^{m} \lambda_i = 1 \Big\}.$$

Da T_m offensichtlich beschränkt und abgeschlossen ist, ist T_m kompakt. Gemäß Satz III.2.2 ist die Abbildung

$$\lambda \mapsto \sum_{i=1}^{m} \lambda_i x_i$$
$$T_m \to \operatorname{conv}(S)$$

surjektiv. Offensichtlich ist sie auch stetig. Hieraus folgt die Behauptung. □

SATZ III.2.4 (Caratheodory). *Sei* $S \subset \mathbb{R}^n$ *und* $k = \dim(\mathrm{aff}(S))$. *Dann gilt*

$$\mathrm{conv}(S) = \Big\{ x : \exists x_0, \dots, x_k \in S \ , \ \exists \lambda_0 \geq 0, \dots, \lambda_k \geq 0 \\ \textit{mit } \sum_{i=0}^{k} \lambda_i = 1 \textit{ und } x = \sum_{i=0}^{k} \lambda_i x_i \Big\}.$$

Jeder Punkt $x \in \mathrm{conv}(S)$ *ist also Konvexkombination von höchstens* $k+1$ *Punkten aus* S.

BEWEIS. Sei $x \in \mathrm{conv}(S)$ beliebig und N die kleinste Zahl, so dass x die Konvexkombination von $N+1$ Punkten aus S ist, d.h.

$$x = \sum_{i=0}^{N} \lambda_i x_i, \qquad x_0, \dots, x_N \in S,$$

$$\lambda_0 \geq 0, \dots, \lambda_N \geq 0, \quad \sum_{i=0}^{N} \lambda_i = 1.$$

Da N minimal ist, sind alle λ_i so gar positiv. Wir nehmen nun an, dass $N > k$ ist. Wegen der Definition von k und den Definitionen I.2.8 (S. 17) von $\mathrm{aff}(S)$ und I.2.6 (S. 17) von $\dim(\mathrm{aff}(S))$ sind dann die Punkte $x_0, \dots, x_N$ affin abhängig, d.h., es gibt Zahlen $\mu_0, \dots, \mu_N$, die nicht alle verschwinden, mit

$$\sum_{i=0}^{N} \mu_i x_i = 0 \quad \text{und} \quad \sum_{i=0}^{N} \mu_i = 0.$$

Damit ist mindestens ein μ_i positiv. Setze

$$\alpha^* = \max\{\alpha \geq 0 : \lambda_i - \alpha \mu_i \geq 0 \text{ für alle } i\} = \min_{\mu_i > 0} \frac{\lambda_i}{\mu_i}.$$

Dann gilt

$$\sum_{i=0}^{N} (\lambda_i - \alpha^* \mu_i) x_i = 0,$$

$$\sum_{i=0}^{N} (\lambda_i - \alpha^* \mu_i) = 1,$$

$$\lambda_i - \alpha^* \mu_i \geq 0$$

für alle i und

$$\lambda_i - \alpha^* \mu_i = 0$$

für mindestens ein i. Dies ist ein Widerspruch zur Minimalität von N. □

DEFINITION III.2.5. Seien $a \in \mathbb{R}^n \backslash \{0\}$, $\alpha \in \mathbb{R}$ und

$$H = \{x \in \mathbb{R}^n : a^t x = \alpha\}$$

eine *Hyperebene* und

$$H_+ = \{x \in \mathbb{R}^n : a^t x \geq \alpha\}$$

und

$$H_- = \{x \in \mathbb{R}^n : a^t x \leq \alpha\}$$

die zugehörigen *Halbräume*. Weiter seien $K_1, K_2 \subset \mathbb{R}^n$ zwei Mengen. Wir sagen (vgl. Abbildung III.2.1):

(1) H *trennt* K_1 und K_2, wenn gilt $K_1 \subset H_-$ und $K_2 \subset H_+$, d.h. $a^t x_1 \leq \alpha \leq a^t x_2$ für alle $x_i \in K_i$.
(2) H *trennt* K_1 und K_2 *strikt*, wenn gilt $K_1 \subset \overset{\circ}{H}_-$ und $K_2 \subset \overset{\circ}{H}_+$, d.h. $a^t x_1 < \alpha < a^t x_2$ für alle $x_i \in K_i$.
(3) H *trennt* K_1 und K_2 *eigentlich*, wenn H die Mengen K_1 und K_2 trennt und $K_1 \cup K_2 \not\subset H$ ist, d.h., $a^t x_1 \leq \alpha \leq a^t x_2$ gilt für alle $x_i \in K_i$ und es gibt mindestens ein Paar $x_i \in K_i$ mit $a^t x_1 < a^t x_2$.

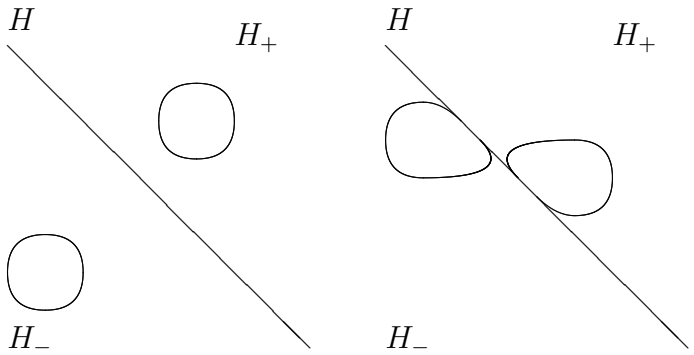

ABBILDUNG III.2.1. Strikte Trennung (links) und eigentliche Trennung (rechts)

BEISPIEL III.2.6. Betrachte

$$K_1 = \{(x, 0) \in \mathbb{R}^2 : x \geq -1\},$$
$$K_2 = \{(x, 0) \in \mathbb{R}^2 : x \leq 1\}.$$

Die einzige Hyperebene, die K_1 und K_2 trennt ist

$$H = \{(x, y) \in \mathbb{R}^2 : y = 0\}$$
$$= \{(x, y) \in \mathbb{R}^2 : \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)^t \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = 0\}.$$

H enthält aber K_1 und K_2. Die Trennung ist also nicht eigentlich.

SATZ III.2.7. *Sei $K \subset \mathbb{R}^n$ nicht leer, konvex und abgeschlossen mit $0 \notin K$. Dann kann $\{0\}$ strikt von K getrennt werden, d.h., es gibt ein $a \in \mathbb{R}^n \backslash \{0\}$ und ein $\alpha > 0$ mit $a^t x > \alpha$ für alle $x \in K$.*

BEWEIS. Sei $x^* \in K$ ein beliebiger Punkt. Nach Voraussetzung ist $\|x^*\| > 0$. Dann ist

$$\inf\{\|x\| : x \in K\} = \inf\{\|x\| : x \in K \cap \overline{B(0, \|x^*\|)}\}.$$

Da $K \cap \overline{B(0, \|x^*\|)}$ kompakt ist, wird das Infimum in einem Punkt $y \in K$ angenommen. Nach Voraussetzung ist $y \neq 0$. Für $x \in K$ und $\lambda \in \mathbb{R}$ sei

$$\begin{aligned} \varphi(\lambda) &= \|\lambda x + (1-\lambda)y\|^2 \\ &= (\lambda x + (1-\lambda)y)^t(\lambda x + (1-\lambda)y) \\ &= \lambda^2\|x-y\|^2 + 2\lambda(x-y)^t y + \|y\|^2. \end{aligned}$$

Für alle $\lambda \in [0,1]$ ist $\lambda x + (1-\lambda)y \in K$ und daher

$$\varphi(\lambda) \geq \|y\|^2 = \varphi(0).$$

Deshalb ist

$$2(x-y)^t y = \varphi'(0) \geq 0$$

und somit

$$x^t y \geq \|y\|^2.$$

Damit folgt die Behauptung mit $a = y$ und $\alpha = \frac{1}{2}\|y\|^2$. □

SATZ III.2.8. *Sei $K \subset \mathbb{R}^n$ nicht leer und konvex mit $0 \notin K$. Dann kann $\{0\}$ von K getrennt werden, d.h., es gibt ein $a \in \mathbb{R}^n \backslash \{0\}$ mit $a^t x \geq 0$ für alle $x \in K$.*

BEWEIS. Setze

$$S = \{x \in \mathbb{R}^n : \|x\| = 1\}$$

und definiere für $x \in K$ die Mengen A_x und C_x durch

$$\begin{aligned} A_x &= \{y \in S : y^t x \geq 0\} \\ C_x &= \mathbb{R}^n \backslash A_x. \end{aligned}$$

1. Beh.: Sind $x_1, \ldots, x_k \in K$ mit k beliebig, so ist

$$A_{x_1} \cap \ldots \cap A_{x_k} \neq \emptyset.$$

Bew.: Sei

$$P = \operatorname{conv}(\{x_1, \ldots, x_k\}).$$

Gemäß Satz III.2.3 ist P kompakt. Wegen $P \subset K$ und $0 \notin K$ ist $0 \notin P$. Wegen Satz III.2.7 gibt es daher ein $y^* \in S$ mit $y^{*t} x > 0$ für alle $x \in P$. Insbesondere gilt $y^{*t} x_i > 0$ für alle i. Also ist $y^* \in A_{x_1} \cap \ldots \cap A_{x_k}$.
2. Beh.: Es gibt ein $y \in S$ mit

$$y \notin \bigcup_{x \in K} C_x.$$

Bew.: Wir nehmen an, es sei

$$S \subset \bigcup_{x \in K} C_x.$$

Da S kompakt ist, gibt es ein $k \in \mathbb{N}$ und $x_1, \ldots, x_k \in K$ mit

$$S \subset \bigcup_{i=1}^{k} C_{x_i}.$$

Also gilt

$$S \cap A_{x_1} \cap \ldots \cap A_{x_k} = \emptyset.$$

Konstruktionsgemäß ist aber $A_x \subset S$ für jedes x. Gemäß der 1. Beh. ist $A_{x_1} \cap \ldots \cap A_{x_k} \neq \emptyset$. Dies liefert einen Widerspruch.
Gemäß der 2. Beh. gibt es ein $y \in S$ mit

$$y \notin \bigcup_{x \in K} C_x.$$

Für dieses y gilt

$$y \in \bigcap_{x \in K} A_x.$$

Dies beweist die Behauptung mit $a = y$. □

Satz III.2.9. *Seien K_1 und K_2 zwei nicht leere, disjunkte, konvexe Mengen in $\mathbb{R}^n$. Dann gibt es eine Hyperebene H, die K_1 und K_2 trennt.*

Beweis. Sei

$$K = K_2 - K_1 = \{x = y_2 - y_1 : y_i \in K_i\}.$$

Wie man sicht leicht überlegt, ist K konvex. Wegen $K_1 \cap K_2 = \emptyset$ ist $0 \notin K$. Gemäß Satz III.2.8 gibt es ein $a \in \mathbb{R}^n \backslash \{0\}$ mit $a^t x \geq 0$ für alle $x \in K$. Setze

$$\alpha = \sup_{y \in K_1} a^t y.$$

Dann ist $\alpha < \infty$ und $H = \{y : a^t y = \alpha\}$ leistet das Gewünschte. □

Für die strikte Trennung von konvexen Mengen benötigen wir den Begriff des relativen Innern einer konvexen Menge. Dieser benutzt die affine Hülle, die wir in Definition I.2.8 (S. 17) eingeführt haben.

Definition III.2.10. Sei $K \subset \mathbb{R}^n$ konvex. Das relative Innere (im topologischen Sinne) des Abschlusses $\overline{K}$ von K bzgl. $\mathrm{aff}(K)$ heißt das *relative Innere* von K und wird mit K^I bezeichnet.

Aus der Definition des topologischen relativen Inneren einer Menge bzgl. einer anderen Menge folgt:

Bemerkung III.2.11. Es ist $x \in \mathbb{R}^n$ in K^I genau dann, wenn gilt $x \in \overline{K}$ und es gibt ein $\varepsilon > 0$ mit $B(x, \varepsilon) \cap \mathrm{aff}(K) \subset K$.

Beispiel III.2.12. Betrachte

$$K = \{x \in \mathbb{R}^2 : x_2 = 0\ ,\ 0 \le x_1 < 1\}.$$

Offensichtlich ist K konvex, $\overset{\circ}{K} = \emptyset$ und

$$\overline{K} = \{x \in \mathbb{R}^2 : x_2 = 0\ ,\ 0 \le x_1 \le 1\}$$
$$\mathrm{aff}(K) = \{x \in \mathbb{R}^2 : x_2 = 0\}.$$

Es ist

$$K^I = \{x \in \mathbb{R}^2 : x_2 = 0\ ,\ 0 < x_1 < 1\}.$$

Satz III.2.13. *Für jede nicht leere konvexe Menge $K \subset \mathbb{R}^n$ ist $K^I \neq \emptyset$.*

Beweis. Sei

$$m = \dim(\mathrm{aff}(K)).$$

Falls $m = 0$ ist, ist $K = \{x_0\}$ und somit $\mathrm{aff}(K) = K$ und $K^I = K$. Also ist o.E. $m \ge 1$. Daher gibt es $m+1$ Punkte $x_0, \dots, x_m \in K$, so dass jedes $x \in \mathrm{aff}(K)$ darstellbar ist in der Form

$$x = \sum_{i=0}^{m} \lambda_i x_i \quad \text{mit} \quad \sum_{i=0}^{m} \lambda_i = 1.$$

Diese Darstellung ist eindeutig. Also ist das LGS

$$\begin{pmatrix} 1 & \dots & 1 \\ x_0 & \dots & x_m \end{pmatrix} \begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

für jedes $x \in \mathrm{aff}(K)$ eindeutig lösbar. Daher hat die Matrix

$$M = \begin{pmatrix} 1 & \dots & 1 \\ x_0 & \dots & x_m \end{pmatrix} \in \mathbb{R}^{(n+1)\times(m+1)}$$

den maximalen Rang $m+1$ und die Matrix M^tM ist regulär. Setze

$$\overline{x} = \frac{1}{m+1} \sum_{i=0}^{m} x_i.$$

Offensichtlich ist $\overline{x} \in K$. Wir wollen zeigen, dass $\overline{x} \in K^I$ ist. Dazu bezeichnen wir mit $\|\cdot\|_\infty$ die Maximumnorm auf $\mathbb{R}^{n+1}$ und mit $|||\cdot|||_\infty$ die zugehörige Matrixnorm. Setze

$$\varepsilon = \left[(m+1)|||(M^tM)^{-1}M^t|||_\infty\right]^{-1}.$$

Sei nun $\widetilde{x} \in \mathrm{aff}(K)$ mit $\|\widetilde{x} - \overline{x}\|_\infty < \varepsilon$ beliebig. Dann ist

$$\widetilde{x} = \sum_{i=0}^{m} \widetilde{\lambda}_i x_i$$

mit

$$M \begin{pmatrix} \widetilde{\lambda}_0 \\ \vdots \\ \widetilde{\lambda}_m \end{pmatrix} = \begin{pmatrix} 1 \\ \widetilde{x} \end{pmatrix}$$

bzw.

$$\begin{pmatrix} \widetilde{\lambda}_0 \\ \vdots \\ \widetilde{\lambda}_m \end{pmatrix} = (M^t M)^{-1} M^t \begin{pmatrix} 1 \\ \widetilde{x} \end{pmatrix}.$$

Wegen

$$\begin{pmatrix} \overline{\lambda}_0 \\ \vdots \\ \overline{\lambda}_m \end{pmatrix} = \frac{1}{m+1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = (M^t M)^{-1} M^t \begin{pmatrix} 1 \\ \overline{x} \end{pmatrix}$$

folgt

$$\begin{aligned} \left\| \begin{pmatrix} \widetilde{\lambda}_0 \\ \vdots \\ \widetilde{\lambda}_m \end{pmatrix} - \begin{pmatrix} \overline{\lambda}_0 \\ \vdots \\ \overline{\lambda}_m \end{pmatrix} \right\|_\infty &= \left\| (M^t M)^{-1} M^t \left(\begin{pmatrix} 1 \\ \widetilde{x} \end{pmatrix} - \begin{pmatrix} 1 \\ \overline{x} \end{pmatrix} \right) \right\|_\infty \\ &\leq |||(M^t M)^{-1} M^t|||_\infty \|\widetilde{x} - \overline{x}\|_\infty \\ &\leq \varepsilon |||(M^t M)^{-1} M^t|||_\infty \\ &= \frac{1}{m+1}. \end{aligned}$$

Also ist $\widetilde{\lambda}_i \geq 0$ für alle i und somit $\widetilde{x} \in K$. Dies beweist

$$B(\overline{x}, \varepsilon) \cap \operatorname{aff}(K) \subset K$$

und damit $\overline{x} \in K^I$. □

LEMMA III.2.14. *Sei $K \subset \mathbb{R}^n$ konvex, $y \in \overline{K}$ und $x \in K^I$. Dann ist die Menge*

$$[x, y) = \{(1-\lambda)x + \lambda y : 0 \leq \lambda < 1\}$$

in K^I enthalten.

