Universität Karlsruhe (TH)
Mathematisches Institut II
Prof. Dr. Andreas Kirsch

Optimierungstheorie

Skript zur Vorlesung im Sommersemester 2005

Dozent: Prof. Dr. Andreas Kirsch
Tel. 608 2050, Zimmer 208.2
Sprechstunde: Donnerstags, 8.30 – 10.30 Uhr
email: kirsch@math.uni-karlsruhe.de

Übungen: Frau PD Dr. Natalia Grinberg
Tel. 608 7574, Zimmer 204.2
Sprechstunde: Montags 10.30 - 11.30 Uhr
email: grinberg@math.uni-karlsruhe.de

`http://www.mathematik.uni-karlsruhe.de/mi2kirsch/de`

Literatur:

- E. Blum, W. Oettli: Mathematische Optimierung. Springer, 1975.
- J.C.G. Boot: Quadratic Programming. North-Holland, 1964.
- L. Brickman: Mathematical Introduction to Linear Programming and Game Theory. Springer, 1989.
- V. Chvátal: Linear Programming. Freeman, 1983.
- L. Collatz, W. Wetterling: Optimierungsaufgaben. Heidelberger Taschenbuch. Springer 1966.
- J. Franklin: Methods of Mathematical Economics. Springer, 1980.
- D. Gale: The Theory of Linear Economic Models. McGrawHill, 1960.
- C. Geiger. C. Kanzow: Theorie und Numerik restringierter Optimierungsaufgaben. Springer, 2003.
- K. Glashoff, S.-A. Gustafson: Einführung in die lineare Optimierung. Wiss. Buchgesellschaft, Darmstadt, 1978.
- G. Hadley: Linear Programming. Addison-Wesley, 1962.
- G. Hämmerlin, K.-H. Hoffmann: Numerische Mathematik. Grundwissen Mathematik 7, Springer, 1989.
- H.W. Hamacher, K. Klamroth: Lineare und Netzwerk-Optimierung. Vieweg, 2002
- R. Henn, H.P. Künzi: Einführung in die Unternehmensforschung II. Heidelberger Taschenbuch, Springer 1968.
- J. Jahn: Introduction to the theory of Nonlinear Optimization. (2nd edition) Springer 1996.
- F. Jarre, J. Stoer: Optimierung. Springer, 2000.
- D. Jungnickel: Optimierungsmethoden. Springer, 2003.
- D.G. Luenberger: Introduction to Linear and Nonlinear Programming. Addison-Wesley, 1965.
- K.G. Murty: Linear and Combinatorical Programming. J. Wiley, 1976.
- F. Noźićka, J. Guddat, H. Hollatz: Theorie der linearen Optimierung. Akademie-Verlag, 1972.
- G. Owen: Spieltheorie. Springer, 1971.
- R. Reemtsen: Lineare Optimierung. Shaker 2001.
- G. Schmeißer, H. Schirmeier: Praktische Mathematik. de Gruyter, 1976.
- A. Schrijver: Theory of Linear and Integer Programming. Wiley, 1986.
- H.R. Schwarz: Numerische Mathematik. Teubner, 1986.
- W.A. Spivey, R.M. Thrall: Linear Optimization. Holt, Rinehart and Winston, 1970.
- W. Vogel: Lineares Optimieren. Akademische Verlagsgesellschaft, 1967.
- H.H. Weber: Lineare Programmierung. Studientext. Akademische Verlagsgesellschaft, 1973.
- H.H. Weber: Einige Erweiterungen der linearen Programmierung. Studientext. Akademische Verlagsgesellschaft, 1975.
- J. Werner: Optimization: Theory and Applications. Vieweg, 1984.
- J. Werner: Numerische Mathematik 2. Vieweg, 1992.

Inhaltsverzeichnis

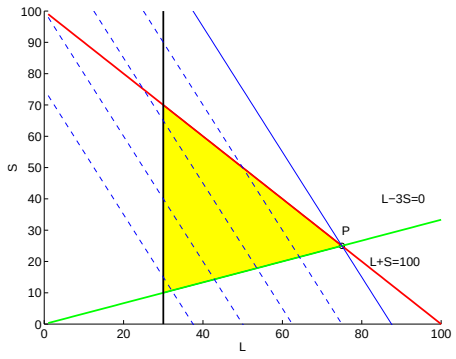
1 Einführung

1.1 Beispiele

(1) Eine Bank habe 100 Millionen Euro zum Anlegen. Ein Teil wird in Darlehen angelegt (L Millionen Euro, L wie loan), ein Teil in Wertpapieren (S Millionen, S wie securities). Darlehen erzielen hohe Rendite (z.B. 10%), sind aber langfristig angelegt, Wertpapiere erreichen z.B. 5% Verzinsung. Die Bank erhält also jährlich $0{,}1\,L + 0{,}05\,S$ an Zinsen. Diese Größe ist unter Berücksichtigung gewisser Nebenbedingungen zu maximieren.

1. Vorzeichenbedingungen: $L \geq 0$, $S \geq 0$
2. Beschränkung des angelegten Geldes: $L + S \leq 100$
3. Liquiditätsbeschränkung: Angenommen, die Bank will (oder muss) wenigstens 25% des angelegten Geldes als Wertpapiere halten, d.h. $S \geq 0,25\,(L+S)$ oder $L-3S \leq 0$
4. Mindestdarlehensanlage: Bestimmte Großkunden wollen Darlehensanlagen, z.B. $L \geq 30$.

Ein Paar (L, S), welches die Nebenbedingungen 1. - 4. erfüllt, heißt **zulässige** Anlage. Gesucht ist diejenige zulässige Anlage, die die Größe $0{,}1\,L+0{,}05\,S$ maximiert. Wir können dieses Problem **graphisch** lösen:



Welcher Punkt im Dreieck ist optimal?

Wir zeichnen die Geraden $\frac{1}{10}L + \frac{1}{20}S$ = constant, d.h. die Geraden $2L + S = c$ für verschiedene Konstanten $c > 0$. Die Gerade mit Steigung -2 ist offensichtlich optimal, wenn sie durch P geht. Damit ist $P = (L^*, S^*)$ optimal!

P ist charakterisiert durch das Gleichungssystem $\left\{ \begin{array}{rcl} L^* + S^* & = & 100 \\ L^* - 3S^* & = & 0 \end{array} \right\}$, also $L^* = 75$, $S^* = 25$ mit Gewinn 8,75 Millionen Euro.

(2) Im Jahr 1945 erschien eine Arbeit von George Stigler: The Cost of Subsistence (Die Kosten des Lebensunterhalts) in: The Journal of Farm Economics.
Was sind die minimalen Kosten einer ausgewogenen Ernährung?
Angenommen, man kann sich durch n Nahrungsmittel ernähren. x_j sei die Menge des j-ten Nahrungsmittels. Ein Diätplan ist ein n-Tupel $(x_1, \ldots, x_n)$. Natürlich muss gelten: $x_j \geq 0$, $j = 1, \ldots, n$. Die Kosten seien c_j pro Einheit. Dann betragen die Gesamtkosten $\sum_{j=1}^{n} c_j x_j$, und dies ist zu minimieren. Zusätzliche Nebenbedingungen können sein: Jede Diät muss eine bestimmte Mindestmenge an Nährstoffen und Vitaminen enthalten. Eine Einheit des j-ten Nahrungsmittels enthalte a_{ij} Einheiten des Nährstoffs i. Die Gesamtheit des aufgenommenen Nährstoffs i ist also $\sum_{j=1}^{n} a_{ij} x_j$. Damit lautet die Nebenbedingung: $\sum_{j=1}^{n} a_{ij} x_j \geq b_i$ für jedes $i = 1, \ldots, m$, wobei b_i die Mindestmenge des Nährstoffes i sei.
1945 konnte man solche Probleme für große n, m noch nicht lösen, die Computer waren noch nicht gut genug!

(3) Transportproblem: m Raffinerien $R_1, \ldots, R_m$ einer Ölgesellschaft beliefern n Tanklager $T_1, \ldots, T_n$. Die Raffinerie R_i kann höchstens r_i Einheiten liefern, das Lager T_j benötigt mindestens t_j Einheiten. Der Transport von R_i nach T_j kostet C_{ij} Euro pro Einheit. Gesucht ist die Anzahl X_{ij} von Einheiten, die von R_i nach T_j geschafft werden. (X_{ij}) ist zulässig, falls $\sum_{j=1}^{n} X_{ij} \leq r_i$ für alle $i = 1, \ldots, m$, und $\sum_{i=1}^{m} X_{ij} \geq t_j$ für alle $j = 1, \ldots, n$. Außerdem muss natürlich gelten: $X_{ij} \geq 0$ für alle i, j. Die Gesamtkosten $\sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij}$ sind zu minimieren!

(4) Diskrete **Tschebyscheffapproximation:** Gegeben sei ein überbestimmtes Gleichungssystem

$$\sum_{j=1}^{n} a_{ij} x_j \stackrel{!}{=} b_i, \quad i = 1, \ldots, m,$$

wobei $a_{ij}, b_i \in \mathbb{R}$ gegeben seien mit $m > n$. Dann besitzt dieses Gleichungssystem i.A. keine Lösung $x \in \mathbb{R}^n$.
Abhilfe: Wir wollen die Fehler

$$\left| \sum_{j=1}^{n} a_{ij} x_j - b_i \right|$$

möglichst klein machen für alle $i = 1, \ldots, m$. Eine Möglichkeit, den Fehler im Gleichungssystem zu messen, ist durch die Maximumnorm gegeben:

$(P1)$ Minimiere $\max_{i=1,\ldots,m} \left| \sum_{j=1}^{n} a_{ij} x_j - b_i \right|$ über alle $x \in \mathbb{R}^n$!

Wir können dies folgendermaßen umformulieren:
$(P2)$ Minimiere $x_0 \in \mathbb{R}$ unter den Nebenbedingungen

$$-x_0 \;\le\; \sum_{j=1}^{n} a_{ij}\, x_j - b_i \;\le\; x_0 \quad \text{für alle } i = 1, \ldots, m.$$

Es ist eine einfache Übung zu zeigen, dass (P_1) und (P2) äquivalent sind, d.h. genauer: Ist x^* Lösung von (P_1), so ist (x_0^*, x^*) Lösung von (P2), wobei

$$x_0^* \;:= \max_{i=1,\ldots,m} \left| \sum_{j=1}^{n} a_{ij}\, x_j^* - b_i \right| .$$

Ist (x_0^*, x^*) Lösung von (P2), so ist x^* Lösung von (P_1).
Diese Umformulierung von (P_1) geht auf E. Stiefel (≈ 1960) zurück!

(5) Ausgleichsproblem mit Nebenbedingung: Wir betrachten dasselbe überbestimmte Gleichungssystem wie oben:

$$\sum_{j=1}^{n} a_{ij}\, x_j \;\stackrel{!}{=}\; b_i\,, \quad i = 1, \ldots, m,$$

suchen aber jetzt nicht-negative Lösungen $x_j \ge 0$ für alle $j = 1, \ldots, n$. Eine Möglichkeit besteht in der Minimierung von

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}\, x_j - b_i \right)^2 \quad \text{unter den Nebenbedingungen} \quad x_j \ge 0\,,\ j = 1, \ldots, n\,.$$

Dies ist ein **quadratisches** Problem, wie wir später sehen werden. Dieses Problem ohne die Vorzeichenbedingungen wurde schon von Gauß (≈ 1800) behandelt und ist unter dem Namen „Methode der kleinsten Quadrate“ bekannt.

1.2 Problemstellung

Die Beispiele 1–4 des letzten Abschnitts lassen sich auf folgende Form bringen:

(P1)
$$\begin{cases} \text{Minimiere } \displaystyle\sum_{j=1}^{n} c_j\, x_j \text{ unter den Nebenbedingungen} \\ x \in \mathbb{R}^n \quad \text{und} \quad \displaystyle\sum_{j=1}^{n} a_{ij}\, x_j \;\le\; b_i, \quad i = 1, \ldots, m\,. \end{cases}$$

Daneben betrachten wir noch Optimierungsprobleme der Gestalt

(P2)
$$\begin{cases} \text{Minimiere } \displaystyle\sum_{j=1}^{n} c_j\, x_j \text{ unter den Nebenbedingungen} \\ x_j \ge 0\,,\ j = 1, \ldots, n\,, \quad \text{und} \quad \displaystyle\sum_{j=1}^{n} a_{ij}\, x_j \;=\; b_i,\ i = 1, \ldots, m\,. \end{cases}$$

(P1) und (P2) nennt man **lineare, endlichdimensionale** Probleme, da sowohl die **Zielfunktion**, als auch die **Nebenbedingungen** durch endlich viele lineare Funktionale beschrieben werden.

Formulierung in der Matrix-Vektor Schreibweise:
Vektoren $a, b, c, x, y, z \in \mathbb{R}^n$ schreiben wir immer als **Spaltenvektoren**!
$(m\times n)-$Matrizen bezeichnen wir mit großen lateinischen Buchstaben, etwa $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \in \mathbb{R}^{m\times n}$. Dann ist $A^\top = (a_{ji})_{\substack{j=1,\dots,n\\ i=1,\dots,m}} \in \mathbb{R}^{n\times m}$ die transponierte Matrix. Wir fassen den Vektor $x \in \mathbb{R}^n$ auch als $(n \times 1)-$Matrix auf, dann ist $x^\top$ gerade der **Zeilenvektor** in $\mathbb{R}^{1\times n}$.

Die verschiedenen Möglichkeiten, Vektoren zu multiplizieren, sind dann Spezialfälle der Matrizenmultiplikation $AB \in \mathbb{R}^{m\times p}$ für $A \in \mathbb{R}^{m\times n}$ und $B \in \mathbb{R}^{n\times p}$.

Das **Skalarprodukt** ist $$x^\top y \;=\; \sum_{j=1}^{n} x_j y_j \quad \text{für } x = (x_j),\ y = (y_j) \in \mathbb{R}^n\,,$$

Das **dyadische Produkt** ist $$x\, y^\top = (x_i y_j)_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \quad \text{für } x \in \mathbb{R}^m \text{ und } y \in \mathbb{R}^n\,.$$

Für $x, y \in \mathbb{R}^n$ führen wir die Halbordnung ein:

$$x \le y \iff x_j \le y_j \text{ für alle } j = 1, \dots, n\,,$$

sowie die Normen $\|\cdot\|_2$ und $\|\cdot\|_\infty$ durch

$$\|x\|_2 \;:=\; \sqrt{x^\top x} \;=\; \sqrt{\sum_{j=1}^{n} x_j^2}\,, \qquad \|x\|_\infty \;:=\; \max_{j=1,\dots,n} |x_j|\,.$$

Damit können wir (P1) und (P2) schreiben als:

(P1) $$\text{Minimiere} \quad c^\top x \quad \text{unter} \quad x \in \mathbb{R}^n\,,\ Ax \le b\,,$$

(P2) $$\text{Minimiere} \quad c^\top x \quad \text{unter} \quad x \ge 0\,,\ Ax = b\,,$$

wobei $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ gegeben sind und $x \in \mathbb{R}^n$ gesucht ist.

Bezeichnung:
Mit M bezeichnen wir im folgenden immer die Menge der **zulässigen Punkte**, d.h.

$$M \;=\; \{x \in \mathbb{R}^n : Ax \le b\} \quad \text{bzw.} \quad M \;=\; \{x \in \mathbb{R}^n : x \ge 0,\ Ax = b\}\,.$$

Definition 1.1 *Sei (P) eines der Probleme (P1) oder (P2). Das Problem heißt **zulässig**, falls $M \neq \emptyset$. Der Vektor $x^* \in \mathbb{R}^n$ heißt **Lösung** von (P), falls:*

*(i) x^* **zulässig** ist, d.h. $x \in M$, und*

*(ii) x^* **optimal** ist, d.h. $c^\top x^* \le c^\top x$ für alle $x \in M$.*

*In diesem Fall heißt das Problem **lösbar**.*

Ferner setzen wir

$$\inf(P) \;:=\; \begin{cases} \inf\{c^\top x : x \in M\}\,, & \text{falls } M \neq \emptyset\,, \\ +\infty\,, & \text{falls } M = \emptyset\,, \end{cases}$$

und

$$\min(P) \;=\; \min\{c^\top x : x \in M\} \;=\; c^\top x^*\,,$$

falls (P) lösbar ist mit Lösung $x^* \in M$. Natürlich kann auch $\inf(P) = -\infty$ sein!

Beide Probleme (P1) und (P2) sind **äquivalent** im folgenden Sinn:

(i) Sei $x \in \mathbb{R}^n$ zulässig für (P1), d.h. $Ax \leq b$. Sei $u_j := \max\{x_j, 0\}$, $v_j := \max\{-x_j, 0\} = -\min\{x_j, 0\}$ und $y := b - Ax$. Dann ist $u \geq 0$, $v \geq 0$, $y \geq 0$, $x = u - v$, und

$$\big[A \,|\, -A \,|\, I\big] \begin{bmatrix} u \\ v \\ y \end{bmatrix} \;=\; b\,,$$

d.h. $\begin{bmatrix} u \\ v \\ y \end{bmatrix}$ ist zulässig für ein Problem in der Form (P2). Hier haben wir die **Blockschreibweise** für Matrizen und Vektoren benutzt. Die Zielfunktion ist jetzt $c^\top u - c^\top v = [c^\top \,|\, -c^\top \,|\, 0] \begin{bmatrix} u \\ v \\ y \end{bmatrix}$. Ist x^* optimal für (P1), so ist das zugehörige $\begin{bmatrix} u^* \\ v^* \\ y^* \end{bmatrix}$ optimal für (P2).

(ii) Andererseits lässt sich (P2) auch schreiben in der Form: Minimiere $c^\top x$ unter den Nebenbedingungen

$$x \in \mathbb{R}^n\,, \quad Ax \leq b\,, \quad -Ax \leq -b\,, \quad -x \leq 0\,,$$

d.h. unter

$$\begin{bmatrix} A \\ -A \\ -I \end{bmatrix} x \;\leq\; \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix}.$$

Dies ist in der Form (P1).

Überführt man ein Problem in der Form (P1) in ein äquivalentes in der Form (P2), so führt man also die **Schlupfvariablen** $y = b - Ax$ ein, vergrößert also die Anzahl der Unbekannten. Umgekehrt vergrößert man die Zahl der **Restriktionen**, wenn man (P2) in (P1) überführt!

Beispiel 1.2

$$\textit{Minimiere} \quad x_1 + 3x_2 + 4x_3 \quad \textit{unter} \quad x_2 \geq 0\,,\; x_3 \geq 0 \quad \textit{und}$$

$$\begin{aligned} x_1 \;+\; 2x_2 \;+\; x_3 \;&=\; 5 \\ 2x_1 \;+\; 3x_2 \;+\; x_3 \;&\geq\; 6 \end{aligned}$$

Umformulierung *auf (P1): Setze* $c = (1,3,4)^\top$ *und* $x = (x_1, x_2, x_3)^\top$ *und*

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -2 & -1 \\ -2 & -3 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 5 \\ -5 \\ -6 \\ 0 \\ 0 \end{pmatrix}.$$

Dann ist das Problem, $c^\top x$ *zu minimieren unter* $Ax \leq b$.

Umformulierung *auf (P2):*

1. Möglichkeit:

Setze $y := 2x_1 + 3x_2 + x_3 - 6\,, \quad x_1 = u - v\,, \quad c = (1,-1,3,4,0)^\top$

$$A = \begin{pmatrix} 1 & -1 & 2 & 1 & 0 \\ 2 & -2 & 3 & 1 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

Mit $z := (u, v, x_2, x_3, y)^\top \in \mathbb{R}^5$ *haben wir dann das Problem:*

$$\textit{Minimiere} \quad c^\top z \quad \textit{unter} \quad z \geq 0 \quad \textit{und} \quad Az = b\,.$$

2. Möglichkeit: *(besser!) Löse in einer Gleichungsnebenbedingung nach einer Variablen auf, die nicht vorzeichen-beschränkt ist und eliminiere diese aus dem System: Es ist* $x_1 = 5 - 2x_2 - x_3$. *Damit lautet unser reduziertes Problem:*

Minimiere $x_2 + 3x_3 + 5$ *unter*

$$-x_2 - x_3 \geq -4\,, \quad x_2 \geq 0\,, \quad x_3 \geq 0\,.$$

Wir setzen

$$c := (1,3,0)^\top, \quad y := 4 - x_2 - x_3\,, \quad A = (1\ 1\ 1)\,, \quad b = 4$$

und haben:

$$\textit{Minimiere} \quad c^\top \begin{pmatrix} x_2 \\ x_3 \\ y \end{pmatrix} \quad \textit{unter} \quad x_2 \geq 0, \quad x_3 \geq 0, \quad y \geq 0\,, \quad A \begin{pmatrix} x_2 \\ x_3 \\ y \end{pmatrix} = b\,.$$

Wir können also unser Optimierungsproblem immer in der Form (P1) oder (P2) schreiben, je nachdem, was uns gerade besser gefällt! Bei praktischen Rechnungen kann eine Form vorteilhafter sein!

Wir betrachten noch einmal **Beispiel 4** und schreiben es in die Form (P1) um. Wir führen wieder eine zusätzliche Variable $x_0 \in \mathbb{R}$ ein und haben:

$$\text{Minimiere} \quad x_0 \quad \text{unter} \quad \|Ax-b\|_\infty \;=\; \max_{i=1,\dots,m} \left| b_i - \sum_{j=1}^n a_{ij}\, x_j \right| \;\le\; x_0\,.$$

Hier haben wir die Schreibweise $\|\cdot\|_\infty$ für die Maximumnorm benuzt. Damit ist $c = (1,0,\dots,0)^\top \in \mathbb{R}^{n+1}$, und die Nebenbedingungen haben die Form

$$\begin{aligned} -\sum_{j=1}^n a_{ij}\, x_j \;-\; x_0 \;&\le\; -b_i\,, \quad i=1,\dots,m, \\ \sum_{j=1}^n a_{ij}\, x_j \;-\; x_0 \;&\le\; b_i\,, \quad i=1,\dots,m\,. \end{aligned}$$

Damit haben wir das Problem, $c^\top x$ zu minimieren unter den Nebenbedingungen $x \in \mathbb{R}^{n+1}$ mit

$$\begin{pmatrix} -1 & -a_{11} & \cdots & -a_{1n} \\ \vdots & \vdots & & \vdots \\ -1 & -a_{m1} & \cdots & -a_{mn} \\ -1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ -1 & a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \;\le\; \begin{pmatrix} -b_1 \\ \vdots \\ -b_m \\ b_1 \\ \vdots \\ b_m \end{pmatrix}$$

Dieses Problem hat $n+1$ Unbekannte und $2m$ Nebenbedingungen.

Das **Beispiel 5** ist ein typisches quadratisches Optimierungsproblem. Die Zielfunktion ist

$$\begin{aligned} f(x) \;&=\; \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}\, x_j - b_i \right)^2 \\ &=\; \sum_{i=1}^m \left[\sum_{j,k=1}^n a_{ij}\, a_{ik}\, x_j\, x_k - 2 \sum_{j=1}^n a_{ij}\, x_j\, b_i + b_i^2 \right] \\ &=\; \sum_{j,k=1}^n x_j\, x_k \sum_{i=1}^m a_{ij}\, a_{ik} \;-\; 2 \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij}\, b_i \;+\; \sum_{i=1}^m b_i^2\,. \end{aligned}$$

Man definiere jetzt $Q \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$ und $\beta \in \mathbb{R}$ durch

$$Q_{jk} \;:=\; \sum_{i=1}^m a_{ij}\, a_{ik}\,, \quad c_j \;:=\; \sum_{i=1}^m a_{ij}\, b_i\,, \quad j,k=1,\dots,n, \quad \beta \;:=\; \sum_{i=1}^m b_i^2\,,$$

also $Q = A^\top A$, $c = A^\top b$, $\beta = b^\top b$, und erhält:

$$f(x) \;=\; x^\top Q\, x \;-\; 2\, c^\top x \;+\; \beta\,.$$

Eine solche Funktion nennt man **quadratisches Funktional**!

2 Konvexe Mengen und Polyeder

2.1 Konvexe Mengen

Die Restriktionenmengen der Optimierungsprobleme (P1) und (P2) haben eine spezielle Form, sie sind z.B. konvex. Die meisten der folgenden Begriffe sind wahrscheinlich schon bekannt:

Definition 2.1 *(a) Eine Menge $V \subseteq \mathbb{R}^n$ heißt* **linearer Teilraum** *oder* **Unterraum**, *wenn mit $x, y \in V$ und $\lambda, \mu \in \mathbb{R}$ auch $\lambda x + \mu y \in V$ gilt.*

(b) Eine Menge $V \subseteq \mathbb{R}^n$ heißt **affiner Teilraum**, *wenn mit $x, y \in V$ und $\lambda \in \mathbb{R}$ auch $(1-\lambda)x + \lambda y \in V$ gilt.*

(c) Eine Menge $M \subseteq \mathbb{R}^n$ heißt **konvex**, *wenn mit $x, y \in M$ und $\lambda \in [0,1]$ auch $(1-\lambda)x + \lambda y \in M$ gilt.*

(d) Eine Menge $K \subseteq \mathbb{R}^n$ heißt **Kegel**, *wenn mit $x \in K$ und $\lambda \geq 0$ auch $\lambda x \in K$ gilt.*

Man skizziere jeweils Beispiele und Gegenbeispiele für diese Begriffe im $\mathbb{R}^2$! Wir bemerken, dass bei uns Kegel immer die „Spitze“ im Ursprung haben. Kegel mit Spitze in $p \in \mathbb{R}^n$ beschreiben wir durch $p + K = \{p + x : x \in K\}$. Das folgende Lemma gibt äquivalente Definitionen für die Begriffe der Definition an:

Lemma 2.2 *(a) $V \subseteq \mathbb{R}^n$ ist genau dann ein linearer Teilraum, wenn für alle $x^{(i)} \in V$ und alle $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, m$, m beliebig, auch $\sum_{i=1}^m \lambda_i x^{(i)} \in V$ gilt.*

(b) $V \subseteq \mathbb{R}^n$ ist genau dann ein affiner Teilraum, wenn für alle $x^{(i)} \in V$ und alle $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, m$, m beliebig, mit $\sum_{i=1}^m \lambda_i = 1$ auch $\sum_{i=1}^m \lambda_i x^{(i)} \in V$ gilt.

(c) $M \subseteq \mathbb{R}^n$ ist genau dann konvex, wenn für alle $x^{(i)} \in M$ und alle $\lambda_i \geq 0$, $i = 1, \ldots, m$, m beliebig, mit $\sum_{i=1}^m \lambda_i = 1$ auch $\sum_{i=1}^m \lambda_i x^{(i)} \in M$ gilt.