BEWEIS. Nach Definition von K^I gibt es ein $\varepsilon > 0$ mit

$$B(x, \varepsilon) \cap \operatorname{aff}(K) \subset K.$$

Sei $z = (1-\lambda)x + \lambda y \in [x, y)$ beliebig. Setze

$$\widetilde{\varepsilon} = \frac{1-\lambda}{1+\lambda} \varepsilon.$$

Wegen $\lambda < 1$ ist $\widetilde{\varepsilon} > 0$. Wegen $y \in \overline{K}$ gilt $y \in K + B(0, \widetilde{\varepsilon})$ und somit

$$\begin{aligned} B(z, \widetilde{\varepsilon}) \cap \operatorname{aff}(K) &= [z + B(0, \widetilde{\varepsilon})] \cap \operatorname{aff}(K) \\ &= [(1-\lambda)x + \lambda y + B(0, \widetilde{\varepsilon})] \cap \operatorname{aff}(K) \\ &\subset [\lambda K + \lambda B(0, \widetilde{\varepsilon}) + (1-\lambda)x + B(0, \widetilde{\varepsilon})] \cap \operatorname{aff}(K) \\ &= [\lambda K + (1-\lambda)x + (1+\lambda)B(0, \widetilde{\varepsilon})] \cap \operatorname{aff}(K) \\ &= \lambda K + [(1-\lambda)x + (1-\lambda)B(0, \frac{1+\lambda}{1-\lambda}\widetilde{\varepsilon})] \cap \operatorname{aff}(K) \\ &= \lambda K + [(1-\lambda)x + (1-\lambda)B(0, \varepsilon)] \cap \operatorname{aff}(K) \\ &= \lambda K + (1-\lambda)[x + B(0, \varepsilon)] \cap \operatorname{aff}(K) \\ &= \lambda K + (1-\lambda)B(x, \varepsilon) \cap \operatorname{aff}(K) \\ &\subset \lambda K + (1-\lambda)K \\ &\subset K. \end{aligned}$$

Also ist $z \in K^I$. □

SATZ III.2.15. *Sei $K \subset \mathbb{R}^n$ konvex. Dann sind folgende Aussagen äquivalent:*

(1) $x \in K^I$.
(2) *Zu jedem $y \in \operatorname{aff}(K)$ gibt es ein $\varepsilon > 0$ mit $x \pm \varepsilon(y - x) \in K$.*

BEWEIS. (1) $\Longrightarrow$ (2): Nach Voraussetzung gibt es ein $\widetilde{\varepsilon} > 0$ mit

$$B(x, \widetilde{\varepsilon}) \cap \operatorname{aff}(K) \subset K.$$

Sei $y \in \operatorname{aff}(K)$ beliebig. O.E. ist $y \neq x$. Setze

$$\varepsilon = \frac{\widetilde{\varepsilon}}{2\|x - y\|}.$$

Dann ist

$$x \pm \varepsilon(y - x) \in B(x, \widetilde{\varepsilon})$$

und

$$x \pm \varepsilon(y - x) \in \operatorname{aff}(K).$$

Also gilt

$$x \pm \varepsilon(y - x) \in B(x, \widetilde{\varepsilon}) \cap \operatorname{aff}(K) \subset K.$$

(2) $\Longrightarrow$ (1): Gemäß Satz III.2.13 gibt es ein $y \in K^I$. Nach Voraussetzung gibt es ein $\varepsilon > 0$ mit

$$\overline{x} = x - \varepsilon(y - x) \in K.$$

Offensichtlich ist

$$x = \frac{1}{1+\varepsilon}\overline{x} + \frac{\varepsilon}{1+\varepsilon}y$$

und daher $x \in [\overline{x}, y)$. Aus Lemma III.2.14 folgt $x \in K^I$. □

SATZ III.2.16. *Seien K_1 und K_2 konvexe Teilmengen des $\mathbb{R}^n$. Dann existiert genau dann eine Hyperebene H, die K_1 und K_2 eigentlich trennt, wenn $K_1^I \cap K_2^I = \emptyset$ ist.*

BEWEIS. „$\Longrightarrow$“ Sei

$$H = \{x \in \mathbb{R}^n : a^t x = \alpha\}$$

eine Hyperebene, die K_1 und K_2 eigentlich trennt. Wir zeigen, dass für alle $x_1 \in K_1^I$ und alle $x_2 \in K_2^I$ gilt

$$a^t x_1 < a^t x_2.$$

Dies beweist $K_1^I \cap K_2^I = \emptyset$.
Zum Beweis dieser Behauptung nehmen wir das Gegenteil an, d.h., es gebe ein Paar $x_i \in K_i^I$ mit

$$a^t x_1 = a^t x_2.$$

Da H die Mengen K_1 und K_2 eigentlich trennt, gibt es $\overline{x}_i \in K_i$ mit

$$a^t \overline{x}_1 < \alpha < a^t \overline{x}_2.$$

Gemäß Satz III.2.16 gibt es ein $\varepsilon > 0$ mit

$$y_1 = x_1 - \varepsilon(\overline{x}_1 - x_1) \in K_1$$

und

$$y_2 = x_2 - \varepsilon(\overline{x}_2 - x_2) \in K_2.$$

Dann folgt

$$\begin{aligned} a^t(y_1 - y_2) &= (1+\varepsilon)a^t(x_1 - x_2) - \varepsilon a^t(\overline{x}_1 - \overline{x}_2) \\ &= -\varepsilon a^t(\overline{x}_1 - \overline{x}_2) \\ &> 0 \end{aligned}$$

im Widerspruch zur Trennung von K_1 und K_2 durch H.
„$\Longleftarrow$“ Nach Voraussetzung ist $0 \notin K_1^I - K_2^I$. Wie man sich leicht überlegt, ist

$$K_1^I - K_2^I = (K_1 - K_2)^I.$$

Setze

$$X = \operatorname{aff}(K_1 - K_2).$$

X ist isomorph zu $\mathbb{R}^m$ mit $m = \dim X$ und $(K_1 - K_2)^I$ ist das topologische Innere von $K_1 - K_2$ in X. Da $K_1 - K_2$ konvex ist, gibt es gemäß Satz III.2.7 eine Hyperebene H in X, die in X den Punkt 0 von $(K_1 - K_2)^I$ trennt. Da $(K_1 - K_2)^I$ in X offen ist, ist diese Trennung strikt. Ebenso folgt, dass die Trennung (ohne Striktheit!) für $K_1 - K_2$ in X gilt. Dies beweist, dass K_1 und K_2 innerhalb von X eigentlich durch H getrennt werden. Indem wir die Hyperebene orthogonal auf ganz $\mathbb{R}^n$ fortsetzen, erhalten wir die eigentliche Trennung in $\mathbb{R}^n$. □

DEFINITION III.2.17. (1) Eine Menge $K \subset \mathbb{R}^n$ heißt ein *Kegel*, wenn für alle $x \in K$ und alle $\lambda \geq 0$ gilt $\lambda x \in K$.

(2) Für eine beliebige Menge $A \subset \mathbb{R}^n$ bezeichnen wir mit $\operatorname{cone}(A)$ den kleinsten konvexen Kegel, der A enthält, d.h.

$$\operatorname{cone}(A) = \bigcap_{\substack{C \text{ konvexer Kegel} \\ A \subset C}} C.$$

BEMERKUNG III.2.18. (1) Ein Kegel K ist genau dann konvex, wenn für alle $x, y \in K$ gilt $x + y \in K$.
(2) Es gilt folgende Charakterisierung

$$\operatorname{cone}(A) = \Big\{ x : \exists N \in \mathbb{N}\ ,\ \exists x_1, \dots, x_N \in A\ ,\ \exists \lambda_1 \geq 0, \dots, \lambda_N \geq 0 \text{ mit } x = \sum_{i=1}^{N} \lambda_i x_i \Big\}.$$

BEWEIS. *ad (1):* „$\Longrightarrow$“: Sei K ein konvexer Kegel und $x, y \in K$. Da K konvex ist, ist

$$\frac{1}{2}x + \frac{1}{2}y \in K.$$

Da K ein Kegel ist, ist

$$x + y = 2\left(\frac{1}{2}x + \frac{1}{2}y\right) \in K.$$

„$\Longleftarrow$“: Sei K ein Kegel und $x, y \in K$. Für beliebiges $\lambda \in [0, 1]$ gilt dann

$$\lambda x \in K, \quad (1 - \lambda)y \in K.$$

Also ist nach Voraussetzung

$$\lambda x + (1 - \lambda)y \in K$$

und damit K konvex.
ad (2): Der Beweis ist analog zu dem von Satz III.2.2. □

DEFINITION III.2.19. Seien $(\cdot, \cdot)$ ein beliebiges Skalarprodukt auf $\mathbb{R}^n$ und $A \subset \mathbb{R}^n$ eine beliebige Menge. Dann heißt

$$A^P = \{y \in \mathbb{R}^n : (y, x) \leq 0 \text{ für alle } x \in A\}$$

der bzgl. $(\cdot, \cdot)$ zu A *polare Kegel*. Die Menge

$$A^D = -A^P$$

heißt der *duale Kegel* zu A (vgl. Abbildung III.2.2).

SATZ III.2.20. *Es gilt:*

(1) A^P *ist ein abgeschlossener konvexer Kegel.*
(2) $A_1 \subset A_2 \Longrightarrow A_1^P \supset A_2^P$.
(3) $A^{PP} = \overline{\operatorname{cone}(A)}$.
(4) $A^{PP} = A \iff A$ *ist ein abgeschlossener Kegel.*
(5) $\left(\overline{\operatorname{cone}(A)}\right)^P = A^P$.

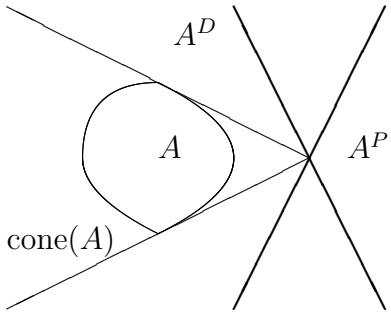

ABBILDUNG III.2.2. Polarer und dualer Kegel einer Menge

(6) *Ist A ein Untervektorraum von $\mathbb{R}^n$, so ist $A^P = A^\perp$.*

BEWEIS. *ad (1):* Gemäß Definition III.2.19 ist

$$A^P = \bigcap_{x\in A} \{y \in \mathbb{R}^n : (y,x) \leq 0\}.$$

Für jedes $x \in A$ ist $\{y \in \mathbb{R}^n : (y,x) \leq 0\}$ ein Halbraum und damit ein abgeschlossener konvexer Kegel. Hieraus folgt die Behauptung.
ad (2): Ist offensichtlich.
ad (3): „$\overline{\text{cone}(A)} \subset A^{PP}$“: Sei $x \in A$. Dann ist

$$(y,x) \leq 0$$

für alle $y \in A^P$ und daher $x \in A^{PP}$. Dies beweist $A \subset A^{PP}$. Wegen (1) folgt hieraus $\overline{\text{cone}(A)} \subset A^{PP}$.
„$A^{PP} \subset \overline{\text{cone}(A)}$“: Angenommen, es gebe ein $x_0 \in A^{PP} \backslash \overline{\text{cone}(A)}$. Gemäß Satz III.2.7[1] gibt es dann ein $a \in \mathbb{R}^n \backslash \{0\}$ und ein $\alpha \in \mathbb{R}$ mit

$$(a,x_0) > \alpha \geq (a,x)$$

für alle $x \in \overline{\text{cone}(A)}$. Da $\overline{\text{cone}(A)}$ ein nicht leerer Kegel ist, folgt $\alpha \geq 0$. Wegen $A \subset \overline{\text{cone}(A)}$ gilt

$$(a,x) \leq 0 \leq \alpha$$

für alle $x \in A$. Also ist $a \in A^P$. Wegen $x_0 \in A^{PP}$ folgt

$$(a,x_0) \leq 0.$$

Dies ist ein Widerspruch.
ad (4): Folgt aus (3).

[1] Satz III.2.7 ist für das euklidische Skalarprodukt formuliert. Jedes Skalarprodukt $(\cdot,\cdot)$ auf $\mathbb{R}^n$ ist von der Form $(x,y) = x^t H y$ mit einer symmetrisch positiv definiten Matrix H. Für jedes solche H gibt es eine reguläre Matrix C mit $H = C^t C$. Die Transformation $x \mapsto Cx$ reduziert daher den allgemeinen Fall auf den Fall des euklidischen Skalarproduktes.

ad (5)*:* Aus (3), (1) und (4) folgt

$$\left(\overline{\operatorname{cone}(A)}\right)^P \overset{(3)}{=} \left(A^{PP}\right)^P = \left(A^P\right)^{PP} \underset{(4)}{\overset{(1)}{=}} A^P.$$

ad (6)*:* Ist A ein Untervektorraum, so ist mit x auch $-x$ in A. Also gilt für alle $x \in A$ und alle $y \in A^P$

$$(y, x) \geq 0$$

und

$$-(y, x) = (y, -x) \geq 0$$

also

$$(y, x) = 0. \qquad \square$$

Satz III.2.21. *Sei $M \subset \mathbb{R}^n$ konvex und $f : M \to \mathbb{R}$ konvex. Dann ist f auf dem Innern $\overset{\circ}{M}$ von M stetig.*

Beweis. Ist $\overset{\circ}{M} = \emptyset$, ist nichts zu zeigen. Sei also $\overset{\circ}{M} \neq \emptyset$ und $x_0 \in \overset{\circ}{M}$. Dann ist insbesondere $\operatorname{aff}(M) = \mathbb{R}^n$. Daher gibt es $n+1$ affin unabhängige Punkte $y_0, \ldots, y_n \in M$ mit

$$x_0 = \frac{1}{n+1} \sum_{i=0}^{n} y_i.$$

Setze

$$Q = \operatorname{conv}(y_0, \ldots, y_n).$$

Für $x = \sum \lambda_i y_i \in Q$ ist wegen der Konvexität von f

$$f(x) = f\left(\sum_{i=0}^{n} \lambda_i y_i\right) \leq \sum_{i=0}^{n} \lambda_i f(y_i) \leq \max_{0 \leq i \leq n} f(y_i).$$

Also ist f auf Q durch

$$M = \max_{0 \leq i \leq n} f(y_i)$$

beschränkt. Da x_0 ein innerer Punkt von Q ist, gibt es ein $\varepsilon > 0$ mit $B(x_0, \varepsilon) \subset Q$. Seien $y \in B(0, \varepsilon)$ und $\sigma \in [0, 1]$ beliebig. Dann ist $x_0 \pm \sigma y \in Q$. Wegen

$$x_0 \pm \sigma y = \sigma(x_0 \pm y) + (1 - \sigma) x_0$$

und $x_0 \pm y \in Q$ folgt

$$f(x_0 \pm \sigma y) \leq \sigma f(x_0 \pm y) + (1 - \sigma) f(x_0) \leq M$$

und somit

$$f(x_0 \pm \sigma y) - f(x_0) \leq \sigma(f(x_0 \pm y) - f(x_0)) \leq \sigma(M - f(x_0)).$$

Wegen

$$x_0 = \frac{\sigma}{1+\sigma}(x_0 \mp y) + \frac{1}{1+\sigma}(x_0 \pm \sigma y)$$

ist auch

$$f(x_0) \leq \frac{\sigma}{1+\sigma} f(x_0 \mp y) + \frac{1}{1+\sigma} f(x_0 \pm \sigma y) \leq M.$$

Multipliziert man dies mit $1+\sigma$, so folgt

$$\begin{aligned} & (1+\sigma) f(x_0) \leq \sigma f(x_0 \mp y) + f(x_0 \pm \sigma y) \\ \Longrightarrow \quad & \sigma(f(x_0) - f(x_0 \mp y)) \leq f(x_0 \pm \sigma y) - f(x_0) \\ \Longrightarrow \quad & f(x_0 \pm \sigma y) - f(x_0) \geq \sigma(f(x_0) - f(x_0 \mp y)) \\ & \geq \sigma(f(x_0) - M). \end{aligned}$$

Also ist

$$|f(x_0 \pm \sigma y) - f(x_0)| \leq \sigma |M - f(x_0)|.$$

Dies beweist die Stetigkeit von f in x_0. □

III.3. Optimalitätskriterien für konvexe Probleme

Wie man sich als Übungsaufgabe leicht überlegt, besitzt eine differenzierbare konvexe Funktion $f : \mathbb{R}^n \to \mathbb{R}$ genau dann in x^* ein Minimum, wenn x^* ein kritischer Punkt ist, d.h., wenn $Df(x^*) = 0$ ist.

Wir wollen dieses Ergebnis auf Probleme mit Nebenbedingungen übertragen. Wir betrachten daher im Folgenden die Aufgabenstellung:

> *Problem (CP):* Gegeben sind eine nicht leere konvexe Menge $C \subset \mathbb{R}^n$, eine konvexe Funktion $f : C \to \mathbb{R}$, $p \geq 0$ konvexe Funktionen $f_1, \ldots, f_p : C \to \mathbb{R}$ und $m - p \geq 0$ affine Funktionen $f_{p+1}, \ldots, f_m : C \to \mathbb{R}$. Gesucht ist ein Minimum von f in C unter den Nebebedingungen
>
> $$\begin{aligned} f_i(x) &\leq 0 \quad 1 \leq i \leq p, \\ f_j(x) &= 0 \quad p+1 \leq j \leq m. \end{aligned}$$

Wir lassen dabei bewusst die Fälle $p = 0$, d.h. keine Ungleichungsnebenbedingungen, und $m = p$, d.h. keine Gleichungsnebenbedingungen, zu.

Zur Vereinfachung setzen wir

$$\begin{aligned} F_1(x) &= \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix}, \\ F_2(x) &= \begin{pmatrix} f_{p+1}(x) \\ \vdots \\ f_m(x) \end{pmatrix}, \\ F(x) &= \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix}, \\ S &= \{x \in C : F_1(x) \leq 0 \ , \ F_2(x) = 0\}. \end{aligned}$$

Dann kann das Problem *(CP)* in der kompakten Form

$$\min\{f(x) : x \in S\} \tag{III.3.1}$$

geschrieben werden.