(d) $K \subseteq \mathbb{R}^n$ ist genau dann ein **konvexer Kegel**, *wenn für alle $x^{(i)} \in K$ und alle $\lambda_i \geq 0$, $i = 1, \ldots, m$, m beliebig, auch $\sum_{i=1}^m \lambda_i x^{(i)} \in K$ gilt.*

Beweis: Insgesamt müssen acht Richtungen bewiesen werden. Drei davon sind trivial, denn gilt (a), (b), (c) des Lemmas, so auch (a), (b), (c) der Definition (man setze $m = 2$). Gilt (d) des Lemmas, so ist K ein Kegel (setze $m = 1$) und auch konvex (setze $m = 2$ und wähle $\lambda_i \in [0,1]$). Die vier Rückrichtungen werden mit vollständiger Induktion bewiesen. Wir führen das nur für die Konvexität durch. Sei also M konvex. Für $m = 2$ reduziert sich die Aussage des Lemmas auf die Definition, ist also richtig. Sei die Aussage des Lemmas also jetzt für m und beliebige $x^{(i)} \in M$ und $\lambda_i \geq 0$, $i = 1, \ldots, m$, mit $\sum_{i=1}^m \lambda_i = 1$ gültig. Es seien jetzt $x^{(1)}, \ldots, x^{(m+1)} \in M$ und $\lambda_1, \ldots, \lambda_{m+1} \geq 0$ mit $\sum_{i=1}^{m+1} \lambda_i = 1$ gegeben. Ohne Einschränkung können wir sogar $\lambda_i > 0$ annehmen. Wir setzen $\mu := \sum_{i=1}^m \lambda_i$ und erhalten

$$\sum_{i=1}^{m+1} \lambda_i x^{(i)} \;=\; \sum_{i=1}^{m} \lambda_i x^{(i)} \;+\; \lambda_{m+1} x^{(m+1)} \;=\; \mu \underbrace{\sum_{i=1}^{m} \frac{\lambda_i}{\mu} x^{(i)}}_{\in M} \;+\; (1-\mu)\, x^{(m+1)} \;\in\; M\,.$$

□

Für eine gegebene (endliche oder unendliche) Teilmenge $S \subseteq \mathbb{R}^n$ können wir den kleinsten linearen Raum (bzw affinen Raum, oder konvexe Menge, oder konvexen Kegel) bestimmen, der S enthält. Dieser ist definiert als Durchschnitt aller linearer Räume (bzw, ...), die S enthalten. Diese Größen heißen dann **lineare Hülle** bzw. **affine Hülle** bzw. **konvexe Hülle** bzw **konvexe Kegelhülle** von S, werden mit $\operatorname{span} S$, $\operatorname{affine} S$, $\operatorname{conv} S$ bzw. $\operatorname{cone} S$ bezeichnet und können wie folgt charakterisiert werden:

Lemma 2.3 *Sei $S \subseteq \mathbb{R}^n$ eine beliebige Menge.*

$$\begin{aligned}
&(a) \quad \operatorname{span} S &&= \left\{ \sum_{i=1}^{m} \lambda_i \, x^{(i)} : x^{(i)} \in S \,,\ \lambda_i \in \mathbb{R} \,,\ m \in \mathbb{N} \right\} , \\
&(b) \quad \operatorname{affine} S &&= \left\{ \sum_{i=1}^{m} \lambda_i \, x^{(i)} : x^{(i)} \in S \,,\ \lambda_i \in \mathbb{R} \,,\ \sum_{i=1}^{m} \lambda_i = 1 \,,\ m \in \mathbb{N} \right\} , \\
&(c) \quad \operatorname{conv} S &&= \left\{ \sum_{i=1}^{m} \lambda_i \, x^{(i)} : x^{(i)} \in S \,,\ \lambda_i \geq 0 \,,\ \sum_{i=1}^{m} \lambda_i = 1 \,,\ m \in \mathbb{N} \right\} , \\
&(d) \quad \operatorname{cone} S &&= \left\{ \sum_{i=1}^{m} \lambda_i \, x^{(i)} : x^{(i)} \in S \,,\ \lambda_i \geq 0 \,,\ m \in \mathbb{N} \right\} .
\end{aligned}$$

Wir verzichten auf den einfachen Beweis. Als Beispiel betrachten wir den Fall, dass S aus endlich vielen Punkten $u^{(1)}, \ldots, u^{(p)} \in \mathbb{R}^n$ besteht. Dann ist z.B.

$$\operatorname{cone} \left\{ u^{(1)}, \ldots, u^{(p)} \right\} \;=\; \left\{ \sum_{i=1}^{p} \lambda_i \, u^{(i)} : \lambda_i \geq 0 \right\} \;=\; \left\{ U\lambda : \lambda \geq 0 \right\} ,$$

wobei wir die λ_i in den Spaltenvektor $\lambda = \left(\lambda_1, \ldots, \lambda_p\right)^\top \in \mathbb{R}^p$ und die Vektoren $u^{(i)}$ in die Matrix $U = \left[u^{(1)} \cdots u^{(p)} \right] \in \mathbb{R}^{n \times p}$ zusammengefasst haben. Solche Kegel heißen „endlich erzeugt“:

Definition 2.4 *(a) Ein konvexer Kegel $K \subseteq \mathbb{R}^n$ heißt* **endlich erzeugt**, *wenn es eine Matrix $U \in \mathbb{R}^{n \times m}$ gibt mit*

$$K \;=\; \left\{ U\lambda : \lambda \geq 0 \right\} \;=\; \left\{ \sum_{i=1}^{m} \lambda_i \, U_{*i} : \lambda_i \geq 0 \text{ für alle } i = 1, \ldots, m \right\} ,$$

wobei wir hier mit $U_{*i} \in \mathbb{R}^n$, den $i-$ten Spaltenvektor von U bezeichnet haben. Wir halten uns damit an die Schreibweise, wie sie z.B. in der Programmiersprache Matlab benutzt wird.

(b) Eine Menge $E \subseteq \mathbb{R}^n$ heißt **(Hyper-)Ebene** *im $\mathbb{R}^n$, wenn E die Form*

$$E \;=\; \left\{ x \in \mathbb{R}^n : a^\top x = \gamma \right\}$$

hat mit einem Vektor $a \in \mathbb{R}^n$, $a \neq 0$, und $\gamma \in \mathbb{R}$.

(c) Eine Menge $H \subseteq \mathbb{R}^n$ heißt **abgeschlossener Halbraum**, *wenn H die Form*

$$H \;=\; \left\{ x \in \mathbb{R}^n : a^\top x \leq \gamma \right\}$$

hat mit einem Vektor $a \in \mathbb{R}^n$, $a \neq 0$, und $\gamma \in \mathbb{R}$. Analog wird der offene Halbraum erklärt.

(d) *Eine Menge $M \subseteq \mathbb{R}^n$ heißt* **polyedral** *(oder auch* **Polyeder***), wenn sie als Durchschnitt von endlich vielen abgeschlossenen Halbräumen darstellbar ist, d.h. wenn es eine Matrix $A \in \mathbb{R}^{m\times n}$ und einen Vektor $b \in \mathbb{R}^m$ gibt mit $M = \{x \in \mathbb{R}^n : Ax \leq b\}$.*

(e) *Beschränkte Polyeder $M \subseteq \mathbb{R}^n$ heißen* **Polytope**.

Wir erkennen sofort: Ebenen und Polyeder sind abgeschlossen und konvex. Ebenen sind sogar affine Räume, Polytope kompakte konvexe Mengen. Unsere Restriktionenmengen M der Optimierungsprobleme sind also Polyeder, aber nicht unbedingt Polytope.

2.2 Der Satz von Weyl und das Farkas Lemma

Zunächst wollen ein berühmtes Ergebnis von H. Weyl zeigen, welches besagt, dass jeder endlich erzeugte konvexe Kegel polyedral ist. Bevor wir den Satz beweisen können, benötigen wir einen Hilfssatz:

Hilfssatz 2.5 *Für gegebene $a^{(j)} \in \mathbb{R}^n \setminus \{0\}$, $j = 1, \ldots, m$, $(m \geq 2)$ definiere die Mengen*

$$K_i \;:=\; \operatorname{cone}\{a^{(1)}, \ldots, a^{(m)}, -a^{(i)}\} \;=\; \left\{ \sum_{j=1}^{m} \lambda_j\, a^{(j)} : \lambda_j \geq 0 \text{ für } j \neq i \right\},$$

$i = 1, \ldots, m$. Dann gilt eine der beiden folgenden Aussagen:

(a) $\operatorname{cone}\{a^{(1)}\} = \cdots = \operatorname{cone}\{a^{(m)}\}$, *d.h. alle $a^{(j)}$ sind positive Vielfache von $a^{(1)}$, oder*

(b) $\operatorname{cone}\{a^{(1)}, \ldots, a^{(m)}\} = \bigcap_{i=1}^{m} K_i$.

Beweis: Wir beweisen die Aussage durch vollständige Induktion nach m. Dazu indizieren wir die Menge K_i zusätzlich durch m, d.h. wir schreiben

$$K_i^{(m)} \;:=\; \left\{ \sum_{j=1}^{m} \lambda_j\, a^{(j)} : \lambda_j \geq 0 \text{ für } j \neq i \right\}, \quad i = 1, \ldots, m\,.$$

Sei zunächst $m = 2$ und $a^{(1)}$ kein positives Vielfaches von $a^{(2)}$. Da die Richtung „$\subseteq$“ immer gilt, brauchen wir nur die Richtung $K_1^{(2)} \cap K_2^{(2)} \subseteq \operatorname{cone}\{a^{(1)}, a^{(2)}\}$ zu zeigen. Dies beweisen wir indirekt. Annahme, es gibt $x \in K_1^{(2)} \cap K_2^{(2)}$ mit $x \notin \operatorname{cone}\{a^{(1)}, a^{(2)}\}$. Dann gibt es Zahlen λ_{ij}, $i = 1, 2$, $j = 1, 2$, mit

$$x \;=\; \lambda_{11}\, a^{(1)} + \lambda_{12}\, a^{(2)} \;=\; \lambda_{21}\, a^{(1)} + \lambda_{22}\, a^{(2)}\,,$$

wobei $\lambda_{12} \geq 0$ und $\lambda_{21} \geq 0$. Es ist $\lambda_{11} < 0$ und $\lambda_{22} < 0$, da ja sonst $x \in \operatorname{cone}\{a^{(1)}, a^{(2)}\}$ gelten würde. Umsortierung dieser Gleichung liefert

$$\underbrace{(\lambda_{21} - \lambda_{11})}_{>0}\, a^{(1)} \;=\; \underbrace{(\lambda_{12} - \lambda_{22})}_{>0}\, a^{(2)}\,,$$

und dies ist ein Widerspruch zur Annahme, dass $a^{(2)}$ kein positives Vielfache von $a^{(1)}$ ist. Damit ist der Fall $m = 2$ bewiesen.

Sei die Aussage des Satzes jetzt für alle Mengen von $m-1$ Vektoren richtig. Es seien m Vektoren $\{a^{(1)},\dots,a^{(m)}\}$ gegeben, die nicht alle durch positive Vielfache von $a^{(1)}$ hervorgehen. Durch Umnummerierung können wir annehmen, dass schon $a^{(1)},\dots,a^{(m-1)}$ nicht alle positive Vielfache von $a^{(1)}$ sind. Dann gilt nach Induktionsvoraussetzung: $\operatorname{cone}\{a^{(1)},\dots,a^{(m-1)}\} = \bigcap_{i=1}^{m-1} K_i^{(m-1)}$. Wir zeigen die Gleichheit $\operatorname{cone}\{a^{(1)},\dots,a^{(m)}\} = \bigcap_{i=1}^{m} K_i^{(m)}$. Die Richtung „$\subseteq$“ gilt trivialerweise. Um die andere Richtung zu zeigen, nehmen wir an, es gebe $x \in \bigcap_{i=1}^{m} K_i^{(m)}$ mit $x \notin \operatorname{cone}\{a^{(1)},\dots,a^{(m)}\}$. Dann gibt es Zahlen λ_{ij}, $i,j=1,\dots,m$, mit

$$x \;=\; \sum_{j=1}^{m} \lambda_{ij}\, a^{(j)} \quad \text{und} \quad \lambda_{ij} \geq 0 \text{ für alle } j \neq i\,. \tag{2.2}$$

Ferner ist $\lambda_{ii} < 0$ für jedes i, denn sonst wäre ja $x \in \operatorname{cone}\{a^{(1)},\dots,a^{(m)}\}$. Sei $i \in \{1,\dots,m-1\}$ festgehalten. Wir multiplizieren (2.2) für $i=m$, nämlich $x = \sum_{j=1}^{m} \lambda_{mj}\, a^{(j)}$, mit λ_{im} und (2.2) mit λ_{mm}, subtrahieren die beiden Gleichungen voneinander und dividieren durch $\lambda_{im} - \lambda_{mm} > 0$. Dann erhalten wir

$$x \;=\; \sum_{j=1}^{m} \frac{\lambda_{mj}\lambda_{im} - \lambda_{ij}\lambda_{mm}}{\lambda_{im} - \lambda_{mm}}\, a^{(j)} \;=\; \sum_{j=1}^{m-1} \mu_{ij}\, a^{(j)} \quad \text{mit} \quad \mu_{ij} \;=\; \frac{\lambda_{mj}\lambda_{im} - \lambda_{ij}\lambda_{mm}}{\lambda_{im} - \lambda_{mm}}\,.$$

Dann ist $\mu_{ij} \geq 0$ für $i,j=1,\dots,m-1$, $i \neq j$. Daher ist

$$x \in \bigcap_{i=1}^{m-1} K_i^{(m-1)} \;\setminus\; \operatorname{cone}\{a^{(1)},\dots,a^{(m-1)}\}\,,$$

ein Widerspruch zur Induktionsvoraussetzung. □

Wir können jetzt den Satz von Weyl zeigen:

Satz 2.6 *(Weyl)*
Jeder endlich erzeugte konvexe Kegel ist polyedral, sogar von der Form $\{x \in \mathbb{R}^n : Ax \leq 0\}$. Damit sind endlich erzeugte konvexe Kegel insbesondere auch abgeschlossen.

Beweis: Sei $m \geq 1$ und $K = \operatorname{cone}\{a^{(1)},\dots,a^{(m)}\}$ ein endlich erzeugter Kegel. Wir zeigen die Behauptung wiederum durch Induktion nach m.
Sei $m=1$, d.h. $K = \operatorname{cone}\{a\}$ mit $a \neq 0$. Wir können a zu einer Orthogonalbasis $\{a, b^{(2)},\dots,b^{(n)}\}$ des $\mathbb{R}^n$ ergänzen. Insbesondere ist dann $\operatorname{span}\{a\} = \{x \in \mathbb{R}^n : x^\top b^{(j)} = 0 \text{ für } j=2,\dots,n\}$ und daher

$$\operatorname{cone}\{a\} \;=\; \{\lambda a : \lambda \geq 0\} \;=\; \{x \in \mathbb{R}^n : x^\top a \geq 0\,,\; x^\top b^{(j)} = 0 \text{ für } j=2,\dots,n\}\,.$$

Durch Umschreiben der Gleichungen in zwei Ungleichungen erkennen wir, dass $\operatorname{cone}\{a\}$ polyedral ist.