Wie man sich leicht überlegt, ist die Menge S konvex, sofern sie nicht leer ist.

SATZ III.3.1. *Sei $C \subset \mathbb{R}^n$ eine nicht leere konvexe Menge, $F : C \to \mathbb{R}^m$ und jede Komponentenfunktion von F konvex. Dann sind folgende Aussagen äquivalent:*

(1) *Es gibt kein $x \in C$ mit $F(x) < 0$.*
(2) *Es gibt ein $z \in \mathbb{R}^m \backslash \{0\}$ mit $z \geq 0$ und*

$$z^t F(x) \geq 0$$

für alle $x \in C$.

BEWEIS. (2) $\Longrightarrow$ (1): Dies ist wegen $z \geq 0$ offensichtlich.
(1) $\Longrightarrow$ (2): Setze

$$A = \{v \in \mathbb{R}^m : \exists x \in C \text{ mit } F(x) < v\}.$$

Da die Komponentenfunktionen von F konvex sind, ist A konvex. Offensichtlich ist $A \neq \emptyset$[2]. Nach Voraussetzung ist $0 \notin A$. Gemäß Satz III.2.8 (S. 120) kann 0 von A getrennt werden, d.h., es gibt ein $z \in \mathbb{R}^m \backslash \{0\}$ mit

$$z^t v \geq z^t 0 = 0$$

für alle $v \in A$. Seien nun $v \in A$, $w \in \mathbb{R}^m$ mit $w \geq 0$ und $\lambda \geq 0$. Dann ist offensichtlich

$$v + \lambda w \in A$$

und daher

$$z^t(v + \lambda w) \geq 0.$$

Da dies für alle $\lambda \geq 0$ gilt, folgt

$$z^t w \geq 0.$$

Da $w \geq 0$ beliebig ist, folgt

$$z \geq 0.$$

Weiter gibt es zu jedem $x \in C$ ein $v \in A$, das beliebig nahe bei $F(x)$ liegt. Daher gilt mit

$$z^t v \geq 0$$

auch

$$z^t F(x) \geq 0. \qquad \square$$

LEMMA III.3.2. *Sei $C \subset \mathbb{R}^n$ nicht leer und konvex und $F : C \to \mathbb{R}^m$ mit konvexen, nicht positiven Komponentenfunktionen. Dann sind folgende Bedingungen äquivalent:*

[2] Betrachte ein $x_0 \in C$ und setze $v = F(x_0) + e$ mit $e = (1, \ldots, 1)^t$. Dann ist $v \in A$.

(1) *Zu jedem* $i \in \{1, \dots, m\}$ *gibt es ein* $x_i \in C$ *mit*

$$F_i(x_i) < 0.$$

(2) (Slater Bedingung) *Es gibt ein* $x \in C$ *mit*

$$F(x) < 0.$$

BEWEIS. (2) $\Longrightarrow$ (1): Ist offensichtlich.
(1) $\Longrightarrow$ (2): Setze

$$\overline{x} = \frac{1}{m} \sum_{i=1}^{m} x_i.$$

Dann gilt für jedes $j \in \{1, \dots, m\}$ wegen der Konvexität von F_j

$$\begin{aligned} F_j(\overline{x}) &\leq \frac{1}{m} \sum_{i=1}^{m} F_j(x_i) \\ &= \frac{1}{m} \sum_{\substack{1 \leq i \leq m \\ i \neq j}} \underbrace{F_j(x_i)}_{\leq 0} + \frac{1}{m} \underbrace{F_j(x_j)}_{<0} \\ &< 0. \end{aligned}$$

□

SATZ III.3.3. *Folgende Voraussetzungen seien erfüllt:*

(1) *Es gibt ein* $x^* \in S \cap C^I$.
(2) *Zu jeder nicht affinen Funktion* f_i *gibt es ein* $x_i \in S$ *mit* $f_i(x_i) < 0$.
(3) f *ist auf* S *nach unten beschränkt, d.h.*

$$\alpha = \inf\{f(x) : x \in S\} > -\infty.$$

Dann gibt es ein $y \in \mathbb{R}^m$ *mit* $y_i \geq 0$ *für* $1 \leq i \leq p$ *und*

$$f(x) + y^t F(x) \geq \alpha \tag{III.3.2}$$

für alle $x \in C$.

BEWEIS. O.E. ist $\alpha = 0$, sonst gehen wir zu $f(x) - \alpha$ über.
Wir machen zunächst folgende einschränkende Voraussetzung:

(2a) Zu jedem $i \in \{1, \dots, p\}$ gibt es ein $x_i \in S$ mit $f_i(x_i) < 0$.

Am Ende des Beweises werden wir zeigen, wie Bedingung (2a) durch die Bedingung (2) ersetzt werden kann.
Gemäß Lemma III.3.2 mit $F = F_1$ und $C = S$ gibt es ein $x^{**} \in S$ mit $F_1(x^{**}) < 0$. Wie im Beweis von Lemma III.3.2 zeigt man, dass man sogar $x^{**} \in S \cap C^I$ wählen kann.
1. Schritt: Wir setzen $f_0 = f$ und zeigen, dass es ein $z = (z_0, \dots, z_m)^t \in \mathbb{R}^{m+1}$ gibt mit $z_i \geq 0$ für alle $i \in \{0, \dots, p\}$ und

$$\sum_{i=0}^{m} z_i f_i(x) \geq 0$$

für alle $x \in C$.
Dazu definieren wir

$$A = \Big\{ v = (v_0, \ldots, v_m)^t \in \mathbb{R}^{m+1} : \exists x \in C \text{ mit } \begin{array}{l} v_0 > f_0(x), \\ v_i \geq f_i(x), 1 \leq i \leq p, \\ v_j = f_j(x), p+1 \leq j \leq m \Big\}. \end{array}$$

Der Vektor

$$v^* = \begin{pmatrix} f_0(x^{**}) + 1 \\ f_1(x^{**}) \\ \vdots \\ f_p(x^{**}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{m+1}$$

ist offensichtlich in A enthalten. Wegen der Konvexität der Funktionen $f_0, \ldots, f_p$ und der Affinität der Funktionen $f_{p+1}, \ldots, f_m$ ist A konvex. Wegen der Voraussetzung (3) und $\alpha = 0$ ist $0 \notin A$. Gemäß Satz III.2.16 (S. 124) kann A eigentlich von 0 getrennt werden. Es gibt also einen Vektor $z = (z_0, \ldots, z_m)^t \in \mathbb{R}^{m+1} \backslash \{0\}$ mit $z^t v \geq 0$ für alle $v \in A$ und ein $\overline{v} \in A$ mit $z^t \overline{v} > 0$. Wie im Beweis von Satz III.3.1 folgt $z_i \geq 0$ für $0 \leq i \leq p$. Ebenso gilt für alle $x \in C$ und alle $\varepsilon > 0$

$$v_\varepsilon = \begin{pmatrix} f_0(x) + \varepsilon \\ F(x) \end{pmatrix} \in A.$$

Also gilt für alle $x \in C$ und alle $\varepsilon > 0$

$$0 \leq z^t v_\varepsilon = z_0(f_0(x) + \varepsilon) + \sum_{i=1}^{m} z_i f_i(x).$$

Durch Grenzübergang $\varepsilon \to 0$ folgt die Zwischenbehauptung.
2. Schritt: Sei z der Vektor aus Schritt 1. Wir wollen zeigen, dass $z_0 > 0$ ist.
Für den Vektor v^* aus dem ersten Schritt gilt $z^t v^* \geq 0$. Wäre $z_0 = 0$, so folgte wegen $f_i(x^{**}) < 0$ für $1 \leq i \leq p$ und $z_i \geq 0$ für $1 \leq i \leq p$ die Beziehung

$$z_0 = z_1 = \ldots = z_p = 0.$$

Aus der Definiton von A folgt dann

$$\sum_{j=p+1}^{m} z_j f_j(x) \geq 0$$

für alle $x \in C$. Da 0 eigentlich von A getrennt werden kann, gibt es ein $\overline{x} \in C$ mit

$$\sum_{j=p+1}^{m} z_j f_j(\overline{x}) > 0.$$

Wegen $x^{**} \in C^I$ ist $x^{**} - \varepsilon(\overline{x} - x^{**}) \in C$ für hinreichend kleines $\varepsilon > 0$. Da $f_{p+1}, \ldots, f_m$ affin sind, folgt für $p+1 \le j \le m$

$$\begin{aligned} f_j(x^{**} - \varepsilon(\overline{x} - x^{**})) &= (1+\varepsilon) \underbrace{f_j(x^{**})}_{=0} - \varepsilon f_j(\overline{x}) \\ &= -\varepsilon f_j(\overline{x}). \end{aligned}$$

Also ist

$$\begin{aligned} \sum_{j=p+1}^{m} z_j f_j(x^{**} - \varepsilon(\overline{x} - x^{**})) &= -\varepsilon \sum_{j=p+1}^{m} z_j f_j(\overline{x}) \\ &< 0. \end{aligned}$$

Dies ist ein Widerspruch. Also muss $z_0 > 0$ sein.
3. Schritt: Sei wieder z der Vektor aus dem 1. Schritt. Wegen $z_0 > 0$ gilt

$$0 \le \sum_{i=0}^{m} z_i f_i(x) = z_0 \left\{ f_0(x) + \sum_{i=1}^{m} \frac{z_i}{z_0} f_i(x) \right\}$$

für alle $x \in C$. Also leistet der Vektor

$$y = \begin{pmatrix} \frac{z_1}{z_0} \\ \vdots \\ \frac{z_m}{z_0} \end{pmatrix}$$

das Gewünschte.
4. Schritt: Wir wollen nun die Bedingung (2a) zur Bedingung (2) abschwächen. Dazu bezeichnen wir mit $I \subset \{1, \ldots, p\}$ die Menge der Indizes, für die f_i affin ist und (2a) nicht gilt. Diese Indizes schlagen wir zu den Indizes $p+1$, …, m. Damit erhalten wir die Existenz eines $y \in \mathbb{R}^m \backslash \{0\}$ mit $y_i \ge 0$ für $i \in \{1, \ldots, p\} \backslash I$, das (III.3.2) erfüllt. Wir müssen noch erreichen, dass auch $y_i \ge 0$ gilt für alle $i \in I$. Die Menge aller y, die (III.3.2) erfüllen, ist konvex. Die Annahme, dass für alle diese y für mindestens ein $i \in I$ die Ungleichung $y_i < 0$ gilt, führt dann aber zu einem Widerspruch zum Trennungssatz III.2.16 (S. 124). □

Beispiel III.3.4. Sei $n = m = p = 1$, $C = \mathbb{R}_+$, $f(x) = -\sqrt{x}$ für $x \in C$ und $f_1(x) = x$. Dann ist

$$S = \{x \in C : x \le 0\} = \{0\}$$

und somit

$$S \cap C^I = \emptyset.$$

Das Problem (III.3.1) besitzt das eindeutige Optimum $x^* = 0$ und der Optimalwert ist 0.
Es gibt aber kein $y \geq 0$ mit

$$-\sqrt{x} + yx \geq 0$$

für alle $x \geq 0$.
Denn $y = 0$ erfüllt diese Bedingung offensichtlich nicht und für $y > 0$ und

$$x_y = \frac{1}{4y^2}$$

erhalten wir

$$-\sqrt{x_y} + yx_y = -\frac{1}{2y} + \frac{1}{4y} = -\frac{1}{4y} < 0.$$

Diese Beispiel zeigt, dass man auf die Voraussetzung (1) in Satz III.3.3 nicht verzichten kann.

Beispiel III.3.5. Sei $n = m = p = 1$, $C = \mathbb{R}$, $f(x) = x$ und $f_1(x) = x^2$. Offensichtlich ist $S = \{0\}$ und das Problem (III.3.1) besitzt das eindeutige Optimum $x^* = 0$ mit dem Optimalwert 0. Die Voraussetzung (2) von Satz III.3.3 ist aber verletzt.
Offensichtlich gibt es kein $y \geq 0$ mit

$$x + yx^2 \geq 0$$

für alle $x \in \mathbb{R}$.
Dieses Beispiel zeigt, dass man auf die Voraussetzung (2) in Satz III.3.3 nicht verzichten kann.

Satz III.3.6. *Die Voraussetzungen und Bezeichnungen seien wie in Satz III.3.3. Zusätzlich sei $C = \mathbb{R}^n$ und die Funktionen f, f_1, …, f_m seien differenzierbar. Außerdem besitze das Problem (III.3.1) eine Lösung x^*. Dann besitzt das System*

$$\begin{aligned} Df(x) + \sum_{i=1}^{m} y_i Df_i(x) &= 0, \\ f_i(x)y_i &= 0, \quad 1 \leq i \leq p, \\ f_i(x) &\leq 0, \quad 1 \leq i \leq p, \\ y_i &\geq 0, \quad 1 \leq i \leq p, \\ f_j(x) &= 0, \quad p+1 \leq j \leq m \end{aligned} \tag{III.3.3}$$

eine Lösung. Die Bedingungen (III.3.3) heißen Karush-Kuhn-Tucker-Bedingungen*, kurz* KKT-Bedingungen.

Beweis. Gemäß Satz III.3.3 gibt es ein $y^* \in \mathbb{R}^m$ mit $y_i^* \geq 0$ für $1 \leq i \leq p$ und

$$f(x) + \sum_{i=1}^{m} y_i^* f_i(x) \geq f(x^*)$$

für alle $x \in \mathbb{R}^n$. Setze

$$\varphi(x) = f(x) - f(x^*) + \sum_{i=1}^{m} y_i^* f_i(x).$$

Die Funktion φ ist konvex und differenzierbar. Für alle $x \in \mathbb{R}^n$ ist $\varphi(x) \geq 0$. Außerdem ist $\varphi(x^*) \leq 0$. Also ist x^* das eindeutige Minimum von φ und es gilt

$$0 = D\varphi(x^*) = Df(x^*) + \sum_{i=1}^{m} y_i^* Df_i(x^*).$$

Also erfüllt (x^*, y^*) die erste Gleichung von (III.3.3). Die letzten drei Gleichungen sind konstruktionsgemäß erfüllt. Die zweite Gleichung ist erfüllt, da wegen der Stetigkeit von φ die Beziehung $\varphi(x^*) = 0$ gilt. □

DEFINITION III.3.7. (1) Sei

$$D = \{y \in \mathbb{R}^m : y_i \geq 0 \text{ für } 1 \leq i \leq p\}.$$

Dann heißt die Funktion $\mathcal{L} : C \times D \to \mathbb{R}$, die durch

$$\mathcal{L}(x, y) = f(x) + \sum_{i=1}^{m} y_i f_i(x) = f(x) + y^t F(x)$$

definiert ist, die *Lagrange-Funktion* zu Problem (III.3.1).
(2) Ein Punkt $(x^*, y^*) \in C \times D$ heißt *Sattelpunkt* von $\mathcal{L}$, wenn für alle $(x, y) \in C \times D$ gilt

$$\mathcal{L}(x, y^*) \geq \mathcal{L}(x^*, y^*) \geq \mathcal{L}(x^*, y).$$

SATZ III.3.8. *Die Voraussetzungen und Bezeichnungen seien wie in Satz III.3.3. Dann ist $x^* \in C$ genau dann eine Optimallösung von (III.3.1), wenn es ein $y^* \in D$ gibt, so dass (x^*, y^*) ein Sattelpunkt der Lagrange-Funktion $\mathcal{L}$ ist.*

BEWEIS. „$\Longrightarrow$“: Gemäß Satz III.3.3 gibt es ein $y^* \in D$ mit

$$\mathcal{L}(x, y^*) = f(x) + y^{*t} F(x) \geq f(x^*)$$

für alle $x \in C$. Für $x = x^*$ folgt hieraus

$$y^{*t} F(x^*) \geq 0.$$

Wegen

$$\begin{pmatrix} f_{p+1}(x^*) \\ \vdots \\ f_m(x^*) \end{pmatrix} = F_2(x^*) = 0$$

folgt

$$\sum_{i=1}^{p} y_i^* f_i(x^*) \geq 0.$$

Wegen $y_i^* \geq 0$ und $f_i(x^*) \leq 0$ für $1 \leq i \leq p$ folgt

$$y_i^* f_i(x^*) = 0$$

für $1 \leq i \leq p$. Also ist

$$y^{*t} F(x^*) = 0$$

und damit

$$f(x^*) = \mathcal{L}(x^*, y^*).$$

Für alle $y \in D$ gilt wegen $f_i(x^*) \leq 0$ für $1 \leq i \leq p$ und $f_j(x^*) = 0$ für $p+1 \leq j \leq m$ die Ungleichung

$$y^t F(x^*) \leq 0.$$

Also ist

$$\begin{aligned} \mathcal{L}(x^*, y^*) &= f(x^*) \\ &\geq f(x^*) + y^t F(x^*) \\ &= \mathcal{L}(x^*, y). \end{aligned}$$

Dies zeigt, dass (x^*, y^*) ein Sattelpunkt von $\mathcal{L}$ ist.
„$\Longleftarrow$“: Für alle $y \in D$ gilt dann

$$\mathcal{L}(x^*, y^*) \geq \mathcal{L}(x^*, y) = f(x^*) + y^t F(x^*).$$

Aus der Definition von D folgt dann $f_i(x^*) \leq 0$ für $1 \leq i \leq p$ und $f_j(x^*) = 0$ für $p+1 \leq j \leq m$. Also ist $x^* \in S$. Angenommen, es ist $y_i^* f_i(x^*) \neq 0$ für ein $i \in \{1, \ldots, p\}$. Dann gilt $y_i^* \geq 0$ und $f_i(x^*) < 0$. Setze $y_i = 0$ und $y_\ell = y_\ell^*$ für $\ell \neq i$. Dann folgt

$$\mathcal{L}(x^*, y) > \mathcal{L}(x^*, y^*)$$

im Widerspruch zur Sattelpunktseigenschaft. Also ist $y_i^* f_i(x^*) = 0$ für alle $1 \leq i \leq m$ und somit

$$f(x^*) = \mathcal{L}(x^*, y^*).$$

Für beliebiges $x \in S$ folgt

$$f(x^*) = \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*) = f(x) + \underbrace{y^{*t}}_{\geq 0} \underbrace{F(x)}_{\leq 0} \leq f(x).$$

Also ist x^* eine Optimallösung von (III.3.1). □

Aus Satz III.3.6, Satz III.3.8 und dem Beweis von Satz III.3.8 folgt:

KOROLLAR III.3.9. *Die Voraussetzungen seien wie in Satz III.3.6. Dann ist $x^* \in \mathbb{R}^n$ genau dann eine Optimallösung von (III.3.1), wenn es ein $y^* \in D \subset \mathbb{R}^m$ gibt, so dass (x^*, y^*) die Bedingungen (III.3.3) erfüllt.*

III.4. Optimalitätskriterien für allgemeine Probleme

Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$. Ist x^* ein lokales Minimum von f, so ist bekanntlich x^* ein kritischer Punkt, d.h. $Df(x^*) = 0$, und die Hesse-Matrix $D^2f(x^*)$ ist positiv semi-definit. Ist umgekehrt x^* ein kritischer Punkt und $D^2f(x^*)$ positiv definit, so ist bekanntlich x^* ein lokales Minimum von f.