Sei die Behauptung jetzt richtig für alle Mengen von $m-1$ Vektoren.
Sei $K = \operatorname{cone}\{a^{(1)},\dots,a^{(m)}\}$, und wir nehmen ohne Einschränkung an, dass nicht alle $a^{(j)}$

positive Vielfache von $a^{(1)}$ sind (sonst folgt Behauptung nach dem soeben Bewiesenen). Nach Lemma 2.5 gilt

$$\operatorname{cone}\left\{a^{(1)},\ldots,a^{(m)}\right\} \;=\; \bigcap_{i=1}^{m} \operatorname{cone}\left\{a^{(1)},\ldots,a^{(m)},-a^{(i)}\right\}.$$

Da der Durchschnitt polyedraler Kegel offenbar wieder polyedral ist, so reicht es zu zeigen, dass $\operatorname{cone}\left\{a^{(1)},\ldots,a^{(m)},-a^{(i)}\right\}$ polyedral ist. Halte also $i \in \{1,\ldots,m\}$ fest und setze zur Abkürzung $a = a^{(i)}$. Der Projektor

$$P \;=\; I \;-\; \frac{1}{\|a\|_2^2}\, a\, a^\top$$

projiziert x orthogonal auf das orthogonale Komplement von a (Skizze!), denn $Px - x \in \operatorname{span}\{a\}$ und $a^\top P x = 0$ für alle $x \in \mathbb{R}^n$. Wir setzen ferner $\overline{a}^{(j)} = Pa^{(j)}$, $j = 1,\ldots,m$. Dann ist $\overline{a}^{(i)} = 0$ und (zeigen Sie dies!)

$$\operatorname{cone}\left\{a^{(1)},\ldots,a^{(m)},-a^{(i)}\right\} \;=\; \left\{x \in \mathbb{R}^n : Px \in \operatorname{cone}\left\{\overline{a}^{(1)},\ldots,\overline{a}^{(i-1)},\overline{a}^{(i+1)},\ldots,\overline{a}^{(m)}\right\}\right\}.$$

Nach Induktionsvoraussetzung ist $\operatorname{cone}\left\{\overline{a}^{(1)},\ldots,\overline{a}^{(i-1)},\overline{a}^{(i+1)},\ldots,\overline{a}^{(m)}\right\}$ polyedral, also

$$\operatorname{cone}\left\{\overline{a}^{(1)},\ldots,\overline{a}^{(i-1)},\overline{a}^{(i+1)},\ldots,\overline{a}^{(m)}\right\} \;=\; \left\{y \in \mathbb{R}^n : Vy \le 0\right\},$$

also

$$\operatorname{cone}\left\{a^{(1)},\ldots,a^{(m)},-a^{(i)}\right\} \;=\; \left\{x \in \mathbb{R}^n : VPx \le 0\right\}.$$

□

Des weiteren benötigen wir den strikten Trennungssatz:

Satz 2.7 *Sei $K \subseteq \mathbb{R}^n$ eine konvexe, abgeschlossene Menge, $K \neq \emptyset$, und $x \notin K$. Dann existiert eine Hyperebene, die x und K trennt, d.h. es existiert $a \in \mathbb{R}^n$, $a \neq 0$, und $\gamma \in \mathbb{R}$ mit*

$$a^\top z \le \gamma < a^\top x \quad \textit{für alle } z \in K.$$

(Skizze!) Ist K sogar ein konvexer Kegel, so kann $\gamma = 0$ gewählt werden.

Beweis: Der Beweis benutzt den **Projektionssatz** in der folgenden Form (siehe Übung):

Sei K eine konvexe abgeschlossene Menge. Zu jedem $x \notin K$ existiert (genau ein) $\hat{x} \in K$ mit $\|\hat{x} - x\|_2 \le \|z - x\|_2$ für alle $z \in K$. Dieser Projektionspunkt $\hat{x}$ ist charakterisiert durch (i) $\hat{x} \in K$ und (ii) $\left(x - \hat{x}\right)^\top \left(z - \hat{x}\right) \le 0$ für alle $z \in K$.

Wir setzen jetzt $a = x - \hat{x} \neq 0$ und $\gamma = a^\top \hat{x}$. Dann ist wegen (ii) $a^\top z \le a^\top \hat{x} = \gamma$ und wegen $a^\top \left(x - \hat{x}\right) = |a|^2 > 0$ auch $\gamma = a^\top \hat{x} < a^\top x$. Sei jetzt K zusätzlich ein Kegel. Dann können wir $z = 0$ wählen und erhalten $\gamma \ge 0$, also $a^\top x > 0$. Angenommen, es gebe $\hat{z} \in K$ mit $a^\top \hat{z} > 0$. Für $z = t\hat{z}$, $t > 0$, wäre dann $z \in K$, also $\gamma \ge a^\top z = t\, a^\top \hat{z} \to \infty$ für $t \to \infty$, ein Widerspruch. Also ist $a^\top z \le 0$ für alle $z \in K$. □

Schließlich beweisen wir jetzt noch das wichtige Lemma von Farkas:

Lemma 2.8 *(Farkas)*
Seien $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$ gegeben. Dann gilt genau eine der beiden folgenden Aussagen:

(i) $Ax = b$, $x \geq 0$, *ist lösbar durch ein* $x \in \mathbb{R}^n$.

(ii) $A^\top y \leq 0$, $b^\top y > 0$, *ist lösbar durch ein* $y \in \mathbb{R}^m$.

Beweis: (a) Angenommen, beide Aussagen würden gelten. Dann:

$$0 \;<\; y^\top b \;=\; y^\top A x \;=\; (A^\top y)^\top x \;\leq\; 0\,,$$

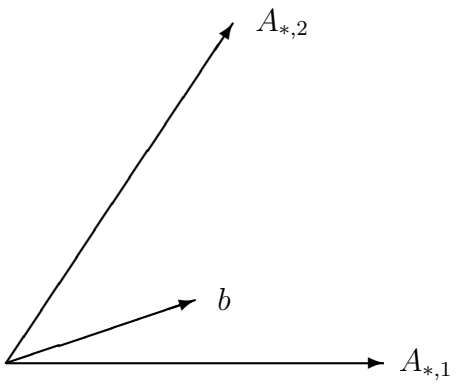
ein Widerspruch.

(b) Jetzt nehmen wir an, (i) gelte nicht. Dies bedeutet, dass

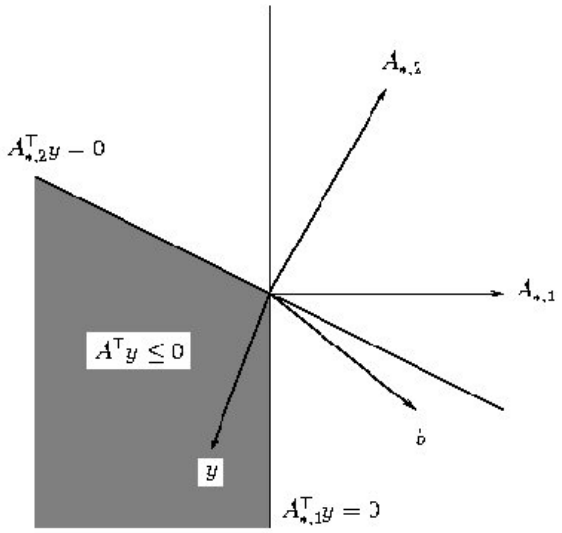
$$b \;\notin\; K \;:=\; \left\{Ax : x \in \mathbb{R}^n\,,\; x \geq 0\right\}.$$

Nach dem Satz von Weyl ist K ein konvexer polyedraler Kegel, insbesondere also auch abgeschlossen. Daher können wir b und K trennen. Es gibt also $y \in \mathbb{R}^m$, $y \neq 0$, mit $y^\top Ax \leq 0 < y^\top b$ für alle $x \geq 0$. Setzt man für x nacheinander die Koordinateneinheitsvektoren ein, so erhält man $A^\top y \geq 0$, und das Lemma ist bewiesen. □

Die folgenden Plots skizzieren die Aussage des Farkas Lemmas für die beiden Fälle: Sei $A = [A_{*,1}, A_{*,2}] \in \mathbb{R}^{2\times 2}$ und $b \in \mathbb{R}^2$.



In diesem Fall ist $Ax = b$, $x \geq 0$ lösbar.



In diesem Fall ist $Ax = b$ mit $x \geq 0$ nicht lösbar, jedoch $A^\top y \leq 0$, $b^\top y > 0$.

2.3 Der Hauptsatz der Polyedertheorie

Wir wollen jetzt eine Art Umkehrung des Satzes von Weyl beweisen, nämlich dass sich jedes geradenfreies Poyeder als konvexe Hülle seiner Ecken und konvexe Kegelhülle seiner extremalen Richtungen schreiben lässt. Diese Begriffe werden in der folgenden Definition erklärt:

Definition 2.9 $M \subseteq \mathbb{R}^n$ *sei beliebige konvexe Menge.*

(a) $x \in M$ *heißt* **Ecke** *oder* **Extremalpunkt** *von* M*, wenn sich* x *nicht als echte Konvexkombination zweier verschiedener Punkte von* M *darstellen lässt, d.h., wenn gilt:*

$$y, z \in M\,, \quad \lambda \in (0,1)\,, \quad x = \lambda y + (1-\lambda) z \quad \Longrightarrow \quad y = z\,.$$

(b) Ein Vektor $u \in \mathbb{R}^n$, $u \neq 0$, heißt **freie Richtung** *von M, wenn es $x \in M$ gibt, so dass der ganze Strahl $\{x + tu : t \geq 0\}$ zu M gehört.*

(c) Eine freie Richtung $u \in \mathbb{R}^n$, $u \neq 0$, heißt **extremale Richtung** *von M, wenn sie sich nicht als echte Konvexkombination zweier linear unabhängiger freier Richtungen schreiben lässt, d.h., wenn gilt:*

$$v, w \text{ freie Richtungen}, \lambda \in (0,1), \ u = \lambda v + (1-\lambda)w \quad \Longrightarrow \quad v, w \text{ linear abhängig}.$$

Strahlen der Form $s = \{x + tu : t \geq 0\} \subseteq M$ mit Ecke $x \in M$ und extremaler Richtung u heißen **Extremalstrahlen**.

Mit $\mathrm{extrP}(M)$ *bzw.* $\mathrm{extrS}(M)$ *bezeichnen wir die Menge aller Extremalpunkte bzw. Vereinigung aller Extremalstrahlen von M.*

Der folgende Satz charakterisiert Ecken und extremale Richtungen für Polyeder. Dazu benötigen wir die Menge der aktiven Indizes.

Satz 2.10 *Sei $M = \{x \in \mathbb{R}^n : Ax \leq b\}$ und für $x \in M$ sei $I(x) := \{i \in \{1, \ldots, m\} : (Ax)_i = b_i\}$ die Menge der* **aktiven Indizes**. *Dann gilt:*

(a) $x \in M$ ist genau dann Ecke von M, wenn $\operatorname{span}\{A_{i*}^\top : i \in I(x)\} = \mathbb{R}^n$, wobei $A_{i*}^\top \in \mathbb{R}^n$ die $i-$te Zeile (als Spaltenvektor aufgefasst) von A ist.

(b) M besitzt höchstens endlich viele Ecken.

(c) $u \in \mathbb{R}^n$, $u \neq 0$, ist genau dann freie Richtung für M, wenn $Au \leq 0$.

(d) $u \in \mathbb{R}^n$, $u \neq 0$, ist genau dann extremale Richtung für M, wenn $Au \leq 0$ und es kein von u linear unabhängiges $v \in \mathbb{R}^n$, $v \neq 0$, gibt mit $Av \leq 0$ und $(Av)_i = 0$ für alle i mit $(Au)_i = 0$.

Beweis: (a) Setze $V = \operatorname{span}\{A_{i*}^\top : i \in I(x)\}$.

Sei zunächst $V \subsetneq \mathbb{R}^n$. Dann gibt es $z \in \mathbb{R}^n$, $z \neq 0$, mit $z \perp V$, d.h. $(Az)_i = 0$ für alle $i \in I(x)$. Wir schreiben $x = \frac{1}{2}(x + \varepsilon z) + \frac{1}{2}(x - \varepsilon z)$. Für $i \in I(x)$ ist $(Ax)_i \pm \varepsilon(Az)_i = b_i$. Für $i \notin I(x)$ ist $(Ax)_i < b_i$, für hinreichend kleine $\varepsilon > 0$ ist also $(Ax)_i \pm \varepsilon(Az)_i \leq b_i$. Damit kann x keine Ecke sein.

Sei jetzt $V = \mathbb{R}^n$ und $x = \lambda y + (1-\lambda)z$ mit $Ay \leq b$, $Az \leq b$, $\lambda \in (0,1)$. Dann gilt für $i \in I(x)$ die Ungleichungskette $b_i = (Ax)_i = \lambda(Ay)_i + (1-\lambda)(Az)_i \leq b_i$, also gilt Gleicheit und daher $(Ay)_i = (Az)_i = b_i$ für alle $i \in I(x)$. Daher steht $y - z$ senkrecht auf V. Wegen $V = \mathbb{R}^n$ ist $y - z = 0$. Daher ist x Ecke.

(b) Wenn x eine Ecke ist, so folgt nach Teil (a), dass es eine Indexmenge $I \subseteq I(x)$ gibt mit n Elementen, so dass die Matrix $A_{I*} = (A_{ij})_{\substack{i \in I \\ j=1,\ldots,n}}$ vollen Rang hat und damit invertierbar ist. Aus $A_{I*}x = b_I$ folgt dann $x = A_{I*}^{-1} b_I$. Hier ist b_I natürlich der Vektor mit Komponenten in I. Daher liegen alle Ecken in der endlichen Menge $\{x = A_{I*}^{-1} b_I : I \subseteq \{1, \ldots, m\},\ A_{I*} \text{ invertierbar}\}$.

(c) Dieser Teil ist sehr einfach und verbleibt als Übung.

(d) Sei u extremal, $v \neq 0$ mit $Av \leq 0$ und $(Av)_i = 0$ für alle $i \in J$, wobei $J := \{i : (Au)_i = 0\}$. Für $i \in J$ ist dann $(Au)_i \pm \varepsilon(Av)_i = 0$. Für $i \notin J$ ist dagegen $(Au)_i < 0$, für

hinreichend kleine $\varepsilon > 0$ ist also $(Au)_i \pm \varepsilon(Av)_i \leq 0$. Daher sind $u \pm \varepsilon v$ freie Richtungen und $u = \frac{1}{2}(u+\varepsilon v)+\frac{1}{2}(u-\varepsilon v)$. Da u extremal ist, müssen $u+\varepsilon v$ und $u-\varepsilon v$ linear abhängig sein und damit auch u und v.

Es gebe jetzt umgekehrt kein von u linear unabhängiges v mit $(Av) \leq 0$ und $(Av)_i = 0$ für $i \in J$. Sei $u = \lambda v + (1-\lambda)w$ mit $\lambda \in (0,1)$ und freien Richtungen v und w. Für $i \in J$ ist $0 = (Au)_i = \lambda(Av)_i + (1-\lambda)(Aw)_i \leq 0$, also $(Av)_i = (Aw)_i = 0$ für $i \in J$. Nach Voraussetzung sind daher sowohl v als auch w Vielfache von u. Daher sind auch v und w linear abhängig. □

Man kann sich an einer Skizze schnell überlegen, dass Polyeder überhaupt keine Ecken oder Extremalstrahlen besitzen müssen.

Sei $M = \{x \in \mathbb{R}^n : Ax \leq b\}$ ein nichtleeres Polyeder. Der Rand ∂M von M ist jedenfalls enthalten in der Menge $\big\{x \in M : (Ax)_i = b_i \text{ für wenigstens ein } i\big\}$. Wir setzen

$$M_i \;=\; \big\{x \in M : (Ax)_i = b_i\big\}\,, \quad i = 1, \ldots, m\,.$$

Dann ist M_i wieder ein Polyeder, wir haben ja nur eine Ungleichung in $Ax \leq b$ durch eine Gleichung ersetzt, d.h. eine weitere Ungleichung hinzugefügt. Nichtleere Mengen der Form $\bigcap\big\{M_i : i \in I\big\}$ mit $I \subseteq \{1, \ldots, m\}$ heißen **Seiten** des Polyeders M. Natürlich sind die Seiten des Polyeders als Durchschnitt endlich vieler Polyeder wieder Polyeder. Für ein beliebiges Polyeder P definieren wir die **Dimension** von P als die Dimension von affine P. (Die Dimension eines affinen Raumes V ist natürlich die Dimension des zugehörigen Unterraumes $V - z$ für $z \in V$.) Damit haben wir für das Polyeder M:

$$\partial M \;=\; \bigcup_{k=0}^{n} \bigcup_{\substack{S \text{ Seite von } M \\ \dim S = k}} S\,.$$

Für den Beweis des Polyedersatzes benötigen wir einige vorbereitende Sätze, z.B. den folgenden einfachen Hilfssatz:

Hilfssatz 2.11 *Sei $M = \{x \in \mathbb{R}^n : Ax \leq b\}$ ein Polyeder und $S \subseteq M$ eine Seite von M. Dann gilt:*

(a) Jede Ecke von S ist auch Ecke von M.

(b) Jeder Extremalstrahl von S ist auch Extremalstrahl von M.

Beweis: (a) Sei x Ecke von S und $y, z \in M$, $\lambda \in (0,1)$ mit $x = \lambda y + (1-\lambda)z$. Die Seite S habe die Darstellung $S = \bigcap\big\{M_i : i \in I\big\}$. Dann ist $b_i = (Ax)_i = \lambda(Ay)_i + (1-\lambda)(Az)_i \leq b_i$ für $i \in I$ und daher Gleichheit $(Ay)_i = (Az)_i = b_i$ für $i \in I$. Dies bedeutet $y, z \in S$. Da x eine Ecke von S ist, so folgt $y = z$.

(b) Sei $\{x + tu : t \geq 0\}$ ein Extremalstrahl von S, d.h. x ist Ecke von S und u ist extremale Richtung von S. Nach (a) ist x auch Ecke von von M. Zu zeigen bleibt, dass u auch extremale Richtung von M ist. Der Beweis dieser Aussage geht fast genauso wie der zu (a) und verbleibt als Übung. □

Wir benötigen die Begriffe des relativen Inneren $\operatorname{int}_{rel} M$ und des relativen Randes $\partial_{rel} M$ einer Menge M. Sei dafür wieder $U(x,\varepsilon) = \big\{y \in \mathbb{R}^n : \|x-y\|_2 < \varepsilon\big\}$ die Kugel um x mit

Radius $\varepsilon > 0$. Wir definieren

$$x \in \operatorname{int}_{rel} M \iff \text{es gibt } \varepsilon > 0 \text{ mit } U(x,\varepsilon) \cap \operatorname{affine} M \subseteq M\,,$$
$$x \in \partial_{rel} M \iff \text{für jedes } \varepsilon > 0 \text{ gilt } U(x,\varepsilon) \cap M \neq \emptyset \text{ und } U(x,\varepsilon) \cap (\operatorname{affine} M \setminus M) \neq \emptyset\,.$$

Dann gilt:

Hilfssatz 2.12 *Sei $M = \{x \in \mathbb{R}^n : Ax \leq b\}$ ein nichtleeres Polyeder mit $k = \dim M$. Dann gilt*

$$\partial_{rel} M \subseteq \bigcup_{j=0}^{k-1} \bigcup_{\substack{S \text{ Seite von } M \\ \dim S = j}} S\,. \tag{2.3}$$

Beweis: Sei $\hat{x} \in \partial_{rel} M$ mit zugehöriger Menge $I(\hat{x}) = \{i : (A\hat{x})_i = b_i\}$ von aktiven Indizes. Setze

$$S = \{x \in M : (Ax)_i = b_i \text{ für alle } i \in I(\hat{x})\} = \bigcap_{i \in I(\hat{x})} M_i\,.$$

Dann ist S eine Seite von M und $\dim M \leq k$. Es ist sogar $\hat{x} \in \operatorname{int}_{rel} S$. Das sieht man folgendermaßen. Zunächst ist $\operatorname{affine} S \subseteq \{x \in \mathbb{R}^n : (Ax)_i = b_i \text{ für alle } i \in I(\hat{x})\}$ wegen $S \subseteq \{x \in \mathbb{R}^n : (Ax)_i = b_i \text{ für alle } i \in I(\hat{x})\}$. Wähle nun $\varepsilon > 0$ so, dass $(Ax)_i \leq b_i$ für alle $i \notin I(\hat{x})$ und alle $x \in \mathbb{R}^n$ mit $\|x - \hat{x}\|_2 \leq \varepsilon$. Dies ist möglich wegen $(A\hat{x})_i < b_i$ für alle $i \notin I(\hat{x})$. Für $x \in U(\hat{x},\varepsilon) \cap (\operatorname{affine} S)$ ist daher $(Ax)_i = b_i$ für alle $i \in I(\hat{x})$ und $(Ax)_i \leq b_i$ für alle $i \notin I(\hat{x})$, also $x \in S$, also $U(\hat{x},\varepsilon) \cap (\operatorname{affine} S) \subseteq S$.

Wäre $\dim S = k$, so $\operatorname{affine} S = \operatorname{affine} M$, also $U(\hat{x},\varepsilon) \cap (\operatorname{affine} M) \subseteq S \subseteq M$, d.h. $\hat{x} \in \operatorname{int}_{rel} M$. Dies widerspricht $\hat{x} \in \partial_{rel} M$. Also ist $\dim S \leq k-1$. □

Satz 2.13 *Sei $M \subseteq \mathbb{R}^n$ konvex und nicht leer. Dann ist $\operatorname{int}_{rel} M \neq \emptyset$, d.h. es gibt $x \in M$ und $\varepsilon > 0$ mit $U(x,\varepsilon) \cap \operatorname{affine} M \subseteq M$.*

Beweis: Wir wählen $z \in M$ beliebig und setzen $\tilde{M} = M - z$. Dann ist $0 \in \tilde{M}$ und $V = \operatorname{affine} \tilde{M} = \operatorname{span} \tilde{M}$. Wir zeigen $\operatorname{int}_{rel} \tilde{M} \neq \emptyset$. Setze dazu $m = \dim V$. Ist $m = 0$, so ist $M = \{z\} = \operatorname{affine} M$, und daher auch $\operatorname{int}_{rel} M = \{z\}$. Ist $m > 0$ so wähle linear unabhängige $b^{(1)}, \ldots, b^{(m)} \in \tilde{M}$. Also liegt das Simplex $S = \operatorname{conv}\{0, b^{(1)}, \ldots, b^{(m)}\}$ in $\tilde{M}$.

Der bijektive Homomorphismus (Koordinatenabbildung) $\psi : \mathbb{R}^m \to \operatorname{span} \tilde{M}$, definiert durch

$$\psi(\alpha_1, \ldots, \alpha_m) = \sum_{j=1}^{m} \alpha_j\, b^{(j)}\,,$$

bildet das Standardsimplex $\operatorname{conv}\{0, e^{(1)}, \ldots, e^{(m)}\}$ auf S ab.[1] Dieses Standardsimplex hat natürlich innere Punkte (z.B. den Schwerpunkt), also auch S (relativ zu V), also auch $\tilde{M}$ (relativ zu V). □

Satz 2.14 *Sei $M \subseteq \mathbb{R}^n$ konvex, nicht leer und geradenfrei (d.h. es gibt keine Gerade $g = \{z + tp : t \in \mathbb{R}\}$ mit $g \subseteq M$). Dann ist $M \subseteq \operatorname{conv} \partial_{rel} M$, also enthalten in der konvexen Hülle des relativen Randes.*

[1] Hier bezeichnen $e^{(j)}$ mit $e^{(j)}_k = \delta_{jk}$, $j,k = 1, \ldots, m$, die Koordinateneinheitsvektoren im $\mathbb{R}^m$.

Beweis: Es ist $\partial_{rel} M \neq \emptyset$, da sonst $M = \operatorname{affine} M$ und dies der Geradenfreiheit widerspräche. Ohne Beschränkung können wir $0 \in \partial M$ annehmen (nur Verschiebung des Koordinatensystems). Dann ist $\operatorname{affine} M = \operatorname{span} M$, und wir nehmen $\operatorname{span} M = \mathbb{R}^n$ an. (Sonst müssten wir uns im Folgenden auf den Unterraum $\operatorname{span} M$ beschränken und alle topologischen Begriffe wir Inneres oder Randpunkt relativ zu diesem Unterraum nehmen.)

Als erstes zeigen wir, dass ∂M nicht konvex ist. Annahme, ∂M ist konvex. Da das Innere von ∂M leer ist, so folgt nach Satz 2.13, dass $\operatorname{span} \partial M \subsetneqq \mathbb{R}^n$. Also existiert ein $z \in M$ mit $z \notin \operatorname{span} \partial M$ (denn sonst wäre $M \subseteq \operatorname{span} \partial M$, d.h. $\mathbb{R}^n = \operatorname{span} M \subseteq \operatorname{span} \partial M \subsetneqq \mathbb{R}^n$, ein Widerspruch). Insbesondere muss $z \in \operatorname{int} M$ sein. Sei jetzt $a \in \operatorname{span} \partial M$, $a \neq 0$, und $g = \{z + ta : t \in \mathbb{R}\}$ die Gerade durch z längs a. Da M geradenfrei ist, so $g \not\subseteq M$. Dies bedeutet, dass es $\hat{t} \in \mathbb{R}$ gibt mit $z + \hat{t}a \in \partial M \subseteq \operatorname{span} \partial M$. Hieraus folgt $z \in \operatorname{span} \partial M$, ein Widerspruch. Damit ist gezeigt, dass ∂M nicht konvex sein kann.

Also gibt es $u, v \in \partial M$ mit $[u, v] \cap \operatorname{int} M \neq \emptyset$ (denn sonst wäre $[u, v] \subseteq \overline{M} \setminus \operatorname{int} M = \partial M$ für alle $u, v \in \partial M$, was die Konvexität von ∂M bedeuten würde). Hier ist $[u, v] = \big\{\lambda u + (1 - \lambda)v : 0 \leq \lambda \leq 1\big\}$ die Strecke zwischen u und v. Sei $y \in [u, v] \cap \operatorname{int} M$, d.h. $y = \lambda u + (1 - \lambda)v$ für ein $\lambda \in (0, 1)$. Ferner sei $\varepsilon > 0$ so klein, dass die Kugel $U(y, \varepsilon) := \big\{x \in \mathbb{R}^n : \|x - y\|_2 < \varepsilon\big\}$ um y mit Radius ε in $\operatorname{int} M$ liegt. Sei jetzt $x \in \operatorname{int} M$ beliebig. Wir zeigen $x \in \operatorname{conv} \partial M$. (Dann wäre der Beweis beendet, denn $\partial \subseteq \operatorname{conv} \partial M$ gilt sowieso).