Wir wollen diese notwendigen und hinreichenden Kriterien so weit wie möglich auf Optimierungsprobleme mit Nebenbedingungen übertragen. Dabei wollen wir die Voraussetzungen des vorigen Abschnittes dahingehend abschwächen, dass die zu minimierende Funktion f und die Funktionen f_i, die die Ungleichungsnebenbedingungen beschreiben, nicht länger konvex und dass die Funktionen f_j, die die Gleichungsnebenbedingungen beschreiben, nicht mehr affin sein müssen. Um unnötige technische Schwierigkeiten zu vermeiden, betrachten wir allerdings nur den Fall, dass der gemeinsame Definitionsbereich der Funktionen f, f_i und f_j ganz $\mathbb{R}^n$ ist[3] und dass die Funktionen ein- oder zweimal stetig differenzierbar sind.

Wir betrachten somit folgende Aufgabenstellung:

> *Problem (NP):* Gegeben sind differenzierbare Funktionen $f : \mathbb{R}^n \to \mathbb{R}$, $f_1, \dots, f_p : \mathbb{R}^n \to \mathbb{R}$ und $f_{p+1}, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$. Gesucht ist ein Minimum von f in $\mathbb{R}^n$ unter den Nebenbedingungen
>
> $$f_i(x) \le 0 \quad 1 \le i \le p,$$
> $$f_j(x) = 0 \quad p+1 \le j \le m.$$

Wieder lassen wir die Sonderfälle $p = 0$ (keine Ungleichungsnebenbedingungen) und $m = p$ (keine Gleichungsnebenbedingungen) zu. Wir benutzen die gleichen Notationen wie im vorigen Paragraphen und setzen insbesondere

$$S = \{x \in \mathbb{R}^n : f_i(x) \le 0\ ,\ 1 \le i \le p\ ,\ f_j(x) = 0\ ,\ p+1 \le j \le m\}.$$

Damit kann das Problem *(NP)* in der kompakten Form

$$\min\{f(x) : x \in S\} \tag{III.4.1}$$

geschrieben werden.

Definition III.4.1. Seien $S \subset \mathbb{R}^n$ eine nicht leere Menge und $x \in S$. Dann heißt

$$T(S;x) = \Big\{v \in \mathbb{R}^n : \exists (\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+\ ,\ \exists (x_k)_{k\in\mathbb{N}} \subset S \text{ mit } \lim_{k\to\infty} x_k = x \text{ und } \lim_{k\to\infty} \lambda_k(x_k - x) = v\Big\}$$

der *Tangentialkegel* von S in x (vgl. Abbildungen III.4.1 und III.4.2).

[3] Eine offene Teilmenge reicht.

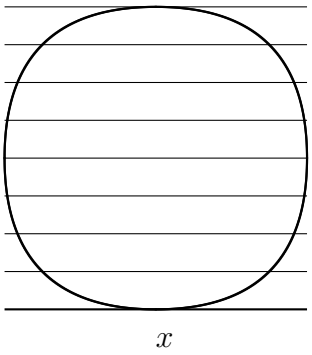

ABBILDUNG III.4.1. Menge S (fett umrandet) aus Beispiel III.4.3 (1) und „angehefteter“ Tangentialkegel $x + T(S;x)$ schraffiert

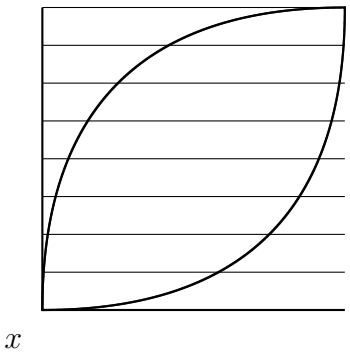

ABBILDUNG III.4.2. Menge S (fett umrandet) aus Beispiel III.4.3 (2) und „angehefteter“ Tangentialkegel $x + T(S;x)$ schraffiert

BEMERKUNG III.4.2. (1) Ist x ein innerer Punkt von S, so ist

$$T(S;x) = \mathbb{R}^n.$$

(2) Ist

$$S = \{x \in \mathbb{R}^n : F(x) = 0\}$$

mit $F \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ eine C^1-Mannigfaltigkeit, so ist $T(S;x)$ der Tangentialraum von S in x.
(3) Sind $S \subset \mathbb{R}^n$ eine nicht leere Menge, x ein Punkt in S und $K \subset \mathbb{R}^n$ ein abgeschlossener Kegel mit $S - x \subset K$, so ist $T(S;x) \subset K$.

BEISPIEL III.4.3. (1) Sei

$$S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$$

und

$$x = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Dann ist

$$T(S;x) = \{v \in \mathbb{R}^2 : v_2 \geq 0\}.$$

Denn für $v \in \mathbb{R}^2$ mit $v_2 \geq 0$ erfüllen v und

$$x_k = x + \frac{1}{k}v, \quad \lambda_k = k$$

für hinreichend großes $k \in \mathbb{N}$ die Bedingungen von Definition III.4.1. Dies beweist $\mathbb{R} \times \mathbb{R}_+ \subset T(S;x)$. Die umgekehrte Inklusion folgt aus Bemerkung III.4.2 (3).

(2) Sei

$$S = \{x \in \mathbb{R}^2 : x_2 \geq x_1^2\ ,\ x_1 \geq 0\} \cap \{x \in \mathbb{R}^2 : x_1 \geq x_2^2\ ,\ x_2 \geq 0\}.$$

Dann ist

$$T(S;0) = \{v \in \mathbb{R}^2 : v_1 \geq 0\ ,\ v_2 \geq 0\}.$$

Denn für $v \in (\mathbb{R}_+)^2$ erfüllen v und

$$x_k = \frac{1}{k}v + \frac{1}{k^2}(1 + \|v\|^2)\begin{pmatrix}1\\1\end{pmatrix}, \quad \lambda_k = k$$

für hinreichend großes $k \in \mathbb{N}$ die Bedingungen von Definition III.4.1. Dies beweist $\mathbb{R}_+ \times \mathbb{R}_+ \subset T(S;x)$. Die umgekehrte Inklusion folgt wieder aus Bemerkung III.4.2 (3).

(3) Sei

$$S = \{x \in \mathbb{R}^2 : x_1 \geq 0\ ,\ 0 \leq x_2 \leq x_1^2\}.$$

Dann ist

$$T(S;0) = \{v \in \mathbb{R}^2 : v_1 \geq 0\ ,\ v_2 = 0\}.$$

Denn ist $(x_k)_{k\in\mathbb{N}} \subset S$ und $(\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+$ mit $\lambda_k x_k \to v$, so ist $v_1 \geq 0$ und $v_2 \geq 0$. Angenommen es sei $v_2 > 0$. Wegen $x_{k,2} \to 0$ folgt dann $\lambda_k \to \infty$ und

$$\lambda_k x_{k,2} \leq \lambda_k x_{k,1}^2 = \frac{1}{\lambda_k}(\lambda_k x_{k,1})^2 \to 0.$$

Dies ist ein Widerspruch. Also ist $v_2 = 0$.
Ist umgekehrt $v \in \mathbb{R}^2$ mit $v_1 \geq 0$ und $v_2 = 0$ beliebig, ist $x_k = \frac{1}{k}v \in S$ und erfüllt $x_k \to 0$ und $kx_k \to v$. Also ist $v \in T(S;0)$.

SATZ III.4.4. *$T(S;x)$ ist ein nicht leerer, abgeschlossener Kegel.*

BEWEIS. Offensichtlich ist $0 \in T(S;x)$.
Die Kegeleigenschaft folgt sofort aus der Definition.
Zum Nachweis der Abgeschlossenheit sei $(v_k)_{k\in\mathbb{N}} \subset T(S;x)$ und $v \in \mathbb{R}^n$ mit

$$\lim_{k\to\infty} v_k = v.$$

Indem wir ggf. zu einer Teilfolge übergehen, können wir

$$\|v_k - v\| \leq \frac{1}{k}$$

für alle $k \in \mathbb{N}^*$ voraussetzen. Zu jedem v_k gibt es definitionsgemäß Folgen $(\lambda_{k,j})_{j\in\mathbb{N}} \subset \mathbb{R}_+$ und $(x_{k,j})_{j\in\mathbb{N}} \subset S$ mit

$$\lim_{j\to\infty} x_{k,j} = x$$

und

$$\lim_{j\to\infty} \lambda_{k,j}(x_{k,j} - x) = v_k.$$

Wähle für jedes k den Index $j(k)$ so, dass für alle $j \geq j(k)$ gilt

$$\|x_{k,j} - x\| \leq \frac{1}{k}$$

und

$$\|\lambda_{k,j}(x_{k,j} - x) - v_k\| \leq \frac{1}{k}.$$

Dann folgt

$$\begin{aligned}\|\lambda_{k,j(k)}(x_{k,j(k)} - x) - v\| &\leq \|\lambda_{k,j(k)}(x_{k,j(k)} - x) - v_k\| + \|v_k - v\| \\ &\leq \frac{2}{k}.\end{aligned}$$

Also gilt

$$\lim_{k\to\infty} x_{k,j(k)} = x$$

und

$$\lim_{k\to\infty} \lambda_{k,j(k)}(x_{k,j(k)} - x) = v.$$

Also ist $v \in T(S;x)$ und $T(S;x)$ somit abgeschlossen. □

Satz III.4.5. *Sei $x^* \in S$ ein lokales Minimum von f und f in x^* differenzierbar. Dann gilt*

$$Df(x^*)v \geq 0$$

für alle $v \in T(S;x^*)$.

Beweis. Nach Voraussetzung gibt es ein $\varepsilon > 0$ und eine in 0 stetige Funktion $r : \mathbb{R} \to \mathbb{R}$ mit

$$r(0) = 0$$

und

$$f(x) = f(x^*) + Df(x^*)(x - x^*) + r(\|x - x^*\|)\|x - x^*\|$$

für alle $x \in B(x^*, \varepsilon) \cap S$. Sei nun $v \in T(S;x^*)$ und $(\lambda_k)_{k\in\mathbb{N}}$ und $(x_k)_{k\in\mathbb{N}}$ wie in Definition III.4.1. Dann gibt es ein $k_0 \in \mathbb{N}$ mit $x_k \in B(x^*, \varepsilon) \cap S$ für alle $k \geq k_0$. Da x^* ein lokales Minimum ist, gilt für alle $k \geq k_0$

$$\begin{aligned}f(x^*) &\leq f(x_k) \\ &= f(x^*) + Df(x^*)(x_k - x^*) + r(\|x_k - x^*\|)\|x_k - x^*\|\end{aligned}$$

und somit

$$0 \leq Df(x^*)(x_k - x^*) + r(\|x_k - x^*\|)\|x_k - x^*\|.$$

Hieraus folgt nach Multiplikation mit λ_k

$$0 \le Df(x^*) \underbrace{[\lambda_k(x_k - x^*)]}_{\to v} + \underbrace{r(\|x_k - x^*\|)}_{\to 0} \underbrace{\|\lambda_k(x_k - x^*)\|}_{\to \|v\|}$$
$$\to Df(x^*)v.$$

Dies beweist die Behauptung. □

Bemerkung III.4.6. Ist x^* ein innerer Punkt von S, ist $T(S; x^*) = \mathbb{R}^n$ und somit $Df(x^*)v \ge 0$ für alle $v \in \mathbb{R}^n$. Hieraus folgt das bekannte Kriterium $Df(x^*) = 0$.

Satz III.4.5 ist für die praktische Rechnung wenig geeignet, da die Berechnung des Tangentialkegels in der Regel recht aufwändig ist. Wir wollen auf einfachere Bedingungen vom Typ der Karush-Kuhn-Tucker-Bedingungen (III.3.3) (S. 134) des vorigen Paragraphen hinaus. Dazu ersetzen wir die Funktion f und die Nebenbedingungen F durch Linearisierungen. Zur Abkürzung schreiben wir im Folgenden für Vektoren $u, v \in \mathbb{R}^m$

$$u \le_K v \quad \Longleftrightarrow \quad \begin{aligned} u_i &\le v_i \quad \text{für } 1 \le i \le p \text{ und} \\ u_j &= v_j \quad \text{für } p+1 \le j \le m. \end{aligned}$$

Sei nun $x^* \in \mathbb{R}^n$, $a \in \mathbb{R}^m$ mit $a \le_K 0$ und $A \in \mathbb{R}^{m \times n}$. Dann definieren wir

$$S_L = \{x \in \mathbb{R}^n : a + A(x - x^*) \le_K 0\}$$

und

$$L(S_L) = \{v \in \mathbb{R}^n : v = \lambda(x - x^*)\ ,\ \lambda \ge 0\ ,\ x \in S_L\}.$$

$L(S_L)$ kann man als eine kegelförmige „Vergrößerung“ von S_L auffassen.

Lemma III.4.7. *Es ist* $L(S_L) \subset T(S_L; x^*)$.

Beweis. Sei $v \in L(S_L)$. Dann gibt es ein $x \in S_L$ und ein $\lambda \ge 0$ mit

$$v = \lambda(x - x^*).$$

Setze für $k \ge 1$

$$x_k = \frac{1}{k}x + \frac{k-1}{k}x^*.$$

Dann gilt

$$\lim_{k\to\infty} x_k = x^*$$

und

$$k\lambda(x_k - x^*) = k\lambda\frac{1}{k}(x - x^*) = v$$

sowie

$$\begin{aligned} a + A(x_k - x^*) &= a + A[\frac{1}{k}(x - x^*)] \\ &= \frac{1}{k}\underbrace{[a + A(x - x^*)]}_{\leq_K 0} + \frac{k-1}{k}\underbrace{a}_{\leq_K 0} \\ &\leq_K 0. \end{aligned}$$

Also ist $x_k \in S_L$ und daher $v \in T(S_L; x^*)$. □

LEMMA III.4.8. *Es sei $x^* \in S$ ein lokales Minimum von f und $a = F(x^*)$, $A = DF(x^*)$ in der Definition von S_L. Weiter gelte $L(S_L) \subset T(S; x^*)$. Dann löst x^* das Minimumproblem*

$$\begin{aligned} &\min\{f(x^*) + Df(x^*)(x - x^*) : x \in S_L\} \\ &= \min\{f(x^*) + Df(x^*)(x - x^*) : F(x^*) + DF(x^*)(x - x^*) \leq_K 0\}. \end{aligned}$$

BEWEIS. Gemäß Satz III.4.5 ist

$$Df(x^*)v \geq 0$$

für alle $v \in T(S; x^*)$ also insbesondere

$$Df(x^*)w \geq 0$$

für alle $w \in L(S_L)$. Sei nun $x \in S_L$ beliebig. Dann ist $x - x^* \in S_L$ und wir erhalten

$$f(x^*) + \underbrace{Df(x^*)(x - x^*)}_{\geq 0} \geq f(x^*).$$

Also löst x^* obiges linearisiertes Optimierungsproblem. □

Das folgende Lemma gibt ein einfaches Kriterium, wann $L(S_L) \subset T(S; x^*)$ ist. Wir verzichten auf den sehr technischen Beweis und verweisen stattdessen auf [2, Satz 9.1.14 und Satz 9.1.19].