Wir zeigen dafür, dass die Strahlen

$$S^{\pm} \;:=\; \big\{x + t(u - v) \;:\; t \gtrless 0\big\}$$

den Rand ∂M schneiden müssen. Annahme $S^+ \subseteq M$. Wir zeigen, dass u eine Darstellung der Form $u = \rho w + (1 - \rho)z$ hat mit einem $\rho \in (0, 1)$, einem $z \in U(y, \varepsilon)$ und einem $w \in S^+$. Mit $w = x + t(u - v) \in S^+$ und daher $u = \rho\big[x + t(u - v)\big] + (1 - \rho)z$ muss z die Form haben

$$\begin{aligned} z &= \frac{1 - \rho\, t}{1 - \rho}\, u \;+\; \frac{\rho\, t}{1 - \rho}\, v \;-\; \frac{\rho}{1 - \rho}\, x\,, \quad \text{also} \\ z - y &= \left(\frac{1 - \rho\, t}{1 - \rho} - \lambda\right) u \;+\; \left(\frac{\rho\, t}{1 - \rho} - 1 + \lambda\right) v \;-\; \frac{\rho}{1 - \rho}\, x \\ &= \left(\frac{1 - \rho\, t}{1 - \rho} - \lambda\right)(u - v) \;+\; \frac{\rho}{1 - \rho}\,(v - x)\,. \end{aligned}$$

Wähle nun $\rho \in (0, 1)$ so klein, dass $\frac{\rho}{1-\rho}\,\|v - x\|_2 < \varepsilon$ und bestimme dann t aus der Gleichung $\frac{1-\rho\, t}{1-\rho} - \lambda = 0$, d.h. $t = \frac{1-\lambda(1-\rho)}{\rho} > 0$. Mit dieser Wahl von ρ und t ist $\|z - y\|_2 = \frac{\rho}{1-\rho}\,\|v - x\|_2 < \varepsilon$, also $z \in U(y, \varepsilon)$. Damit ist die Darstellung $u = \rho w + (1 - \rho)z$ mit einem $\rho \in (0, 1)$, einem $z \in U(y, \varepsilon)$ und einem $w \in S^+$ gezeigt. Es liegt also u echt zwischen w und z, damit $u \in \operatorname{int} M$, ein Widerspruch.

Daher ist $S^+ \cap \partial M \neq \emptyset$. Genauso zeigt man, dass auch $S^- \cap \partial M \neq \emptyset$. Mit $y^{\pm} \in S^{\pm} \cap \partial M$ folgt dann $x \in [y^-, y^+] \subseteq \operatorname{conv} \partial M$, und der Beweis ist endlich beendet. □

Damit können wir schließlich den Hauptsatz der Polyedertheorie beweisen:

Satz 2.15 *Sei $M \subseteq \mathbb{R}^n$ ein geradenfreies Polyeder. Dann ist*

$$M \;=\; \operatorname{conv}\big\{\operatorname{extrP}(M) \;\cup\; \operatorname{extrS}(M)\big\}\,. \tag{2.4}$$

In anderen Worten: M ist konvexe Hülle seiner Ecken und Extremalstrahlen.

Beweis: Wir führen einen Induktionsbeweis über $m = \dim M$.

$m = 1$: Dann ist affine M eine Gerade und M kann als konvexe Menge nur ein Intervall $[u, v]$ auf der Geraden oder ein Strahl sein. In beiden Fällen ist die Aussage des Satzes richtig.

Es gelte (2.4) jetzt für alle geradenfreien Polyeder M' mit $\dim M' \leq m-1$, und es sei M ein geradenfreies Polyeder der Form $M = \left\{x \in \mathbb{R}^n : Ax \leq b\right\}$ mit $\dim M = m$. Da die Inklusion „$\supseteq$“ von (2.4)) trivial ist, muss wegen Satz 2.14 und (2.3) nur gezeigt werden, dass jede Seite S von M mit $\dim S \leq m-1$ in $\operatorname{conv}\left\{\operatorname{extrP}(M) \cup \operatorname{extrS}(M)\right\}$ liegt. Da S ebenfalls ein geradenfreies Polyeder ist, folgt nach Induktionsvoraussetzung und Hilfssatz 2.11

$$S \;=\; \operatorname{conv}\left\{\operatorname{extrP}(S) \;\cup\; \operatorname{extrS}(S)\right\} \;\subseteq\; \operatorname{conv}\left\{\operatorname{extrP}(M) \;\cup\; \operatorname{extrS}(M)\right\}.$$

Damit ist der Satz bewiesen. □

Da ein Polytop als beschränkte Menge keine Extremalstrahlen besitzt, folgt sofort:

Korollar 2.16 *Jedes (nichtleere) Polytop M ist die konvexe Hülle seiner Ecken.*

Die Voraussetzung der Geradenfreiheit ist q̈uivalent zur Existenz von Ecken:

Satz 2.17 *Sei $M \subseteq \mathbb{R}^n$ ein nichtleeres Polyeder. M besitzt genau dann Ecken, wenn M keine Gerade enthält, also geradenfrei ist.*

Beweis: Ist M geradenfrei, so liefert der Hauptsatz 2.15 insbesondere auch die Existenz von Ecken. Umgekehrt besitze jetzt M eine Ecke $\hat{x}$, und wir nehmen an, M enthalte auch eine Gerade $g = \{z + tu : t \in \mathbb{R}\}$. Dann gilt $A\hat{x} \leq b$ und $A(z + tu) \leq b$ für alle $t \in \mathbb{R}$. Aus der letzten Ungleichung folgt $Au = 0$ (weshalb?) und daher $A(\hat{x} + tu) \leq b$. Daher liegt die gesamte Gerade durch $\hat{x}$ mit Richtung u in M, ein Widerspruch dazu, dass $\hat{x}$ Ecke ist. □

In diesem Kapitel haben wir uns an Skripten und das Lehrbuch von J. Werner (Optimization Theory and Applications) gehalten sowie in Abschnitt 2.3 an die Vorlesungsmitschrift von meinem Kollegen W. Weil aus dem WS 2000/2001, dem ich dafür danke.

3 Existenz- und Dualitätstheorie für lineare Optimierungsaufgaben

3.1 Ein Existenzsatz

Am Anfang des ersten Kapitels haben wir gesehen, dass es gleichgültig ist, ob wir das Optimierungsproblem in der Form (P1) oder (P2) vorliegen haben – oder in einer allgemeineren Form mit Gleichungen und Ungleichungen und Vorzeichenbedingungen nur für einige Komponenten. Wir können es immer auf eine der beiden Standardformen transformieren. Wir gehen jetzt von der Normalform (P2) aus, d.h. betrachten das Problem:

$$\text{(P)} \qquad \text{Minimiere} \quad c^\top x \quad \text{auf } M \;=\; \{x \in \mathbb{R}^n : Ax = b,\; x \geq 0\}\,,$$

wobei $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ und $c \in \mathbb{R}^n$ gegeben sind. Zunächst bemerken wir, dass natürlich auch dieses M ein Polyeder ist. Es hat aber die zusätzliche Eigenschaft, **geradenfrei** zu sein. M ist ja enthalten im Polyeder $\{x \in \mathbb{R}^n : x \geq 0\}$, welches keine Gerade enthält. Nach Satz 2.17 besitzt M also Ecken. Außerdem ist der Hauptsatz der Polyedertheorie anwendbar.

Wir fragen jetzt nach der Existenz einer Lösung. Einfache geometrische Überlegungen (vielleicht besser für Probleme in der Form (P1)) zeigen, dass nicht immer eine Lösung existieren muss. Falls M beschränkt ist, so existiert jedenfalls eine Lösung, da dann M kompakt und die Zielfunktion stetig ist. Die Voraussetzung der Beschränktheit ist aber zu stark. Es gilt die folgende wichtige Existenzaussage:

Satz 3.1 ***(Existenzsatz)***
Es sei das Problem (P) gegeben, und es sei

$$\mu^* \;:=\; \inf\{c^\top x : x \in M\} \;>\; -\infty\,.$$

(a) Falls (P) überhaupt zulässig ist, d.h. falls $M \neq \emptyset$, so ist (P) auch lösbar, d.h. es existiert ein $x^* \in M$ mit $c^\top x^* = \mu^*$, d.h. $c^\top x^* \leq c^\top x$ für alle $x \in M$.
(b) Falls (P) lösbar ist, so existiert auch eine Ecke als Lösung.

Beweis: (a) Wir definieren die folgende Menge $\Lambda \subseteq \mathbb{R} \times \mathbb{R}^m$, die uns auch später noch bei der Herleitung des dualen Problems begegnen wird:

$$\Lambda \;:=\; \{(c^\top x, Ax - b) \in \mathbb{R} \times \mathbb{R}^m : x \in \mathbb{R}^n\,,\; x \geq 0_n\}\,.$$

Hier haben wir der Deutlichkeit halber 0_n für den Nullvektor im $\mathbb{R}^n$ geschrieben. Diese Menge Λ ist ein endlich erzeugter konvexer Kegel in $\mathbb{R}^{m+1}$ mit Spitze im Punkt $\left[\begin{smallmatrix} 0 \\ -b \end{smallmatrix}\right]$, denn wir können Λ schreiben in der Form

$$\Lambda \;=\; \begin{bmatrix} 0 \\ -b \end{bmatrix} + \Lambda_0 \quad \text{mit} \quad \Lambda_0 \;=\; \left\{ \begin{bmatrix} c^\top \\ A \end{bmatrix} x : x \in \mathbb{R}^n\,,\; x \geq 0_n \right\} \;\subseteq \mathbb{R}^{m+1}\,.$$

Als endlich erzeugter konvexer Kegel ist Λ_0 polyedral, also abgeschlossen (Satz von Weyl). Damit ist auch Λ abgeschlossen. Sei jetzt (x_k) eine Folge in M mit $c^\top x_k \to \mu^*$ für $k \to \infty$.

Bezeichnen wir mit 0_m den Nullvektor im $\mathbb{R}^m$, so ist $\big(c^\top x_k, 0_m\big) \in \Lambda$ und $\big(c^\top x_k, 0_m\big) \longrightarrow \big(\mu^*, 0_m\big)$ für $k \to \infty$. Wegen der Abgeschlossenheit von Λ ist auch $(\mu^*, 0_m) \in \Lambda$, d.h. es gibt $x^* \geq 0_n$ mit $\mu^* = c^\top x^*$ und $0_m = Ax^* - b$. Damit ist x^* optimal!

(b) Sei $x^* \in M$ optimal. Da M geradenfrei ist, so können wir den Hauptsatz 2.15 anwenden. Es gibt also $b^j \in \mathrm{extrP}(M) \cup \mathrm{extrS}(M)$ und $\lambda_j \geq 0$, $j = 1, \ldots, p$, mit $\sum_{j=1}^p \lambda_j = 1$ und $x^* = \sum_{j=1}^p \lambda_j b^{(j)}$. Ohne Beschränkung können wir $\lambda_j > 0$ für alle j annehmen. Daher ist

$$c^\top x^* \;=\; \sum_{j=1}^p \lambda_j \underbrace{c^\top b^{(j)}}_{\geq c^\top x^*} \;\geq\; \sum_{j=1}^p \lambda_j\, c^\top x^* \;=\; c^\top x^*\,.$$

Hieraus folgt (Widerspruchsargument) $c^\top b^{(j)} = c^\top x^*$ für alle $j = 1, \ldots, p$. Tritt unter den $b^{(j)}$ eine Ecke auf, sind wir fertig. Ist $b^{(j)}$ ein Element aus einem Extremalstrahl $s = \{z + tu : t \geq 0\}$ mit Ecke z und extremaler Richtung u, so folgt $c^\top u \geq 0$ wegen $c^\top x^* \leq c^\top(z + tu)$ für alle $t \geq 0$. Ist nun $b^{(j)} = z + \hat{t}u$ für ein $\hat{t} \geq 0$, so folgt

$$c^\top x^* \;=\; c^\top b^{(j)} \;=\; c^\top z + \hat{t}\, c^\top u \;\geq\; c^\top z \;\geq\; c^\top x^*$$

und daher Gleichheit $c^\top z = c^\top x^*$. Also ist die Ecke z optimal. □

Bemerkung: Natürlich gilt der Existenzteil (a) des Satzes auch für alle anderen Formen von linearen Optimierungsproblemen, da sie ja auf die Form (P) transformierbar sind. Insbesondere benötigt man für Maximierungsaufgaben, dass $\sup\{c^\top x : x \in M\} < +\infty$. Allerdings erfordert der Teil (b) wirklich die Existenz einer Ecke von M (der Hauptsatz muss anwendbar sein!). Man überlege sich (graphisch) ein Beispiel eines linearen Optimierungsproblems, das lösbar ist, ohne dass die zulässige Menge Ecken besitzt.

3.2 Das duale Problem

Es sei wieder das Problem (P) gegeben:

(P) $\qquad$ Minimiere $\quad c^\top x \quad$ auf $M = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}$.

Wir nennen (P) das **primale Problem** und wollen zunächst das duale Problem motivieren und herleiten. Dazu erinnern wir noch einmal an den endlich erzeugten konvexen Kegel

$$\Lambda \;=\; \big\{(c^\top x, Ax - b) \in \mathbb{R} \times \mathbb{R}^m : x \in \mathbb{R}^n\,,\ x \geq 0_n\big\} \tag{3.1}$$

und betrachten das folgende Problem (siehe Skizze):

(P') $\qquad$ Minimiere $\quad \rho \quad$ unter $(\rho, y) \in M' \;:=\; \Lambda \cap \big(\mathbb{R} \times \{0_m\}\big)$

und zeigen den Zusammenhang mit (P):

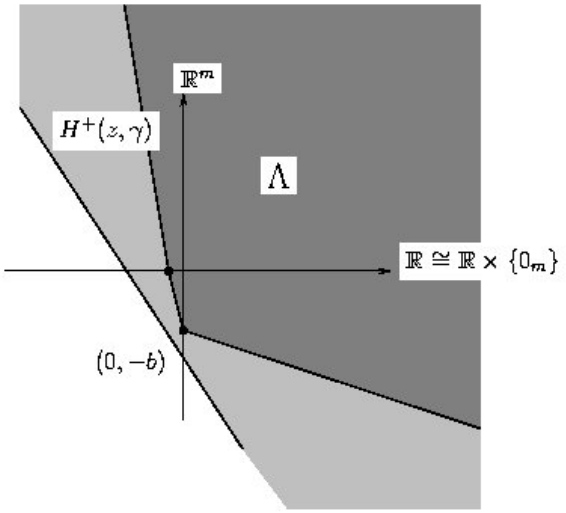
Lemma 3.2 *Es gilt:*

(a) Es ist $x \in M$ genau dann, wenn $(x, 0_m) \in M'$.

(b) $x^* \in M$ ist genau dann Lösung von (P) mit Wert $\mu^* = c^\top x^*$, wenn $(\mu^*, 0_m)$ Lösung ist von (P') mit Wert μ^*.

In diesem Sinn sind also (P) und (P') äquivalent.

Der **Beweis** verbleibt als Übung (siehe analogen Beweis von Lemma 3.3)!