LEMMA III.4.9. *In der Definition von S_L sei $a = F(x^*)$ und $A = DF(x^*)$. Dann ist $L(S_L) \subset T(S; x^*)$ genau dann, wenn die Gradienten $Df_{p+1}(x^*)$, …, $Df_m(x^*)$ linear unabhängig sind und es ein $s \in \mathbb{R}^n$ gibt mit $Df_j(x^*)s = 0$ für alle $j \in \{p+1, \ldots, m\}$ und $Df_i(x^*)s < 0$ für alle Indizes $i \in \{1, \ldots, m\}$ mit $f_i(x^*) = 0$.*

Aus den Lemmata III.4.8 und III.4.9 und Satz III.3.6 (S. 134) folgt:

SATZ III.4.10 (KKT-Bedingungen). *Es sei $x^* \in S$ ein lokales Minimum von f. Die Gradienten $Df_{p+1}(x^*)$, …, $Df_m(x^*)$ seien linear unabhängig und es gebe ein $s \in \mathbb{R}^n$ gibt mit $Df_j(x^*)s = 0$ für alle $j \in \{p+1, \ldots, m\}$ und $Df_i(x^*)s < 0$ für alle Indizes $i \in \{1, \ldots, m\}$ mit $f_i(x^*) = 0$. Dann gibt es einen Vektor $y^* \in \mathbb{R}^m$, so dass die Karush-Kuhn-Tucker-Bedingungen (III.3.3) (S. 134) aus Satz III.3.6 (S. 134) erfüllt sind.*

KOROLLAR III.4.11. *Die Voraussetzungen von Satz III.4.10 seien erfüllt. Dann gibt es einen Vektor* $y^* \in (\mathbb{R}_+)^p \times \mathbb{R}^{m-p}$, *so dass* (x^*, y^*) *ein Sattelpunkt der Lagrange-Funktion*

$$\mathcal{L}(x, y) = f(x) + y^{*t} F(x)$$

ist.

Satz III.4.10 bzw. Korollar III.4.11 geben notwendige Bedingungen für eine Lösung von Problem (III.4.1). Der folgende Satz gibt hinreichende Bedingungen. Sein Beweis ist technisch aufwändig, da genau Buch geführt werden muss, welche Ungleichungsnebenbedingungen zu Gleichungsnebenbedingungen entarten, und weil man Ergebnisse zur Linearisierung von Tangentialkegeln benötigt, die diejenigen von Lemma III.4.9 ergänzen. Wir verzichten daher auf den Beweis und verweisen stattdessen auf [**2**, Satz 9.2.2 und Satz 9.2.8].

SATZ III.4.12. *Für* $(x, y) \in S \times ((\mathbb{R}_+)^p \times \mathbb{R}^{m-p})$ *definiere die Mengen* $I(x)$ *und* $\widetilde{I}(x, y)$ *durch*

$$I(x) = \{i : 1 \le i \le p\ ,\ f_i(x) = 0\}$$

und

$$\widetilde{I}(x, y) = \{i \in I(x) : y_i > 0\}.$$

Folgende Voraussetzungen seien erfüllt:

(1) $(x^*, y^*) \in S \times ((\mathbb{R}_+)^p \times \mathbb{R}^{m-p})$ *ist ein Sattelpunkt der Lagrange-Funktion* $\mathcal{L}$.

(2) *Die Gradienten* $Df_k(x^*)$, $k \in \widetilde{I}(x^*, y^*) \cup \{p+1, \dots, m\}$, *sind linear unabhängig.*

(3) *Es gibt ein* $s \in \mathbb{R}^n$ *mit*

$$Df_k(x^*)s = 0$$

für alle $k \in \widetilde{I}(x^*, y^*) \cup \{p+1, \dots, m\}$ *und*

$$Df_j(x^*)s < 0$$

für alle $j \in I(x^*) \backslash \widetilde{I}(x^*, y^*)$.

(4) *Mit*

$$L(x^*) = \Big\{ s \in \mathbb{R}^n : Df_k(x^*)s = 0 \quad \text{für } k \in \widetilde{I}(x^*, y^*) \cup \{p+1, \dots, m\},\ Df_j(x^*)s \le 0 \quad \text{für } j \in I(x^*) \backslash \widetilde{I}(x^*, y^*) \Big\}$$

gilt für alle $s \in L(x^*) \backslash \{0\}$

$$\begin{aligned} 0 &< s^t D_x^2 \mathcal{L}(x^*, y^*) s \\ &= s^t D^2 f(x^*) s + \sum_{i=1}^{m} y_i^* s^t D^2 f_i(x^*) s. \end{aligned}$$

Dann ist x^* ein lokales Minimum von f, d.h., es gibt ein $\varepsilon > 0$ mit

$$f(x^*) < f(x)$$

für alle $x \in B(x^*, \varepsilon) \cap S \backslash \{x^*\}$.

III.5. Projektionsverfahren

In diesem Abschnitt wollen wir das Abstiegsverfahren aus Algorithmus III.1.2 (S. 110) auf Probleme mit Nebenbedingungen übertragen. Dazu betrachten wir das Problem (III.3.1) (S. 130)

$$\min\{f(x) : x \in S\}$$

mit einer nicht leeren, abgeschlossenen, konvexen Menge S und einer C^1-Funktion f. Ein wichtiger Spezialfall ist der eines Polyeders, d.h.

$$S = S_P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Innerhalb dieses Spezialfalles ist der Sonderfall einer quadratischen Zielfunktion, d.h.

$$f(x) = f_q(x) = \frac{1}{2} x^t C x + c^t x + \gamma,$$

von besonderer Bedeutung.

Die Grundidee unseres Ansatzes besteht darin, in Algorithmus III.1.2 (S. 110) die Iterierten auf die Menge S zu projizieren. Dazu benötigen wir einige Eigenschaften der Projektion auf konvexe Mengen.

LEMMA III.5.1. *Sei $S \subset \mathbb{R}^n$ eine nicht leere, abgeschlossene, konvexe Menge. Dann gilt:*

(1) *Zu jedem $x \in \mathbb{R}^n$ gibt es genau ein $P_S(x) \in S$ mit*

$$\|x - P_S(x)\| \leq \|x - y\|$$

für alle $y \in S$.

(2) *Für alle $y \in S$ gilt*

$$(x - P_S(x))^t (y - P_S(x)) \leq 0.$$

(3) *Für alle $x, y \in \mathbb{R}^n$ gilt*

$$(P_S(y) - P_S(x))^t (y - x) \geq \|P_S(y) - P_S(x)\|^2.$$

(4) *Für alle $x, y \in \mathbb{R}^n$ gilt*

$$\|P_S(y) - P_S(x)\| \leq \|x - y\|.$$

BEWEIS. *ad (1):* Sei $x \in \mathbb{R}^n$ beliebig. Nach Voraussetzung gibt es ein $y_0 \in S$. Für jedes $y \in S$ mit

$$\|y - y_0\| \geq 2\|x - y_0\|$$

gilt

$$\begin{aligned} \|y - x\| &\geq \|y - y_0\| - \|x - y_0\| \\ &\geq \|x - y_0\|. \end{aligned}$$

Also ist

$$\inf\{\|y-x\| : y \in S\} = \inf\{\|y-x\| : y \in S \cap \overline{B(y_0, 2\|x-y_0\|)}\}.$$

Da $S \cap \overline{B(y_0, 2\|x-y_0\|)}$ kompakt ist, gibt es ein $y^* \in S$ mit minimalem Abstand.
Wir müssen also noch die Eindeutigkeit beweisen. Seien dazu $y_1, y_2 \in S$ zwei Elemente mit minimalem Abstand. Da S konvex ist, ist

$$\frac{1}{2}(y_1 + y_2) \in S.$$

Dann folgt

$$\|x-y_1\|^2 = \|x-y_2\|^2$$

und somit

$$\begin{aligned}
\frac{1}{2}\|x-y_1\|^2 + \frac{1}{2}\|x-y_2\|^2 &= \|x-y_2\|^2 \\
&\leq \|x - \frac{1}{2}(y_1+y_2)\|^2 \\
&= \|\frac{1}{2}(x-y_1) + \frac{1}{2}(x-y_2)\|^2 \\
&= \frac{1}{4}\|x-y_1\|^2 + \frac{1}{4}\|x-y_2\|^2 \\
&\quad + \frac{1}{2}(x-y_1)^t(x-y_2) \\
&= \frac{1}{2}\|x-y_1\|^2 + \frac{1}{2}\|x-y_2\|^2 \\
&\quad - \|\frac{1}{2}(x-y_1) - \frac{1}{2}(x-y_2)\|^2 \\
&= \frac{1}{2}\|x-y_1\|^2 + \frac{1}{2}\|x-y_2\|^2 \\
&\quad - \frac{1}{4}\|y_1-y_2\|^2.
\end{aligned}$$

Also ist

$$\|y_1-y_2\|^2 \leq 0$$

und damit

$$y_1 = y_2.$$

ad (2): Seien $x \in \mathbb{R}^n$ und $y \in S$. Dann ist

$$ty + (1-t)P_S(x) \in S$$

für alle $t \in [0,1]$. Daher ist die Funktion $\varphi : [0,1] \to \mathbb{R}$ mit

$$\varphi(t) = \|x - ty - (1-t)P_S(x)\|^2$$

wohl definiert und differenzierbar und hat in 0 ein Minimum. Damit folgt

$$\begin{aligned} 0 &\leq \varphi'(0) \\ &= 2(x - ty - (1-t)P_S(x))^t(P_S(x) - y)\Big|_{t=0} \\ &= 2(x - P_S(x))^t(P_S(x) - y). \end{aligned}$$

Dies beweist die Behauptung.
ad (3): Aus (2) folgt für $x, y \in \mathbb{R}^n$

$$(x - P_S(x))^t(P_S(y) - P_S(x)) \leq 0$$

und

$$(y - P_S(y))^t(P_S(x) - P_S(y)) \leq 0.$$

Addition dieser Ungleichungen liefert

$$\begin{aligned} 0 &\geq (x - P_S(x))^t(P_S(y) - P_S(x)) + (y - P_S(y))^t(P_S(x) - P_S(y)) \\ &= (x - P_S(x))^t(P_S(y) - P_S(x)) + (P_S(y) - y)^t(P_S(y) - P_S(x)) \\ &= (x - y + P_S(y) - P_S(x))^t(P_S(y) - P_S(x)). \end{aligned}$$

Hieraus folgt die Behauptung.
ad (4): Folgt aus (3) und der Cauchy-Schwarzschen Ungleichung. □

Diese Überlegungen führen auf Algorithmus III.5.1.

Algorithmus III.5.1 Projektionsverfahren

Gegeben: Punkt $x \in S$, Parameter $\beta, \mu \in (0,1)$, $\gamma > 0$
Gesucht: stationärer Punkt $x \in S$ von f

1: **while** $\min\{Df(x)v : v \in T(S;x), \|v\| = 1\} < 0$ **do**
2: $\quad\delta \leftarrow 1$, $z \leftarrow P_S(x - \delta\gamma Df(x))$
3: $\quad$**while** $f(z) > f(x) + \mu Df(x)(z - x)$ **do**
4: $\quad\quad\delta \leftarrow \delta\beta$
5: $\quad$**end while**
6: $\quad x \leftarrow z$
7: **end while**

Bemerkung III.5.2. (1) Gemäß Satz III.4.5 (S. 140) gilt für jedes lokale Minimum x^* von f

$$Df(x^*)v \geq 0$$

für alle $v \in T(S;x^*)$. In Zeile 1 von Algorithmus III.5.1 wird dieses notwendige Kriterium nachgeprüft.
(2) In Algorithmus III.1.2 (S. 110) für Probleme ohne Nebenbedingungen wird im k-ten Schritt f entlang der Halbgeraden $\{x_k - tDf(x_k) : t > 0\}$ minimiert. In Algorithmus III.5.1 wird diese Halbgerade auf S projiziert und durch die Kurve $\{P_S(x_k - tDf(x_k)) : t > 0\}$ ersetzt.

Falls Algorithmus III.5.1 nicht Zeile 1 abbricht, kann man zeigen, dass für hinreichend kleines $t > 0$ gilt

$$Df(x_k)[P_S(x_k - tDf(x_k))] < 0.$$

Die Schleife in Zeile 3 wird also nach endlich vielen Schritten beendet.
(3) Falls $S = S_P$ ein Polyeder ist, kann man einen Startwert x_0 mit dem Simplexalgorithmus aus Kapitel I bestimmen (vgl. Satz I.4.10 (S. 36) und Algorithmus I.4.3 (S. 37)).
(4) Da Algorithmus III.5.1 ein Gradientenverfahren ist, muss man wie im nicht restringierten Fall aus §III.1 mit einer langsamen Konvergenz rechnen.
(5) Die Berechnung der Projektion P_S ist auch bei Polyedern häufig sehr aufwändig.

Zum Nachweis der Konvergenz von Algorithmus III.5.1 benötigen wir das folgende technische Hilfsergebnis:

LEMMA III.5.3. *Seien $x, d \in \mathbb{R}^n$. Definiere die Funktion $\psi : \mathbb{R}_+ \to \mathbb{R}$ durch*

$$\psi(t) = \frac{1}{t}\|P_S(x + td) - x\|.$$

Die Funktion ψ ist monoton fallend.

BEWEIS. Sei $0 < \beta < \alpha$. Falls

$$P_S(x + \alpha d) = P_S(x + \beta d)$$

ist, ist nichts zu zeigen. Sei also

$$P_S(x + \alpha d) \neq P_S(x + \beta d).$$

Wir setzen zur Abkürzung

$$u = P_S(x + \alpha d) - x, \quad v = P_S(x + \beta d) - x.$$

Dann ist

$$u - v = P_S(x + \alpha d) - P_S(x + \beta d).$$

Aus Lemma III.5.1 (2) mit $y = P_S(x + \alpha d)$ folgt

$$(x + \beta d - P_S(x + \beta d))^t (P_S(x + \alpha d) - P_S(x + \beta d)) \leq 0 \tag{III.5.1}$$

bzw. nach Multiplikation mit -1

$$\begin{aligned} 0 &\leq (\underbrace{P_S(x + \beta d) - (x + \beta d)}_{=v - \beta d})^t (\underbrace{P_S(x + \alpha d) - P_S(x + \beta d)}_{=u - v}) \\ &= (v - \beta d)^t (u - v) \\ &= v^t(u - v) - \beta d^t(u - v) \end{aligned}$$

und somit

$$v^t(u - v) \geq \beta d^t(u - v). \tag{III.5.2}$$

Aus Lemma III.5.1 (3) folgt weiter

$$\begin{aligned}\|u-v\|^2 &= \|P_S(x+\alpha d) - P_S(x+\beta d)\|^2 \\ &\le (P_S(x+\alpha d) - P_S(x+\beta d))^t((x+\alpha d)-(x+\beta d)) \\ &= (u-v)^t d(\alpha-\beta).\end{aligned}$$

Da nach Voraussetzung $\alpha > \beta$ und $P_S(x+\alpha d) \neq P_S(x+\beta d)$ ist, folgt

$$(u-v)^t d > 0. \tag{III.5.3}$$

Aus der Cauchy-Schwarzschen Ungleichung

$$u^t v \le \|u\|\|v\|$$

folgt

$$u^t v\big[\|u\| + \|v\|\big] \le \|u\|\|v\|\big[\|u\| + \|v\|\big]$$

bzw. nach Umsortieren

$$\|u\|u^t v - \|u\|\|v\|^2 \le \|u\|^2\|v\| - u^t v\|v\|.$$

Wegen

$$\|u\|^2 = u^t u, \quad \|v\|^2 = v^t v$$

kann man diese Ungleichung in der Form schreiben

$$\begin{aligned}\|u\|v^t(u-v) &= \|u\|[\underbrace{v^t u}_{=u^t v} - v^t v] \\ &= \|u\|u^t v - \|u\|\|v\|^2 \\ &\le \|u\|^2\|v\| - u^t v\|v\| \\ &= \|v\|[u^t u - u^t v] \\ &= \|v\|u^t(u-v).\end{aligned}$$

Wegen (III.5.3) können wir diese Ungleichung durch $v^t(u-v)$ dividieren und die Ungleichung bleibt erhalten. Wegen (III.5.3) ist auch $v \neq 0$, so dass wir ebenso durch $\|v\|$ dividieren können, ohne die Ungleichung zu ändern. Insgesamt folgt

$$\frac{\|P_S(x+\alpha d) - x\|}{\|P_S(x+\beta d) - x\|} = \frac{\|u\|}{\|v\|} \le \frac{u^t(u-v)}{v^t(u-v)}.$$

Wegen (III.5.2) ist

$$\frac{u^t(u-v)}{v^t(u-v)} \le \frac{u^t(u-v)}{\beta d^t(u-v)}.$$

Vertauschen wir in (III.5.1) die Rollen von α und β erhalten wir

$$\begin{aligned}0 &\ge (x+\alpha d - P_S(x+\alpha d))^t(P_S(x+\beta d) - P_S(x+\alpha d)) \\ &= (P_S(x+\alpha d) - (x+\alpha d))^t(P_S(x+\alpha d) - P_S(x+\beta d)) \\ &= (u-\alpha d)^t(u-v) \\ &= u^t(u-v) - \alpha d^t(u-v)\end{aligned}$$

bzw.

$$u^t(u-v) \le \alpha d^t(u-v).$$

Insgesamt folgt

$$\frac{\|P_S(x+\alpha d)-x\|}{\|P_S(x+\beta d)-x\|} \le \frac{u^t(u-v)}{v^t(u-v)} \le \frac{\alpha d^t(u-v)}{\beta d^t(u-v)} = \frac{\alpha}{\beta}.$$

Also ist

$$\psi(\alpha) \le \psi(\beta),$$

was die Monotonie von ψ beweist. □

SATZ III.5.4. *Folgende Voraussetzungen seien erfüllt:*

(a) *Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist auf S stetig differenzierbar.*
(b) *Die Ableitung Df ist auf S gleichmäßig stetig.*
(c) *f ist auf S nach unten beschränkt, d.h.*

$$\inf\{f(x) : x \in S\} > -\infty.$$

Dann gilt für jede von Algorithmus III.5.1 erzeugte Folge $(x_k)_{k\in\mathbb{N}}$:

(1) $\lim\limits_{k\to\infty} \dfrac{1}{\beta^{m_k}\gamma}\|x_{k+1}-x_k\| = 0.$
(2) $\lim\limits_{k\to\infty} \|x_{k+1}-x_k\| = 0.$
(3) *Für jeden Häufungspunkt x^* von $(x_k)_{k\in\mathbb{N}}$ ist*

$$Df(x^*)v \ge 0$$

für alle $v \in T(S;x^*)$.