In der Skizze besteht das Problem (P') also darin, den Punkt auf der $\mathbb{R}-$Achse zu finden, der am weitesten links in Λ liegt. Den selben Punkt können wir auch folgendermaßen finden: Wir betrachten alle Hyperebenen $H(z,\gamma)$ im $\mathbb{R}\times\mathbb{R}^m$ (in der Skizze also Geraden), die die $\mathbb{R}-$Achse in genau einem Punkt schneiden und die Λ im nicht-negativen Halbraum H^+ enthalten. Von diesen Hyperebenen suchen wir diejenige, die die $\mathbb{R}-$Achse am weitesten rechts schneidet. Dieses sei das verbal formulierte duale Problem.
Wir wollen dies nun übersetzen. Hyperebenen in $\mathbb{R} \times \mathbb{R}^m$ haben die Form

$$\left\{ (t,y) \in \mathbb{R} \times \mathbb{R}^m : \begin{bmatrix} t \\ y \end{bmatrix}^\top \begin{bmatrix} s \\ z \end{bmatrix} = \gamma \right\} \;=\; \left\{ (t,y) \in \mathbb{R} \times \mathbb{R}^m : t\,s \;+\; y^\top z = \gamma \right\}$$

für gegebenen Vektor $(s,z) \in \mathbb{R} \times \mathbb{R}^m$ und Zahl $\gamma \in \mathbb{R}$. Diese Hyperebene schneidet $\mathbb{R} \times \{0_m\}$ genau dann in genau einem Punkt, wenn $s \neq 0$ ist. Nach Division durch s können wir $s = 1$ annehmen, also haben die zugelassenen Hyperebenen die Form

$$H(z,\gamma) \;:=\; \left\{ (t,y) \in \mathbb{R} \times \mathbb{R}^m : t \;+\; y^\top z = \gamma \right\}.$$

Der Schnittpunkt mit $\mathbb{R} \times \{0_m\}$ ist gerade $(\gamma, 0_m)$, und der nicht-negative Halbraum hat die Form

$$H^+(z,\gamma) \;:=\; \left\{ (t,y) \in \mathbb{R} \times \mathbb{R}^m : t \;+\; y^\top z \geq \gamma \right\}.$$

Damit können wir das vorläufige duale Problem (D') formulieren als:

(D') Maximiere γ unter $\Lambda \subseteq H^+(z,\gamma)$ für ein $z \in \mathbb{R}^m$.

Analog zu Lemma 3.2 gilt:

Lemma 3.3 *Es gilt:*

(a) $\Lambda \subseteq H^+(z,\gamma)$ *ist äquivalent zu*

$$A^\top z + c \geq 0_n \quad \textit{und} \quad b^\top z + \gamma \leq 0\,. \tag{3.2}$$

(b) Ist $(\gamma, z) \in \mathbb{R}\times\mathbb{R}^m$ zulässig für (D'), so ist $y = -z$ zulässig für das folgende Problem, das wir (D) nennen:

$$\text{(D)} \qquad \text{Maximiere} \quad b^\top y \quad \text{auf } N \;=\; \left\{y \in \mathbb{R}^m : A^\top y \le c\right\}$$

Ist $(\gamma^*, z^*) \in \mathbb{R} \times \mathbb{R}^m$ Lösung von (D') mit Wert γ^*, so ist $y^* := -z^*$ Lösung von (D) mit Wert γ^*.

(c) Ist y zulässig für (D), so ist $(\gamma, -y)$ zulässig für (D') für alle $\gamma \le b^\top y$. Ist y^* optimal für (D), so ist $(b^\top y^*, -y^*)$ optimal für (D').

In diesem Sinn sind also (D) und (D') äquivalent.

Beweis: (a) $\Lambda \subseteq H^+(z,\gamma)$ ist äquivalent zu $c^\top x + (Ax - b)^\top z \ge \gamma$ für alle $x \ge 0_n$, d.h.

$$\left(c + A^\top z\right)^\top x \ge \gamma + b^\top z \text{ für alle } x \ge 0_n\,. \tag{3.3}$$

Wir bemerken zunächst, dass (3.2) die Bedingungen (3.3) implizieren. Es gelte umgekehrt (3.3). Für $x = 0_n$ folgt hieraus erst einmal $\gamma + b^\top z \le 0$. Es verbleibt $A^\top z + c \ge 0_n$ zu zeigen. Angenommen, dies gelte nicht. Dann gibt es Index $j_0 \in \{1,\ldots,n\}$ mit $\left(A^\top z\right)_{j_0} + c_{j_0} < 0$. Setze $x_j = 0$ für $j \ne j_0$ und $x_{j_0} = t$ für ein $t > 0$. Dann ist nach (3.3):

$$t\left(A^\top z + c\right)_{j_0} \;\ge\; \gamma + b^\top z\,.$$

Für $t \to \infty$ liefert dies jedoch einen Widerspruch.

(b) Sei $(\gamma, z) \in \mathbb{R}\times\mathbb{R}^m$ zulässig für (D'), d.h. $\Lambda \subseteq H^+(z,\gamma)$. Nach (a) ist $A^\top z + c \ge 0_n$ und $b^\top z + \gamma \le 0$. Für $y = -z$ gilt dann $A^\top y \le c$ und $b^\top y \ge \gamma$. Daher ist y zulässig für (D). Ist jetzt $(\gamma^*, z^*) \in \mathbb{R} \times \mathbb{R}^m$ optimal für (D'), so ist $y^* = -z^*$ zulässig für (D) und $b^\top y^* \ge \gamma^*$. Für beliebiges $y \in N$ ist $(b^\top y, -y)$ zulässig für (D'). Daher ist $b^\top y \le \gamma^* \le b^\top y^*$. Daher ist y^* optimal für (D).

(c) Ist y zulässig für (D), so ist offenbar $(\gamma, -y)$ zulässig für (D') für jedes $\gamma \le b^\top y$. Ist y^* optimal für (D), so ist $(b^\top y^*, -y^*)$ zulässig für (D'). Ist (γ, z) ebenfalls zulässig für (D'), so ist $-z$ zulässig für (D) und $b^\top z + \gamma \le 0$, also $\gamma \le -b^\top z \le b^\top y^*$. Dies bedeutet genau, dass $(b^\top y^*, -y^*)$ optimal für (D') ist. □

Dieses lineare Optimierungsproblem (D) liegt in der zweiten Normalform vor. Wir wollen zunächst das duale Problem zu (D) ermitteln: Daher schreiben wir wie gewohnt $y = y^+ - y^-$ mit $y^+, y^- \ge 0$, und führen Schlupfvariable $z = c - A^\top y \ge 0$ ein. Damit lautet (D):

$$\text{Minimiere} \quad \begin{bmatrix} -b \\ b \\ 0_n \end{bmatrix}^\top \begin{bmatrix} y^+ \\ y^- \\ z \end{bmatrix} \quad \text{unter den Nebenbedingungen}$$

$$\begin{bmatrix} y^+ \\ y^- \\ z \end{bmatrix} \;\ge\; 0_{2m+n} \quad \text{und} \quad \left[-A^\top \mid A^\top \mid -I_n\right] \begin{bmatrix} y^+ \\ y^- \\ z \end{bmatrix} \;=\; -c\,.$$

Das dazu duale Problem lautet:

Maximiere $(-c)^\top x$ unter den Nebenbedingungen

$$\begin{bmatrix} -A \\ A \\ -I_n \end{bmatrix} x \;\leq\; \begin{bmatrix} -b \\ b \\ 0_n \end{bmatrix}, \quad \text{also} \quad Ax = b \quad \text{und} \quad x \geq 0_n\,.$$

Daher ist das duale Problem zu (D) wieder (P).

3.3 Die Dualitätssätze

In diesem Abschnitt wollen wir Beziehungen zwischen (P) und (D) aufstellen. Wir formulieren noch einmal die beiden Probleme:

(P) $$\text{Minimiere} \quad c^\top x \quad \text{auf } M = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}\,,$$

(D) $$\text{Maximiere} \quad b^\top y \quad \text{auf } N = \left\{y \in \mathbb{R}^m : A^\top y \leq c\right\}$$

Wir beginnen mit einem ersten ganz einfachen, aber sehr wichtigen Satz.

Satz 3.4 ***(schwacher Dualitätssatz)***
Die Probleme (P) und (D) seien beide zulässig, d.h. $M \neq \emptyset$ und $N \neq \emptyset$. Dann gilt:
(a) Sind $x \in M$ und $y \in N$, so ist $b^\top y \leq c^\top x$. Dies bedeutet, dass die Zielfunktionswerte des dualen Problems immer unterhalb der des primalen Problems liegen.
(b) Beide Probleme (P) und (D) besitzen Lösungen.
(c) Gilt ferner für $x^* \in M$ und $y^* \in N$, dass $b^\top y^* = c^\top x^*$, so ist x^* eine Lösung von (P) und y^* eine Lösung von (D).

Beweis: (a) Sei $x \in M$, $y \in N$. Dann ist

$$b^\top y \;=\; (Ax)^\top y \;=\; \underbrace{x^\top}_{\geq 0}\underbrace{(A^\top y)}_{\leq c} \;\leq\; x^\top c\,.$$

(b) Dies folgt aus dem Existenzsatz 3.1, denn mit (a) ist ja $\inf(P) \geq b^\top y > -\infty$ für jedes $y \in N$. Genauso schliesst man für (D).
(c) Ist etwa $x \in M$ beliebig, so ist nach (a): $c^\top x \geq b^\top y^* = c^\top x^*$. Dies bedeutet, dass x^* optimal ist. Genauso schliesst man für (D). □

Folgerung: Mit diesem Satz können wir die primalen Kosten sofort nach unten abschätzen. Er liefert uns auch einen Test auf Optimalität.

Satz 3.5 ***(starker Dualitätssatz)***
Gegeben seien (P) und (D). Dann gilt:

(1) Sind (P) und (D) zulässig (d.h. $M \neq 0$ und $N \neq 0$), so besitzen (P) und (D) optimale Lösungen $x^* \in M$ und $y^* \in N$ und $b^\top y^* = c^\top x^*$, d.h. $\min(P) = \max(D)$.

(2) Ist (P) zulässig, (D) nicht zulässig, so ist $\inf(P) = -\infty$.

(3) Ist (D) zulässig, (P) nicht zulässig, so ist $\sup(D) = +\infty$.

Beweis: (1) Die Existenz von Lösungen von (P) und (D) folgt aus dem schwachen Dualitätssatz 3.4. Außerdem liefert er $\max(D) \leq \min(P)$. Wir nehmen $\max(D) < \min(P)$ an. Dann ist $\big(\max(D), 0_m\big) \notin \Lambda$, wobei wir wieder den Kegel Λ von (3.1) benutzen. Da dieser ja abgeschlossen ist, können wir wieder den Trennungssatz anwenden und den Punkt $\big(\max(D), 0_m\big) \in \mathbb{R} \times \mathbb{R}^m$ von Λ trennen. Es existiert also $(s, z) \in \mathbb{R} \times \mathbb{R}^m$ und $\gamma \in \mathbb{R}$ mit

$$s\, c^\top x \;+\; z^\top (Ax - b) \;\geq\; \gamma \;>\; s \max(D) \quad \text{für alle } x \geq 0_n\,.$$

Für $x \in M$ erhalten wir $Ax = b$, also $s\, c^\top x > s \max(D)$ und hieraus $s > 0$ mit dem schwachen Dualitätssatz. Nach Division durch s können wir ohne Einschränkung $s = 1$ annehmen. Wir haben also

$$\big(c + A^\top z\big)^\top x \;-\; z^\top b \;\geq\; \gamma \;>\; \max(D) \quad \text{für alle } x \geq 0_n\,.$$

Wie in der Herleitung von (D) erhalten wir hieraus $c + A^\top z \geq 0$ und $-z^\top b > \max(D)$. Daher ist $y := -z \in N$ und $y^\top b > \max(D)$, ein Widerspruch. Daher ist $\max(D) = \min(P)$.

(2) Sei jetzt (P) zulässig, also $M \neq \emptyset$, und (D) nicht zulässig, also $N = \emptyset$. Letzteres bedeutet, dass es kein $y \in \mathbb{R}^m$ gibt mit $A^\top y \leq c$. Dies impliziert, dass es keine Vektoren $y^+, y^- \in \mathbb{R}^m$, $z \in \mathbb{R}^n$ gibt mit $y^+ \geq 0_m$, $y^- \geq 0_m$, $z \geq 0_n$ und $A^\top(y^+ - y^-) + z = c$. Dies bedeutet, dass das System

$$\big[\, A^\top \,|\, -A^\top \,|\, I_n \,\big] \begin{bmatrix} y^+ \\ y^- \\ z \end{bmatrix} \;=\; c\,, \quad \begin{bmatrix} y^+ \\ y^- \\ z \end{bmatrix} \;\geq\; 0_{2m+n}\,,$$

keine Lösung besitzt. Das Farkas-Lemma 2.8 liefert die Existenz einer Lösung $z \in \mathbb{R}^n$ von

$$\begin{bmatrix} A \\ -A \\ I_n \end{bmatrix} z \;\leq\; 0_{2m+n}\,, \quad c^\top z \;>\; 0\,.$$

Dies bedeutet, dass $Az = 0_m$ und $z \leq 0_n$, sowie $c^\top z > 0$. Sei nun $\hat{x} \in M$ beliebig. Dann ist auch $x_t := \hat{x} - tz \in M$ für alle $t > 0$ und $c^\top x_t = c^\top \hat{x} - t\, c^\top z \longrightarrow -\infty$, $t \to +\infty$. Also ist $\inf(P) = -\infty$.

(3) Dies folgt genauso wie (2). Damit ist alles bewiesen. □

Folgerung: Sind $x^* \in M$ und $y^* \in N$ optimal, so gilt die **Komplementaritätsbedingung**:

$$x_j^* \;=\; 0 \quad \text{oder} \quad (A^\top y^*)_j \;=\; c_j \quad \text{für jedes } j = 1, \ldots, n\,.$$

Dies folgt aus der Gleichungskette

$$0 \;=\; c^\top x^* - b^\top y^* \;=\; c^\top x^* - x^{*\top} A^\top y^* \;=\; x^{*\top}[c - A^\top y^*]$$

und $x^* \geq 0$ sowie $c - A^\top y^* \geq 0$.