BEWEIS. *ad (1):* Sei m_k die Zahl der Durchläufe der Schleife in Zeile 3 von Algorithmus III.5.1 im k-ten Durchlauf der Schleife in Zeile 1. Wir setzen zur Abkürzung

$$\alpha_k = \beta^{m_k}\gamma.$$

Dann lautet die Behauptung

$$\lim_{k\to\infty} \frac{1}{\alpha_k}\|x_{k+1}-x_k\| = 0.$$

Wir nehmen an, dies sei nicht der Fall. Dann gibt es ein $\varepsilon > 0$ und eine unendliche Indexmenge K mit

$$\frac{1}{\alpha_k}\|x_{k+1}-x_k\| \ge \varepsilon$$

für alle $k \in K$. Für alle $k \in K$ gilt dann auch

$$\begin{aligned} \frac{1}{\alpha_k}\|x_{k+1}-x_k\|^2 &\ge \max\{\varepsilon\|x_{k+1}-x_k\|^2\ ,\ \varepsilon^2\alpha_k\} \\ &= \varepsilon\max\{\|x_{k+1}-x_k\|^2\ ,\ \varepsilon\alpha_k\}. \end{aligned} \tag{III.5.4}$$

Da $(f(x_k))_{k\in\mathbb{N}}$ monoton fällt und gemäß Voraussetzung (c) nach unten beschränkt ist, ist $(f(x_k))_{k\in\mathbb{N}}$ konvergent. Damit folgt aus Schritt (3) von Algorithmus III.5.1

$$\lim_{k\to\infty} Df(x_k)(x_{k+1}-x_k)=0. \tag{III.5.5}$$

Wegen $x_k\in S$ folgt aus Lemma III.5.1 (3)

$$\begin{aligned}
&\|x_{k+1}-x_k\|^2\\
&=\|P_S(x_k-\alpha_k Df(x_k))-x_k\|^2\\
&=\|P_S(x_k-\alpha_k Df(x_k))-P_S(x_k)\|^2\\
&\le (P_S(x_k-\alpha_k Df(x_k))-P_S(x_k))^t(x_k-\alpha_k Df(x_k)-x_k)\\
&=(x_{k+1}-x_k)^t(\alpha_k Df(x_k))\\
&=\alpha_k Df(x_k)(x_k-x_{k+1}).
\end{aligned}$$

Also gilt auch

$$\lim_{k\to\infty}\frac{1}{\alpha_k}\|x_{k+1}-x_k\|^2=0.$$

Damit folgt aus (III.5.4)

$$\lim_{\substack{k\to\infty\\ k\in K}}\alpha_k=0 \quad\text{und}\quad \lim_{\substack{k\to\infty\\ k\in K}}\|x_{k+1}-x_k\|=0.$$

Mithin gilt für große $k\in K$

$$m_k>0.$$

O.E. können wir daher $m_k>0$ für alle $k\in K$ annehmen. Wegen der Schleife in Zeile 3 von Algorithmus III.5.1 gilt mit der Abkürzung

$$\overline{x}_{k+1}=P_S\left(x_k-\frac{\alpha_k}{\beta}Df(x_k)\right)$$

für alle $k\in K$

$$f(\overline{x}_{k+1})>f(x_k)+\mu Df(x_k)(\overline{x}_{k+1}-x_k).$$

Wegen

$$\frac{\alpha_k}{\beta}>\alpha_k$$

folgt aus Lemma III.5.3 mit $x=x_k$, $d=-Df(x_k)$ die Abschätzung

$$\begin{aligned}
\frac{1}{\alpha_k}\|x_{k+1}-x_k\|^2&=\|x_{k+1}-x_k\|\underbrace{\frac{\|x_{k+1}-x_k\|}{\alpha_k}}_{=\psi(\alpha_k)}\\
&\ge\underbrace{\|x_{k+1}-x_k\|}_{\ge\varepsilon\alpha_k}\underbrace{\frac{\|\overline{x}_{k+1}-x_k\|}{\frac{\alpha_k}{\beta}}}_{=\psi(\frac{\alpha_k}{\beta})}\\
&\ge\varepsilon\beta\|\overline{x}_{k+1}-x_k\|.
\end{aligned}$$

Lemma III.5.1 (3) mit

$$x = x_k - \alpha_k Df(x_k), \quad y = x_k - \frac{\alpha_k}{\beta} Df(x_k)$$

liefert

$$\begin{aligned} 0 &\le \|\overline{x}_{k+1} - x_{k+1}\|^2 \\ &\le \underbrace{(-\frac{\alpha_k}{\beta} + \alpha_k)}_{=-\alpha_k \frac{1-\beta}{\beta}} Df(x_k)(\overline{x}_{k+1} - x_{k+1}) \end{aligned}$$

und somit

$$\begin{aligned} Df(x_k)(x_k - \overline{x}_{k+1}) &\ge Df(x_k)(x_k - x_{k+1}) \\ &\ge \frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 \\ &\ge \varepsilon\beta\|\overline{x}_{k+1} - x_k\|^2 \\ &> 0. \end{aligned} \tag{III.5.6}$$

Wegen

$$\lim_{k\to\infty} Df(x_k)(x_k - x_{k+1}) = 0$$

folgt hieraus

$$\lim_{\substack{k\to\infty \\ k\in K}} \|\overline{x}_{k+1} - x_k\| = 0.$$

Außerdem ist die Größe

$$\rho_k = \frac{f(x_k) - f(\overline{x}_{k+1})}{Df(x_k)(\overline{x}_{k+1} - x_k)}$$

wohl definiert und erfüllt für alle $k \in K$ die Ungleichung

$$\rho_k < \mu < 1.$$

Wegen der gleichmäßigen Stetigkeit von Df gibt es andererseits eine in 0 stetige Funktion r mit

$$r(0) = 0$$

und

$$|\rho_k - 1| \le \frac{r(\|\overline{x}_{k+1} - x_k\|)\|\overline{x}_{k+1} - x_k\|}{Df(x_k)(\overline{x}_{k+1} - x_k)}.$$

Wegen (III.5.6) folgt

$$|\rho_k - 1| \le \frac{1}{\varepsilon\beta} r(\|\overline{x}_{k+1} - x_k\|) \underset{\substack{k\to\infty \\ k\in K}}{\longrightarrow} 0.$$

Dies ist ein Widerspruch. Also gilt (1).
ad (2): Folgt wegen $m_k \ge 0$ und $\beta < 1$ aus (1).
ad (3): Sei nun x^* ein Häufungspunkt von $(x_k)_{k\in\mathbb{N}}$ und $(x_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge mit

$$x_{n_k} \underset{k\to\infty}{\longrightarrow} x^*.$$

Sei $z \in S$ beliebig. Dann folgt aus Lemma III.5.1 (2) mit $x = x_k - \alpha_k Df(x_k)$ und $y = z$

$$(x_{k+1} - (x_k - \alpha_k Df(x_k)))^t (x_{k+1} - z) \le 0.$$

Durch Umformen erhalten wir

$$\begin{aligned}\alpha_k Df(x_k)(x_{k+1} - z) &\le (x_{k+1} - x_k)^t (z - x_{k+1}) \\ &= (x_{k+1} - x_k)^t (z - x_k) \\ &\quad + \underbrace{(x_{k+1} - x_k)^t (x_k - x_{k+1})}_{\le 0} \\ &\le (x_{k+1} - x_k)^t (z - x_k) \\ &\le \|x_{k+1} - x_k\| \|z - x_k\|\end{aligned}$$

und

$$\begin{aligned}Df(x_k)(x_k - z) &= Df(x_k)(x_k - x_{k+1}) + Df(x_k)(x_{k+1} - z) \\ &\le Df(x_k)(x_k - x_{k+1}) + \frac{1}{\alpha_k} \|x_{k+1} - x_k\| \|z - x_k\|.\end{aligned}$$

Wegen (III.5.5) folgt

$$Df(x^*)(x^* - z) \le 0$$

für alle $z \in S$. Zusammen mit der Definition III.4.1 (S. 137) des Tangentialkegels folgt hieraus die Behauptung. □

III.6. Penalty-Verfahren

Wir betrachten zunächst das allgemeine Problem

$$\min\{f(x) : x \in S\} \tag{III.6.1}$$

mit einer nicht leeren, abgeschlossenen Teilmenge S von $\mathbb{R}^n$ und einer stetigen Funktion f. Später werden wir uns auf das Problem (III.4.1) (S. 137) aus §III.4 beschränken, d.h.

$$\begin{aligned}S = \Big\{ x \in \mathbb{R}^n : f_i(x) \le 0, 1 \le i \le p, \\ f_j(x) = 0, p+1 \le j \le m \Big\}.\end{aligned} \tag{III.6.2}$$

Definition III.6.1. Sei S eine abgeschlossene, nicht leere Teilmenge von $\mathbb{R}^n$. Eine Funktion $\ell : \mathbb{R}^n \to \mathbb{R}$ mit

$$\begin{aligned}\ell(x) &> 0 \quad \text{falls } x \notin S, \\ \ell(x) &= 0 \quad \text{falls } x \in S\end{aligned}$$

heißt *Straffunktion* (engl. *penalty function*) zu S.

Beispiel III.6.2. Für die Menge S aus (III.6.2) ist

$$\ell(x) = \sum_{i=1}^{p} (f_i(x)^+)^\alpha + \sum_{j=p+1}^{m} |f_j(x)|^\alpha$$

mit $\alpha > 0$ und

$$z^+ = \max\{z, 0\}$$

eine Straffunktion.

Die Idee der Penalty-Verfahren, die Algorithmus III.6.1 zugrunde liegt, ist folgende:

> Für $r > 0$ suchen wir ein lokales Minimum $x(r)$ der Funktion
>
> $$p(x, r) = f(x) + r\ell(x)$$
>
> in ganz $\mathbb{R}^n$, wobei ℓ eine Straffunktion zu S ist. Dies ist ein Optimierungsproblem ohne Nebenbedingungen. Falls $\ell(x(r)) = 0$ ist, ist $x(r)$ auch ein lokales Minimum von f in S und wir sind fertig. Nun betrachten wir eine monoton wachsende Folge $(r_k)_{k\in\mathbb{N}}$ mit $r_k \to \infty$ und berechnen die zugehörigen $x(r_k)$. Falls dieser Prozess nicht mit einem $x(r_p) \in S$ abbricht, hoffen wir, dass zumindest $(x(r_k))_{k\in\mathbb{N}}$ gegen ein Optimum von f in S konvergiert.

Algorithmus III.6.1 Allgemeines Penalty-Verfahren

Gegeben: Startwert $r > 0$, Funktion $f : \mathbb{R}^n \to \mathbb{R}$, Menge $S \subset \mathbb{R}^n$, Straffunktion ℓ zu S
Gesucht: Minimum $x \in S$ von f

1: **repeat**
2: $\quad x \leftarrow \operatorname{argmin}\{f(x) + r\ell(x) : x \in \mathbb{R}^n\}$
3: $\quad r \leftarrow 2r$
4: **until** $x \in S$

Der Satz III.6.3 beweist die Konvergenz von Algorithmus III.6.1.

Satz III.6.3. *Sei $f : \mathbb{R}^n \to \mathbb{R}$ stetig, $x^* \in S$ ein striktes lokales Minimum von f in S und $\ell : \mathbb{R}^n \to \mathbb{R}$ eine stetige Straffunktion zu S. Dann gibt es ein $r_0 > 0$ so, dass die Funktion*

$$p(x, r) = f(x) + r\ell(x)$$

für jedes $r \geq r_0$ ein lokales Minimum $x(r)$ besitzt, das für $r \to \infty$ gegen x^* konvergiert.

Beweis. Für $\varepsilon > 0$ sei

$$C_\varepsilon = \{x \in \mathbb{R}^n : \|x - x^*\| = \varepsilon\} = \partial B(x^*, \varepsilon).$$

Falls $S \cap C_\varepsilon \neq \emptyset$ ist, ist $S \cap C_\varepsilon$ kompakt. Daher können wir die Zahl $\delta(\varepsilon)$ durch

$$\delta(\varepsilon) = \begin{cases} \min\limits_{x\in S\cap C_\varepsilon} f(x) - f(x^*) & \text{falls } S \cap C_\varepsilon \neq \emptyset \\ 1 & \text{falls } S \cap C_\varepsilon = \emptyset \end{cases}$$

definieren. Da x^* ein striktes lokales Minimum ist, ist für hinreichend kleines ε im Fall $S \cap C_\varepsilon \neq \emptyset$

$$\delta(\varepsilon) > 0.$$

Es gibt daher ein $k_0 \geq 0$ mit

$$\delta(2^{-k}) > 0$$

für alle $k \geq k_0$. O.E. ist $k_0 = 0$.
Für $\rho > 0$ und $k \in \mathbb{N}$ setzen wir

$$S_{\rho,k} = \{x \in C_{2^{-k}} : \exists z \in S \cap C_{2^{-k}} \text{ mit } \|z - x\| \leq \rho\}.$$

Die Menge $S_{\rho,k}$ kann leer sein, enthält aber auf jeden Fall die Menge $S \cap C_{2^{-k}}$. Mit der Konvention

$$\min_{x \in \emptyset} f(x) = \infty$$

folgt daher wegen der gleichmäßigen Stetigkeit von f auf der kompakten Menge $C_{2^{-k}}$, dass es zu jedem $k \in \mathbb{N}$ ein $\rho_k > 0$ gibt mit

$$\min_{x \in S_{\rho_k,k}} f(x) - f(x^*) \geq \frac{\delta(2^{-k})}{2}.$$

Auf $\overline{C_{2^{-k}} \backslash S_{\rho_k,k}}$ gilt $\ell(x) > 0$. Wegen der Kompaktheit dieser Menge und der Stetigkeit von ℓ gibt es daher ein $\lambda_k > 0$ mit

$$\ell(x) \geq \lambda_k$$

für alle $x \in \overline{C_{2^{-k}} \backslash S_{\rho_k,k}}$.
Setze

$$M_k = \min \left\{ 0, \min_{x \in C_{2^{-k}}} f(x) - f(x^*) \right\}$$

und

$$\widehat{r}_k = -\frac{M_k}{\lambda_k}.$$

Für $x \in C_{2^{-k}} \cap S_{\rho_k,k}$ und $r \geq \widehat{r}_k$ gilt dann

$$p(x,r) = f(x) \geq f(x^*) + \frac{\delta(2^{-k})}{2} > f(x^*) = p(x^*,r).$$

Für $x \in C_{2^{-k}} \backslash S_{\rho_k,k}$ und $r \geq \widehat{r}_k$ gilt andererseits

$$p(x,r) \geq f(x) - M \geq f(x^*) = p(x^*,r).$$

Daher besitzt $p(\cdot, r)$ für jedes $r \geq \widehat{r}_k$ ein lokales Minimum in $\overline{B(x^*, 2^{-k})}$.
Wir setzen nun

$$r_0 = \widehat{r}_0 + 1$$

und für $k \geq 1$

$$r_k = \max\{r_{k-1}, \widehat{r}_k\} + 1.$$

Dann ist die Folge $(r_k)_{k\in\mathbb{N}}$ monoton wachsend mit $r_k \to \infty$. Außerdem besitzt die Funktion $p(\cdot, r)$ für jedes $r \in [r_k, r_{k+1})$ ein lokales Minimum $x(r)$ mit

$$\|x(r) - x^*\| \le 2^{-k}.$$

Dies beweist die Behauptung. □

In jedem Schritt von Algorithmus III.6.1 ist ein Minimierungsproblem ohne Nebenbedingungen zu lösen. Wir sind daher besonders an differenzierbaren Straffunktionen interessiert. In Hinblick auf Satz III.6.3 sind wir andererseits an sogenannten *exakten* Straffunktionen interessiert, bei denen es ein $\overline{r}$ gibt, so dass für alle $r \ge \overline{r}$ das Minimum von $p(\cdot, r)$ auch ein Minimum von f ist. Denn für derartige Straffunktionen bricht Algorithmus III.6.1 nach endlich vielen Schritten ab. Leider widersprechen sich diese beiden Wünsche in der Regel.

Wir suchen daher nach einer Modifikation der Penalty-Methode, die diesen Widerspruch nicht aufweist. Dazu beschränken wir uns auf Mengen S der Form (III.6.2), die wir in §III.4 untersucht haben. Wir setzen voraus, dass die zugehörige Lagrange-Funktion

$$\mathcal{L}(x, y) = f(x) + \sum_{i=1}^{m} y_i f_i(x)$$

einen Sattelpunkt (x^*, y^*) besitzt, der die Bedingungen von Satz III.4.12 (S. 143) erfüllt.

Definition III.6.4. Die Funktion $\Lambda : \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}^*_+)^m \to \mathbb{R}$, die durch

$$\begin{aligned}\Lambda(x, y, r) = f(x) &+ \sum_{i=1}^{p} \frac{1}{2} r_i \left[\left(f_i(x) + \frac{y_i}{r_i} \right)^+ \right]^2 \\ &+ \sum_{j=p+1}^{m} \frac{1}{2} r_j \left[f_j(x) + \frac{y_j}{r_j} \right]^2 \\ &- \sum_{k=1}^{m} \frac{1}{2} \frac{y_k^2}{r_k}\end{aligned}$$

definiert ist, heißt *erweiterte Lagrange-Funktion* (engl. *augmented Lagrangian*) zu Problem (III.6.1), (III.6.2). Dabei ist

$$z^+ = \max\{z, 0\}.$$

Bemerkung III.6.5. Wegen

$$\frac{1}{2} r \left[f + \frac{y}{r} \right]^2 - \frac{1}{2} \frac{y^2}{r} = \frac{1}{2} r f^2 + y f$$

gilt im Fall $p = 0$, d.h. beim Fehlen von Ungleichungsnebenbedingungen

$$\Lambda(x, y, r) = \mathcal{L}(x, y) + \sum_{j=1}^{m} \frac{1}{2} r_j f_j(x)^2.$$

Dies erklärt den Namen „erweiterte Lagrange-Funktion“.

SATZ III.6.6. *Folgende Aussagen sind äquivalent:*

(1) (x^*, y^*) *ist ein Sattelpunkt der Lagrange-Funktion* $\mathcal{L}$*, d.h., es gilt*

$$\begin{aligned} D_x\mathcal{L}(x^*, y^*) &= 0, \\ f_i(x^*) &\le 0, \quad 1 \le i \le p, \\ y_i^* &\ge 0, \quad 1 \le i \le p, \\ y_i^* f_i(x^*) &= 0, \quad 1 \le i \le p, \\ f_j(x^*) &= 0, \quad p+1 \le j \le m. \end{aligned}$$

(2) *Es gibt ein* $r \in (\mathbb{R}_+^*)^m$*, so dass* (x^*, y^*) *ein stationärer Punkt von* $\Lambda(\cdot, \cdot, r)$ *ist, d.h.*

$$\begin{aligned} D_x\Lambda(x^*, y^*, r) &= 0, \\ D_y\Lambda(x^*, y^*, r) &= 0. \end{aligned}$$

BEWEIS. Aus der Definition von Λ folgt

$$\begin{aligned} D_x\Lambda(x, y, r) &= Df(x) + \sum_{i=1}^{p} r_i \left[f_i(x) + \frac{y_i}{r_i} \right]^+ Df_i(x) \\ &\qquad + \sum_{j=p+1}^{m} r_j \left[f_j(x) + \frac{y_j}{r_j} \right] Df_j(x), \\ \frac{\partial}{\partial y_i}\Lambda(x, y, r) &= \left(f_i(x) + \frac{y_i}{r_i} \right)^+ - \frac{y_i}{r_i} \qquad , 1 \le i \le p, \\ \frac{\partial}{\partial y_j}\Lambda(x, y, r) &= f_j(x) \qquad , p+1 \le j \le m. \end{aligned}$$

„(1) $\Longrightarrow$ (2)“: Für $p+1 \le j \le m$ ist für jedes $r \in (\mathbb{R}_+^*)^m$

$$\frac{\partial}{\partial y_j}\Lambda(x^*, y^*, r) = f_j(x^*) = 0.$$

Für $1 \le i \le p$ und $r \in (\mathbb{R}_+^*)^m$ gilt

$$\begin{aligned} \frac{\partial}{\partial y_i}\Lambda(x^*, y^*, r) &= \left(f_i(x^*) + \frac{y_i^*}{r_i} \right)^+ - \frac{y_i^*}{r_i} \\ &= \max\left\{ f_i(x^*) + \frac{y_i^*}{r_i}, 0 \right\} - \frac{y_i^*}{r_i} \\ &= \max\left\{ f_i(x^*), -\frac{y_i^*}{r_i} \right\} \\ &= 0. \end{aligned}$$

Sei

$$I = \big\{ i \in \{1, \dots, p\} : f_i(x^*) = 0 \big\}.$$

Für $i \in \{1, \dots, p\} \backslash I$ gilt

$$f_i(x^*) < 0 \quad \text{und} \quad y_i^* = 0$$

und daher

$$\left(f_i(x^*) + \frac{y_i^*}{r_i} \right)^+ = 0.$$

Damit folgt

$$\begin{aligned}
D_x \Lambda(x^*, y^*, r) = Df(x^*) &+ \sum_{\substack{i=1 \\ i \in I}}^{p} r_i \underbrace{\left[f_i(x^*) + \frac{y_i^*}{r_i} \right]^+}_{= \frac{y_i^*}{r_i}} Df_i(x^*) \\
&+ \sum_{\substack{i=1 \\ i \notin I}}^{p} r_i \underbrace{\left[f_i(x^*) + \frac{y_i^*}{r_i} \right]^+}_{=0} Df_i(x^*) \\
&+ \sum_{j=p+1}^{m} r_j \underbrace{\left[f_j(x^*) + \frac{y_j^*}{r_j} \right]}_{= \frac{y_j^*}{r_j}} Df_j(x^*) \\
= Df(x^*) &+ \sum_{\substack{i=1 \\ i \in I}}^{p} y_i^* Df_i(x^*) + \sum_{\substack{i=1 \\ i \notin I}}^{p} r_i \underbrace{y_i^*}_{=0} Df_i(x^*) \\
&+ \sum_{j=p+1}^{m} y_j^* Df_j(x^*) \\
= D_x \mathcal{L}(x^*, y^*) & \\
= 0. &
\end{aligned}$$

„(2) $\Longrightarrow$ (1)“: Für $p+1 \leq j \leq m$ gilt

$$0 = \frac{\partial}{\partial y_j} \Lambda(x^*, y^*, r) = f_j(x^*).$$

Dies ist die letzte Bedingung von (1).
Für $1 \leq i \leq p$ haben wir

$$\begin{aligned}
0 &= \frac{\partial}{\partial y_i} \Lambda(x^*, y^*, r) \\
&= \left(f_i(x^*) + \frac{y_i^*}{r_i} \right)^+ - \frac{y_i^*}{r_i} \\
&= \max \left\{ f_i(x^*) + \frac{y_i^*}{r_i}, 0 \right\} - \frac{y_i^*}{r_i} \\
&= \max \left\{ f_i(x^*), -\frac{y_i^*}{r_i} \right\}.
\end{aligned}$$

Also gilt

$$f_i(x^*) \leq 0 \quad \text{und} \quad -\frac{y_i^*}{r_i} \leq 0$$

und somit

$$y_i^* \geq 0.$$

Außerdem können nicht $f_i(x^*)$ und y_i^* gleichzeitig von Null verschieden sein. Dies beweist die Bedingungen zwei bis vier von (1).
Damit folgt wie im ersten Teil des Beweises

$$0 = D_x\Lambda(x^*, y^*, r) = D_x\mathcal{L}(x^*, y^*)$$

also die erste Bedingung von (1). □

Bemerkung III.6.7. Beim Beweis der Richtung „(1) $\Longrightarrow$ (2)“ haben wir im obigen Beweis so gar gezeigt, dass

$$D_x\Lambda(x^*, y^*, r) = 0 \quad \text{und} \quad D_y\Lambda(x^*, y^*, r) = 0$$

ist für *alle* $r \in (\mathbb{R}_+^*)^m$.

Satz III.6.6 legt folgende Vorgehensweise nahe:

> Zu gegebenem y, r finde ein $x = x(y, r)$, das $\Lambda(\cdot, y, r)$ minimiert. Dann ist
>
> $$D_x\Lambda(x, y, r) = 0.$$
>
> Versuche nun, bei festem x, r ein $\widetilde{y}$ zu finden, das die zweite Gleichung
>
> $$D_y\Lambda(x, \widetilde{y}, r) = 0$$
>
> erfüllt.

Leider führt dieses naive Vorgehen in der Regel nicht zum Ziel. Stattdessen hilft folgende Beobachtung: Nach Berechnung von $x = x(y, r)$ gilt

$$\begin{aligned} 0 &= D_x\Lambda(x, y, r) \\ &= Df(x) + \sum_{i=1}^{p} r_i \left[f_i(x) + \frac{y_i}{r_i} \right]^+ Df_i(x) \\ &\qquad + \sum_{j=p+1}^{m} r_j \left[f_j(x) + \frac{y_j}{r_j} \right] Df_j(x). \end{aligned}$$

Andererseits lauten die KKT-Bedingungen (III.3.3) (S. 134) für das ursprüngliche Problem

$$0 = D_x\mathcal{L}(x, \widetilde{y}) = Df(x) + \sum_{k=1}^{m} \widetilde{y}_k Df_k(x).$$

Dies legt es nahe, y durch

$$y_i^{\text{neu}} = r_i \left(f_i(x) + \frac{y_i}{r_i} \right)^+ \quad \text{für } 1 \le i \le p,$$
$$y_j^{\text{neu}} = r_j \left(f_j(x) + \frac{y_j}{r_j} \right) \quad \text{für } p+1 \le j \le m$$

zu ersetzen. Da dann

$$D_x(\Lambda(x, y^{\text{neu}}, r) \neq 0$$

sein wird, ist dieser Prozess zu iterieren.

Diese Idee führt auf Algorithmus III.6.2.

Algorithmus III.6.2 Penalty-Verfahren mit erweiterter Lagrange-Funktion

Gegeben: Vektoren $r \in (\mathbb{R}_+^*)^m$, $y \in (\mathbb{R}_+)^p \times \mathbb{R}^{m-p}$.
Gesucht: stationärer Punkt der Lagrange-Funktion Λ

1: **repeat**
2: $x \leftarrow \operatorname{argmin}\{\Lambda(x, y, r) : x \in \mathbb{R}^n\}$
3: **for** $i = 1, \ldots, p$ **do**
4: $y_i \leftarrow (r_i f_i(x) + y_i)^+$
5: **end for**
6: **for** $j = p+1, \ldots, m$ **do**
7: $y_j \leftarrow r_j f_j(x) + y_j$
8: **end for**
9: **until** (x, y) erfüllt KKT-Bedingungen

Man kann zeigen [2, Korollar 11.2.23]:

Satz III.6.8. *Für hinreichend großes $\rho > 0$ konvergiert Algorithmus III.6.2 mit*

$$r = \rho e = \begin{pmatrix} \rho \\ \vdots \\ \rho \end{pmatrix}$$

gegen eine Sattelpunkt der Lagrange-Funktion $\mathcal{L}$. Die Konvergenz ist linear. Die Konvergenzgeschwindigkeit ist umso besser, je größer ρ ist.

Ein anderer Penalty-Ansatz geht von der Beobachtung aus, dass im allgemeinen Ungleichungsnebenbedingungen wesentlich schwerer zu behandeln sind als Gleichungsnebenbedingungen. Daher werden nur Ungleichungen mit Straftermen belegt. Andererseits werden in den Straftermen Ungleichungen der Form $f_i(x) \le 0$ zu $f_i(x) < d_i$ mit $d_i > 0$ abgeschwächt. Wesentliches Hilfsmittel bei diesem Ansatz sind *Barriere-Funktionen*.

Definition III.6.9. Eine Funktion $\mathcal{B} : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ heißt *Barriere-Funktion* (engl. *barrier function*), wenn sie folgende Bedingungen erfüllt:

(1) $\mathcal{B}(t) = \infty$ für alle $t \leq 0$.
(2) $\mathcal{B}$ ist monoton fallend.
(3) $\mathcal{B}$ ist konvex.
(4) $\mathcal{B}$ ist auf $\mathbb{R}_+^*$ stetig differenzierbar.
(5) $\lim\limits_{t\to 0+} \mathcal{B}(t) = \infty$.
(6) $\lim\limits_{t\to 0+} \mathcal{B}'(t) = -\infty$.

Beispiel III.6.10. Die Funktionen

$$\mathcal{B}(t) = \begin{cases} -\ln t & \text{für } t > 0 \\ \infty & \text{für } t \leq 0 \end{cases}$$

$$\mathcal{B}(t) = \begin{cases} t^{-\alpha} & \text{für } t > 0 \\ \infty & \text{für } t \leq 0 \end{cases}$$

mit $\alpha > 0$ sind Barriere-Funktionen.

Für eine gegebene Barriere-Funktion $\mathcal{B}$ definieren wir die Funktion $\Phi : \mathbb{R}^n \times \mathbb{R}_+^* \times (\mathbb{R}_+)^p \to \mathbb{R}$ durch

$$\Phi(x, \mu, d) = f(x) + \mu \sum_{i=1}^{p} \mathcal{B}(d_i - f_i(x))$$

und betrachten das Hilfsproblem

$$\min_x \{\Phi(x, \mu, d) : f_j(x) = 0\ ,\ p+1 \leq j \leq m\}\,. \tag{III.6.3}$$

Lemma III.6.11. *Sind die Funktionen f und $f_1, \ldots, f_p$ konvex, so ist die Funktion $\Phi(\cdot, \mu, d)$ für jedes $\mu > 0$ und $d \in \times(\mathbb{R}_+)^p$ ebenfalls konvex.*

Beweis. Sind g, h konvexe Funktionen und $\lambda \geq 0$, $\mu \geq 0$, so ist auch $\lambda g + \mu h$ konvex. Wir müssen daher nur zeigen, dass für eine konvexe Funktion g, eine reelle Zahl d und eine Barriere-Funktion $\mathcal{B}$ die Funktion $x \mapsto \mathcal{B}(d - g(x))$ konvex ist.
Seien dazu $x, y \in \mathbb{R}^n$ und $\lambda \in [0, 1]$. Aus der Konvexität von g folgt dann

$$\begin{aligned} d - g(\lambda x + (1-\lambda)y) &\geq d - \lambda g(x) - (1-\lambda)g(y) \\ &= \lambda[d - g(x)] + (1-\lambda)]d - g(y)]. \end{aligned}$$

Damit folgt

$$\begin{aligned} &\mathcal{B}(d - g(\lambda x + (1-\lambda)y)) \\ &\leq \mathcal{B}(\lambda[d - g(x)] + (1-\lambda)]d - g(y)]) \quad \text{Monotonie von } \mathcal{B} \\ &\leq \lambda\mathcal{B}(d - g(x)) + (1-\lambda)\mathcal{B}(d - g(y)) \quad \text{Konvexität von } \mathcal{B}. \end{aligned}$$ □

Sind f und $f_1, \ldots, f_p$ konvex und $f_{p+1}, \ldots, f_m$ affin, ist das Hilfsproblem (III.6.3) wegen Lemma III.6.11 für jedes μ und d ein konvexes Optimierungsproblem, wie wir es in §III.2 betrachtet haben. Man kann zeigen [2, Lemma 12.1.3]:

SATZ III.6.12. *Die Funktionen f und $f_1, \ldots, f_p$ seien konvex und die Funktionen $f_{p+1}, \ldots, f_m$ seien affin. Die Menge S aus (III.6.2) sei nicht leer und beschränkt. Für die Barriere-Funktion $\mathcal{B}$ gelte*

$$\lim_{t\to\infty} \mathcal{B}'(t) = 0.$$

Dann besitzt das Hilfsproblem (III.6.3) für jedes $\mu > 0$ und jedes $d \in (\mathbb{R}_+^*)^p$ eine Optimallösung $x(\mu, d)$ und es gilt

$$\lim_{\lambda\to 0+} \inf_{x^*} \Big\{ \|x^* - x(\lambda\mu, \lambda d)\| : x^* \textit{ ist eine Optimallösung von (III.6.1), (III.6.2)} \Big\} = 0.$$

Satz III.6.12 führt auf Algorithmus III.6.3.

Algorithmus III.6.3 Barriere-Verfahren für konvexe Optimierungsprobleme

---

Gegeben: Konvexe Funktionen f, f_1, ..., f_p, affine Funktionen f_{p+1}, ..., f_m, Barriere-Funktion $\mathcal{B}$, Vektor $x \in \mathbb{R}^n$ mit $f_j(x) = 0$ für $p+1 \le j \le m$, Parameter μ, Toleranz $\varepsilon > 0$

Gesucht: Minimum x von f in S

1: $e \leftarrow \infty$
2: **while** $e > \varepsilon$ **do**
3: $\quad d \leftarrow$ Vektor in $(\mathbb{R}_+^*)^p$ mit $f_i(x) < d_i$ für $1 \le i \le p$
4: $\quad \lambda \leftarrow$ Zahl in $(0,1)$ mit $f_i(x) < \lambda d_i$ für $1 \le i \le p$
5: $\quad y \leftarrow \operatorname{argmin}\{\Phi(x, \mu, d) : x \in \mathbb{R}^n, \ , \ f_j(x) = 0 \ , \ p+1 \le j \le m\}$
▷ Liniensuche mit Startwert x
6: $\quad e \leftarrow \|x - y\|$, $x \leftarrow y$
7: **end while**

III.7. SQP-Vefahren

Wir betrachten wieder das Problem (III.4.1) (S. 137) aus §III.4. Gemäß Satz III.4.10 (S. 142) ist eine notwendige Bedingung für ein lokales Minimum x^* die Existenz eines Lagrange-Multiplikators y^*, so dass die Karush-Kuhn-Tucker-Bedingungen (III.3.3) (S. 134)

$$\begin{aligned} Df(x^*) + \sum_{i=1}^{m} y_i^* Df_i(x^*) &= 0, \\ f_i(x^*) y_i^* &= 0, \quad 1 \le i \le p, \\ f_j(x^*) &= 0, \quad p+1 \le j \le m, \\ f_i(x^*) &\le 0, \quad 1 \le i \le p, \\ y_i^* &\ge 0, \quad 1 \le i \le p \end{aligned}$$

erfüllt sind. Die Gleichungsnebenbedingungen können wir in der Form

$$\Phi(x^*, y^*) = 0$$

schreiben mit $\Phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ und

$$\Phi(x,y) = \begin{pmatrix} Df(x) + \sum_{i=1}^{m} y_i Df_i(x) \\ y_1 f_1(x) \\ \vdots \\ y_p f_p(x) \\ f_{p+1}(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

Falls f und die f_i zweimal differenzierbar sind, folgt sofort

$$D\Phi(x,y) = \begin{pmatrix} D_x^2\mathcal{L}(x,y) & Df_1(x) & \dots & Df_p(x) & Df_{p+1}(x) & \dots & Df_m(x) \\ y_1 Df_1(x) & f_1(x) & & & & & \\ \vdots & & \ddots & & & 0 & \\ y_p Df_p(x) & & & f_p(x) & & & \\ Df_{p+1}(x) & & & & 0 & & \\ \vdots & & 0 & & & \ddots & \\ Df_m(x) & & & & & & 0 \end{pmatrix}$$

mit

$$D_x^2\mathcal{L}(x,y) = D^2 f(x) + \sum_{i=1}^{m} y_i D^2 f_i(x).$$

Unter den Voraussetzungen von Satz III.4.12 (S. 143) kann man zeigen, dass $D\Phi(x^*, y^*)$ invertierbar ist. Daher liegt der Versuch nahe, eine Lösung von $\Phi(x,y) = 0$ mit dem Newton-Verfahren zu bestimmen. Neben den bekannten Problemen des Newton-Verfahrens stellt sich hier zusätzlich die Schwierigkeit, dass die Ungleichungsnebenbedingungen $f_i(x) \le 0$, $y_i \ge 0$ für $1 \le i \le p$ zu erfüllen sind. Dies kann man bei naiver Anwendung des Newton-Verfahrens nicht erwarten. Daher wird der Newton-Schritt modifiziert.

Um diese Modifikation zu beschreiben definieren wir die Funktion $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{n\times m} \to \mathbb{R}^n \times \mathbb{R}^m$ durch

$$\Psi(x,y,A) = \begin{pmatrix} A & Df_1(x) & \dots & Df_p(x) & Df_{p+1}(x) & \dots & Df_m(x) \\ y_1 Df_1(x) & f_1(x) & & & & & \\ \vdots & & \ddots & & & 0 & \\ y_p Df_p(x) & & & f_p(x) & & & \\ Df_{p+1}(x) & & & & 0 & & \\ \vdots & & 0 & & & \ddots & \\ Df_m(x) & & & & & & 0 \end{pmatrix}.$$

Dann ist

$$D\Phi(x,y) = \Psi(x,y,D_x^2\mathcal{L}(x,y))$$

und der übliche Newton-Schritt hat die Form

$$\begin{aligned} \Psi(x_k, y_k, D_x^2\mathcal{L}(x_k,y_k)) \begin{pmatrix} \Delta x_k \\ \Delta y_k \end{pmatrix} &= -\Phi(x_k, y_k) \\ x_{k+1} &= x_k + \Delta x_k \\ y_{k+1} &= y_k + \Delta y_k. \end{aligned}$$

Die Modifikation des Newton-Schrittes besteht nun in folgendem Ansatz

$$\begin{aligned}
\Psi(x_k, y_{k+1}, D_x^2\mathcal{L}(x_k, y_k)) \begin{pmatrix} \Delta x_k \\ \Delta y_k \end{pmatrix} &= -\Phi(x_k, y_k) \\
x_{k+1} &= x_k + \Delta x_k \\
y_{k+1} &= y_k + \Delta y_k \\
y_{k+1,i} &\geq 0 \qquad 1 \leq i \leq p \\
f_i(x_k) + Df_i(x_k)\Delta x_k &\leq 0 \qquad 1 \leq i \leq p.
\end{aligned}$$

Es sind also die Ungleichungsnebenbedingungen hinzugekommen und $\Psi(x_k, y_k, D_x^2\mathcal{L}(x_k, y_k))$ ist durch $\Psi(x_k, y_{k+1}, D_x^2\mathcal{L}(x_k, y_k))$ ersetzt. Dadurch wird $(\Delta x_k, \Delta y_k)$ nicht mehr durch ein lineares Gleichungssystem bestimmt sondern durch ein nichtlineares System.

Um dieses besser zu verstehen, setzen wir zur Abkürzung

$$B_k = D_x^2\mathcal{L}(x_k, y_k)$$

und schreiben das nicht lineare System aus:

$$\begin{aligned}
B_k\Delta x_k + \sum_{i=1}^{m} \Delta y_{k,i} Df_i(x_k) &= -Df(x_k) - \sum_{i=1}^{m} y_{k,i} Df_i(x_k) \\
y_{k+1,i} Df_i(x_k)\Delta x_k + f_i(x_k)\Delta y_{k,i} &= -y_{k,i} f_i(x_k) \qquad 1 \leq i \leq p \\
Df_j(x_k)\Delta x_k &= -f_j(x_k) \qquad p+1 \leq j \leq m
\end{aligned}$$

bzw.

$$\begin{aligned}
Df(x_k) + B_k\Delta x_k + \sum_{i=1}^{m} y_{k+1,i} Df_i(x_k) &= 0 \\
y_{k+1,i}\Big[f_i(x_k) + Df_i(x_k)\Delta x_k\Big] &= 0 \qquad 1 \leq i \leq p \\
f_j(x_k) + Df_j(x_k)\Delta x_k &= 0 \qquad p+1 \leq j \leq m.
\end{aligned}$$

Zusammen mit den Ungleichungsnebenbedingungen

$$\begin{aligned}
y_{k+1,i} &\geq 0 \quad 1 \leq i \leq p \\
f_i(x_k) + Df_i(x_k)\Delta x_k &\leq 0 \quad 1 \leq i \leq p
\end{aligned}$$

sind dies die Karush-Kuhn-Tucker-Bedingungen für das Hilfsproblem

$$\min_s \Big\{ Df(x_k)s + \frac{1}{2}s^t B_k s : f_i(x_k) + Df_i(x_k)s \leq 0, 1 \leq i \leq p, \\ f_j(x_k) + Df_j(x_k)s = 0, p+1 \leq j \leq m \Big\}.$$

$(\Delta x_k, \Delta y_k)$ lösen die Karush-Kuhn-Tucker-Bedingungen für dieses Hilfsproblem. Die zu minimierende Funktion ist quadratisch, die Nebenbedingungen sind affin. Dies gab dem Verfahren den Namen S*equential* Q*uadratic* P*rogramming* oder kurz *SQP*. Es wird durch Algorithmus III.7.1 realisiert.

Algorithmus III.7.1 SQP-Verfahren

Gegeben: Vektoren $x \in \mathbb{R}^n$, $y \in (\mathbb{R}_+^*)^p \times \mathbb{R}^{m-p}$, Toleranz $\varepsilon > 0$
Gesucht: Minimum x von f in S

1: $e \leftarrow \infty$
2: **while** $e > \varepsilon$ **do**
3: $\quad B \leftarrow D_x^2 \mathcal{L}(x, y)$
4: $\quad (s, z) \leftarrow$ Lösung für die KKT-Bedingungen des Hilfsproblems

$$\min_s \Big\{ Df(x)s + \frac{1}{2} s^t B s : f_i(x) + Df_i(x)s \le 0, 1 \le i \le p, \\ f_j(x) + Df_j(x)s = 0, p+1 \le j \le m \Big\}$$

5: $\quad e \leftarrow \|s\| + \|z\|$, $x \leftarrow x + s$, $y \leftarrow y + z$
6: **end while**

Bemerkung III.7.1. (1) Bei der praktischen Durchführung von Algorithmus III.7.1 wird wie bei Quasi-Newton-Verfahren $D_x^2 \mathcal{L}(x, y)$ durch eine Approximation B ersetzt.
(2) Ist $B = D_x^2 \mathcal{L}(x, y)$, konvergiert Algorithmus III.7.1 lokal quadratisch. Bei geeigneter Approximation von $D_x^2 \mathcal{L}(x, y)$ ist die Konvergenz noch linear.

III.8. Die Simplexmethode von Nelder und Mead

Dieses Verfahren ist nicht zu verwechseln mit dem Simplexalgorithmus aus §I.4. Es wird auf allgemeine, nichtlineare Optimierungsprobleme angewandt. Sein Name rührt daher, dass es mit n-dimensonalen Simplizes arbeitet. Für dieses Verfahren existiert keine Konvergenztheorie und es konvergiert in der Regel sehr langsam. Dennoch ist es in der Praxis sehr beliebt, da es sehr einfach ist, wenige Funktionsauswertungen benötigt, ohne Ableitungen auskommt und relativ schnell eine grobe Approximation des gesuchten Optimums liefert.

Zur Beschreibung der Idee von Algorithmus III.8.1 betrachten wir zunächst das nicht restringierte Problem

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

Wir nehmen an, dass wir $n+1$ affin unabhängige Punkte $x_0, \ldots, x_n \in \mathbb{R}^n$ kennen. Wir sortieren diese Punkte so, dass

$$f(x_0) \le \ldots \le f(x_n)$$

ist. Die Punkte $x_0, \ldots, x_{n-1}$ erzeugen eine Hyperebene, die den $\mathbb{R}^n$ in zwei Teilräume aufspaltet. In dem einen liegt der „schlechte“ Punkt x_n. Die Hoffnung ist, in dem anderen Teilraum einen besseren Punkt zu finden. Dazu wird x_n am Schwerpunkt von $x_0, \ldots, x_{n-1}$ gespiegelt; das Ergebnis sei x_n'. Falls

$$f(x_n') \le f(x_n)$$

ist, wird x_n durch x'_n ersetzt. Falls

$$f(x'_n) > f(x_n)$$

ist, lässt man den Simplex „schrumpfen“. Falls $f(x'_n)$ „sehr klein“ ist, „expandiert“ man den Simplex.

Bemerkung III.8.1. Für restringierte Probleme der Form

$$\min\{f(x) : x \in K\}$$

wählt man zu Beginn von Algorithmus III.8.1 die Punkte $x_0, \ldots, x_n$ in K und ersetzt die Funktionswerte f_r, f_e bzw. f_c durch ∞, falls der betreffende Punkt nicht in K liegt.

III.9. Globale Optimierung

Alle bisher betrachteten Algorithmen liefern bestenfalls ein *lokales Minimum*. In manchen Anwendungen benötigt man aber ein oder so gar alle *globalen Minima*. Diesen Problemkreis nennt man *globale Optimierung*.

Alle Algorithmen der globalen Optimierung haben folgende gemeinsame Struktur:

- Untersuche mehrere Kandidaten für globale Minima.
- Ersetze gegebenenfalls einen Kandidaten durch das Ergebnis einer *lokalen Suche*, d.h. wende einen der bisher betrachteten Algorithmen mit dem gegebenen Kandidaten als Startnäherung an.
- Iteriere gegebenenfalls über eine Liste von Kandidaten.
- Störe gegebenenfalls Kandidaten.

Dia Algorithmen unterscheiden sich durch

- die Auswahl der Kandidaten,
- die Methoden zur Verbesserung der Kandidaten,
- die Art der Störung,
- den Grad an Zufall,
- den Aufwand für die lokale Suche.

Im Folgenden beschreiben wir diese Punkte kurz.

Man unterscheidet zwei Varianten für die Wahl der Kandidaten: *deterministisch* und *zufällig*. Bei der deterministischen Variante überdeckt man die zulässige Menge $S \subset \mathbb{R}^n$ durch ein gleichmäßiges Gitter (vgl. den linken Teil von Abbildung III.9.1). Bei der Zufallsvariante dagegen überdeckt man S durch ein Gitter zufälliger Punkte bezüglich eines gewählten Wahrscheinlichkeitsmaßes wie z. B. der Gleichverteilung (vgl. den rechten Teil von Abbildung III.9.1). Bei beiden Varianten konstruiert man gegebenenfalls mehrere Listen von Kandidaten.

In der Regel werden mehrere Varianten zur Verbesserung von Kandidaten genutzt:

Algorithmus III.8.1 Simplexmethode von Nelder und Mead für Probleme ohne Nebenbedingungen

Gegeben: $n+1$ affin unabhängige Punkte $x_0, \ldots, x_n \in \mathbb{R}^n$, Funktion f, Toleranz $\varepsilon > 0$
Gesucht: Minimum von f

for $i = 0, 1, \ldots, n$ **do** $f_i \leftarrow f(x_i)$ ▷ Initialisierung
end for
Sortiere die Punkte so um, dass $f_0 \leq \ldots \leq f_n$ ist.
$d \leftarrow \infty$
while $d > \varepsilon$ **do**
    $\overline{f} \leftarrow \dfrac{1}{n+1} \displaystyle\sum_{i=0}^{n} f(x_i)$, $d \leftarrow \dfrac{1}{n+1} \displaystyle\sum_{i=0}^{n} \big(f(x_i) - \overline{f}\big)^2$
    $c \leftarrow \dfrac{1}{n} \displaystyle\sum_{i=0}^{n-1} x_i$, $x_r \leftarrow 2c - x_n$, $f_r \leftarrow f(x_r)$
    if $f_0 \leq f_r \leq f_{n-1}$ **then** ▷ Reflexionsschritt
        $x_n \leftarrow x_r$, $f_n \leftarrow f_r$
    end if
    if $f_r < f_0$ **then** ▷ Expansionsschritt
        $x_e \leftarrow 2x_r - c$, $f_e \leftarrow f(x_e)$
        if $f_e < f_r$ **then**
            $x_r \leftarrow x_e$
        end if
        $x_n \leftarrow x_r$, $f_n \leftarrow f_r$
    end if
    if $f_r > f_{n-1}$ **then** ▷ Kontraktionsschritt
        if $f_r \geq f_n$ **then**
            $x_c \leftarrow \frac{1}{2}(x_n + c)$
        else
            $x_c \leftarrow \frac{1}{2}(x_r + c)$
        end if
        $f_c \leftarrow f(x_c)$
        if $f_c < \min\{f_n, f_r\}$ **then**
            $x_n \leftarrow x_c$, $f_n \leftarrow f_c$
        else
            for $i = 1, \ldots, n$ **do**
                $x_i \leftarrow \frac{1}{2}(x_0 + x_i)$, $f_i \leftarrow f(x_i)$
            end for
        end if
    end if
    Sortiere die Punkte so um, dass $f_0 \leq \ldots \leq f_n$ ist.
end while

- Benutze eine lokale Suche mit einem der in den vorigen Abschnitten betrachteten Verfahren.

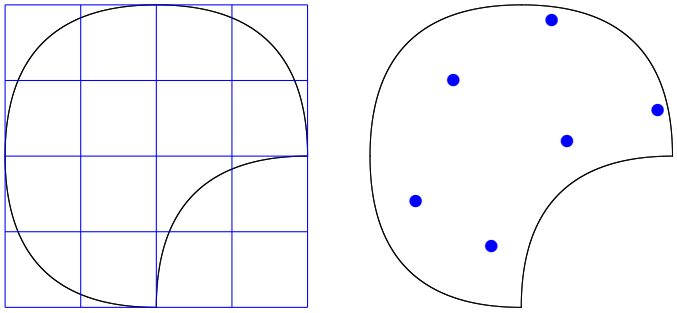

ABBILDUNG III.9.1. Deterministisches (links) und zufälliges (rechts) Gitter für eine zulässige Menge S

- Störe Kandidaten durch die unten beschriebene Technik.
- Akzeptiere mit einer kleinen Wahrscheinlichkeit eine Vergrößerung der Zielfunktion f. Bei dem sog. *simulated annealing* wird z. B. ein Punkt x' mit $f(x') > f(x)$ dem Punkt x mit Wahrscheinlichkeit $e^{\frac{f(x)-f(x')}{T}}$ vorgezogen. Dabei ist T die sog. *Abkühlzeit* (engl. *cooling time*), die auf Basis *heuristischer* Kriterien gewählt und im Laufe des Algorithmus angepasst wird.
- Gegebenfalls werden Kandidatenlisten auch mit *Verzweigungstechniken* (engl. *branch and bound*) angepasst.

Eine beliebte Störungstechnik geht wie folgt vor:

- Normalisiere alle Punte so, dass ihre Koordinaten durch einen String mit N Bits dargestellt werden können.
- Für jeden Kandidaten wähle zufällig ein Bit seiner Darstellung aus und schalte es zufällig um.

Mit dieser Technik kann man leicht Störungen mit einem großen Abstand in der euklidischen Norm konstruieren.

BEISPIEL III.9.1. Für $N = 4$ und

$$x = 15 = 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

sind

$$x' = 11 = 1 \cdot 2^3 + \mathbf{0} \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

und

$$x'' = 7 = \mathbf{0} \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

zulässige Störungen von x, während

$$\widetilde{x} = 9 = 1 \cdot 2^3 + \mathbf{0} \cdot 2^2 + \mathbf{0} \cdot 2^1 + 1 \cdot 2^0$$

keine zulässige Störung ist.

Bei der Auswahl der Optimierungsmethode sollte man nie vergessen, dass jeder Algorithmus seine Vor- und Nachteile hat und dass es keinen effizienten universellen Black-Box-Algorithmus gibt, der alle denkbaren Probleme löst.

Literaturverzeichnis

[1] K.-H. Borgwardt, *Optimierung, Operations Research, Spieltheorie*, Birkhäuser, Basel, 2001, Mathematische Grundlagen.
[2] F. Jarre and J. Stoer, *Optimierung*, Springer, Berlin, 2004.

Index