Beispiel 3.6 *(Diskretes Tschebyscheff-Problem)*
Es seien $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$, wobei wir $m > n$ annehmen. Wir wollen das überbestimmte Gleichungssystem $Ax \approx b$ lösen. Als Ersatzproblem betrachten wir:

$$\text{Minimiere} \quad \|Ax - b\|_\infty \quad \text{auf} \quad \mathbb{R}^n\,!$$

Dies schreiben wir in bekannter Weise als lineares Optimierungsproblem in der Form

$$\text{Minimiere} \quad x_0 \quad \text{unter} \quad \begin{aligned} Ax - b &\leq x_0\, e\,, \\ -Ax + b &\leq x_0\, e\,, \end{aligned}$$

wobei wieder $e = (1,\ldots,1)^\top \in \mathbb{R}^m$. Dies können wir äquivalent umformen zu:

$$(D) \qquad \text{Maximiere} \quad c^\top \tilde{x} \quad \text{unter} \quad \begin{bmatrix} -e & A \\ -e & -A \end{bmatrix} \tilde{x} \;\leq\; \begin{bmatrix} b \\ -b \end{bmatrix}\,,$$

wobei $\tilde{x} = (x_0, x_1, \ldots, x_n)^\top \in \mathbb{R}^{n+1}$ und $c = -(1,0,\ldots,0)^\top \in \mathbb{R}^{n+1}$. Dieses ist das duale Problem zu

$$(P) \qquad \text{Minimiere} \quad b^\top y - b^\top z \quad \text{unter} \quad y \geq 0\,,\; z \geq\; 0\,, \quad \begin{bmatrix} -e^\top & -e^\top \\ A^\top & -A^\top \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \;=\; c\,,$$

d.h.

$$\text{Minimiere} \quad b^\top(y-z) \quad \text{unter} \quad y, z \geq 0 \text{ und } e^\top(y+z) = 1 \text{ sowie } A^\top(y-z) = 0\,.$$

Offenbar sind (P) und (D) zulässig, also lösbar mit $\inf(P) = \sup(D)$. Wir machen die zusätzliche Voraussetzung, dass das überbestimmte Gleichungssystem $Ax = b$ keine Lösung besitzt. Dies bedeutet $x_0^* > 0$. Aus der Komplementaritätsbedingung folgt

$$\begin{aligned} y_i^* = 0 \quad &\text{oder} \quad (Ax^* - b)_i = x_0^* \quad \text{für jedes } i = 1,\ldots,m\,, \\ z_i^* = 0 \quad &\text{oder} \quad (Ax^* - b)_i = -x_0^* \quad \text{für jedes } i = 1,\ldots,m\,. \end{aligned}$$

Setzen wir $I^\pm := \big\{i \in \{1,\ldots,m\} : (Ax^* - b)_i = \pm x_0^*\big\}$, so ist $I^+ \cap I^- = \emptyset$ und

$$y_i^* \;=\; 0 \quad \text{für alle } i \notin I^+ \quad \text{und} \quad z_i^* \;=\; 0 \quad \text{für alle } i \notin I^-\,.$$

Definiere u_i für $i \in I := I^+ \cup I^-$ durch

$$u_i \;:=\; \begin{cases} y_i^*, & i \in I^+\,, \\ z_i^*, & i \in I^-\,. \end{cases}$$

Dann ist $u_i \geq 0$ und $\sum_{i\in I} u_i = 1$ und $y_i^* - z_i^* = \operatorname{sign}\big[(Ax^* - b)_i\big]\, u_i$ für alle $i \in I$. Die Bedingung $A^\top(y^* - z^*) = 0$ hat dann die Form

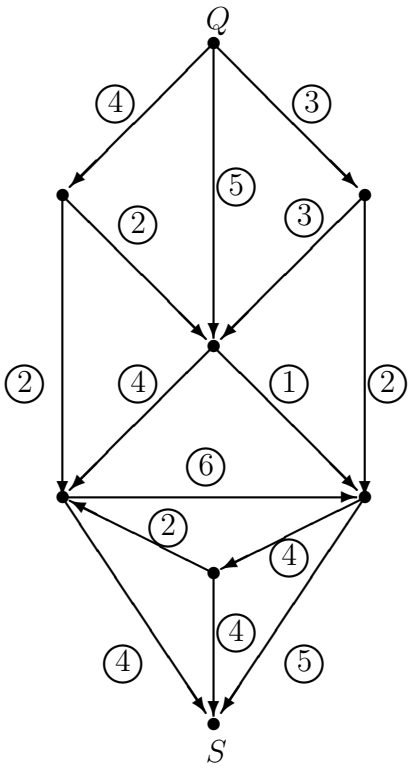
$$\sum_{i\in I} u_i \operatorname{sign}\big[(Ax^* - b)_i\big]\, a_{i*}^\top \;=\; 0\,, \quad \text{also} \quad 0 \in \operatorname{conv}\big\{\sigma_i a_{i*}^\top : i = 1,\ldots,m\big\}$$

mit $\sigma_i = \operatorname{sign}\big[(Ax^* - b)_i\big]$ für $i = 1,\ldots,m$. Aus diesen Bedingungen kann man dann einen Algorithmus basteln, der das Ausgleichsproblem löst. (Im kontinuierlichen Fall ist dies der berühmte Remez-Algorithmus der Approximationstheorie.) Allerdings liefert dies der Simplexalgorithmus in einfacherer Weise.

4 Anwendungen in der Graphen- und Spieltheorie

4.1 Das Max-Flow-Min-Cut-Theorem der Flussmaximierung

Wir betrachten jetzt **gerichtete** und **bewertete** Graphen. Wir stellen uns vor, die Kanten seien Öl-Pipelines, die an gewissen Punkten zusammenstoßen und sich verzweigen. Bei Q (wie Quelle) werde Öl gefördert, zur Ecke S (wie Senke) soll so viel wie möglich transportiert werden. Dabei müssen Kapazitätsbeschränkungen der Kanten berücksichtigt werden. Solch ein Graph könnte etwa so aussehen:



Dieser Graph wird durch die Kapazitätsmatrix $C \in \mathbb{R}^{n\times n}$ beschrieben. Die Ecken (oder Knoten) des Graphen entsprechen den Punkten $1, \ldots, n$, und c_{ij} gibt an, wieviel maximal von Knoten i nach Knoten j transportiert werden kann. Natürlich setzen wir $c_{ij} \geq 0$ für alle i, j voraus. $c_{ij} = 0$ bedeutet hierbei, dass es überhaupt keine Kante von i nach j gibt. Die Ecken seien so nummeriert, dass $i = 1$ die Quelle und $i = n$ die Senke ist. Wir setzen voraus, dass keine Kanten in die Quelle hinein und keine Kanten aus der Senke herausführen, d.h. $c_{i1} = c_{ni} = 0$ für alle $i = 1, \ldots, n$. Wir setzen auch voraus, dass 1 bzw. n die einzige Quelle bzw. Senke ist, d.h. $\sum_{i=1}^{n} c_{ij} > 0$ für alle $j = 2, \ldots, n$, und $\sum_{j=1}^{n} c_{ij} > 0$ für alle $i = 1, \ldots, n-1$. Außerdem sei $c_{ii} = 0$ für alle $i = 1, \ldots, n$. Dies bedeutet, dass die Knoten selbst keine Kapazitäten haben sollen. Schließlich setzen wir noch $c_{ij}c_{ji} = 0$ voraus für alle i, j, d.h. auf jeder Kante kann der Fluss höchstens in eine Richtung fließen. Eine Matrix $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ mit diesen Eigenschaften nennen wir eine **Kapazitätsmatrix**.

Jetzt können wir definieren:

Definition 4.1 *Sei $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ eine Kapazitätsmatrix. Ein* **Fluss** *ist eine Matrix $X = (x_{ij}) \in \mathbb{R}^{n\times n}$ mit $0 \le x_{ij} \le c_{ij}$ für alle $i, j = 1, \ldots, n$ und*

$$\sum_{i=1}^{n} x_{ij} \;=\; \sum_{k=1}^{n} x_{jk} \quad \text{für alle } j = 2, \ldots, n-1\,. \tag{4.1}$$

Die Bedingung (4.1) bedeutet, dass alles, was in den Knoten j hineinfließt, auch wieder hinausfließt. Zunächst zeigen wir, dass alles bei der Senke „ankommt“, was aus der Quelle hinausfließt:

Lemma 4.2 *Es gilt*

$$\sum_{j=1}^{n} x_{1j} \;=\; \sum_{k=1}^{n} x_{kn}\,.$$

Diese Größe heißt der **Wert** *des Flusses X und wird mit $W(X)$ bezeichnet.*

Beweis: Wir summieren (4.1) auf und beachten, dass $x_{nj} = x_{j1} = 0$ ist für alle j:

$$\sum_{j=2}^{n-1} \left[x_{1j} \;+\; \sum_{i=2}^{n-1} x_{ij} \right] \;=\; \sum_{j=2}^{n-1} \left[x_{jn} \;+\; \sum_{k=2}^{n-1} x_{jk} \right]\,.$$

Die Doppelsummen sind gleich und heben sich weg, also

$$\sum_{j=2}^{n-1} x_{1j} \;=\; \sum_{j=2}^{n-1} x_{jn}\,.$$

Addition von x_{1n} und $x_{11} = 0$ und $x_{nn} = 0$ auf beiden Seiten liefert die Behauptung. □

Damit können wir das Fluss-Optimierungsproblem formulieren:

$$\text{(P)} \qquad \text{Maximiere} \quad W(X) = \sum_{k=1}^{n} x_{kn} \quad \text{unter}$$

$$0 \le x_{ij} \le c_{ij} \text{ für alle } i, j = 1, \ldots, n \quad \text{und} \quad \sum_{i=1}^{n} x_{ij} \;=\; \sum_{k=1}^{n} x_{jk} \text{ für alle } j = 2, \ldots, n-1\,.$$

Jetzt kommen wir zur Definition eines Schnittes.

Definition 4.3 *Sei $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ eine Kapazitätsmatrix. Ein* **Schnitt** *(J^-, J^+) ist eine Zerlegung $J^+ \cup J^- = \{1, \ldots, n\}$ mit $J^+ \cap J^- = \emptyset$ und $1 \in J^-$ und $n \in J^+$. Die* **Kapazität** *eines Schnittes (J^-, J^+) sei definiert durch*

$$K(J^-, J^+) \;:=\; \sum_{\substack{i\in J^-\\ j\in J^+}} c_{ij}\,.$$

Dann können wir zeigen:

Satz 4.4 *Sei $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ eine Kapazitätsmatrix. Für jeden Schnitt (J^-, J^+) und jeden Fluss X gilt:*

$$W(X) \;=\; \sum_{\substack{i\in J^-\\ j\in J^+}} x_{ij} \;-\; \sum_{\substack{i\in J^-\\ j\in J^+}} x_{ji} \;\le\; K(J^-, J^+)\,. \tag{4.2}$$

Beweis: Da $x_{i1} = 0$ und $n \notin J^-$, so ist mit (4.1):

$$\sum_{j\in J^-}\sum_{i=1}^{n} x_{ij} \;=\; \sum_{j\in J^-\setminus\{1\}}\sum_{i=1}^{n} x_{ij} \;=\; \sum_{j\in J^-\setminus\{1\}}\sum_{k=1}^{n} x_{jk} \;=\; \sum_{j\in J^-}\sum_{k=1}^{n} x_{jk} \;-\; \underbrace{\sum_{k=1}^{n} x_{1k}}_{=\,W(X)}\,,$$

also

$$\begin{aligned} W(X) \;&=\; \sum_{j\in J^-}\sum_{k=1}^{n} x_{jk} \;-\; \sum_{j\in J^-}\sum_{i=1}^{n} x_{ij} \\ &=\; \sum_{j,k\in J^-} x_{jk} \;+\; \sum_{\substack{j\in J^-\\ k\in J^+}} x_{jk} \;-\; \sum_{j,i\in J^-} x_{ij} \;-\; \sum_{\substack{j\in J^-\\ i\in J^+}} x_{ij} \\ &=\; \sum_{\substack{j\in J^-\\ k\in J^+}} x_{jk} \;-\; \sum_{\substack{j\in J^-\\ i\in J^+}} x_{ij} \\ &\le\; \sum_{\substack{j\in J^-\\ k\in J^+}} c_{jk} \;=\; K(J^-, J^+)\,. \end{aligned}$$

□

Damit haben wir einen schwachen Dualitätssatz bewiesen: Der Maximierung des Flusses steht die Minimierung der Kapazität des Schnittes gegenüber. Unser Ziel ist es, den entsprechenden starken Dualitätssatz zu beweisen, d.h.

$$\max\big\{W(X) : X \text{ Fluss}\big\} \;=\; \min\big\{K(J^-, J^+) : (J^-, J^+) \text{ Schnitt}\big\}\,.$$

Dies geschieht durch Anwendung des starken Dualitätssatzes der linearen Optimierung. Um das Fluss-Maximierungsproblem als lineares Optimierungsproblem in der Standardform zu schreiben, fassen wir die Matrizen C und X als Vektoren $\hat{c}$ und $\hat{x}$ im $\mathbb{R}^{n^2}$ auf, indem wir die Spalten untereinander schreiben. Wir setzen also $\hat{x}_{i+(j-1)n} = x_{ij}$ für $i, j = 1, \ldots, n$ und genauso $\hat{c}$. Mit $\hat{d} \in \mathbb{R}^{n^2}$, definiert durch

$$\hat{d} \;=\; \big(0,\ldots,0|0,\ldots,0|\cdots|0,\ldots,0|1,\ldots,1\big)^\top\,,$$

ist $W(X) = \sum_{k=1}^{n} x_{kn} = \hat{d}^\top\hat{x}$. Hier haben wir eine offensichtliche Blockschreibweise benutzt.

Die Matrizen $A, B \in \mathbb{R}^{(n-2)\times n^2}$ seien definiert durch

$$A = \left[\begin{array}{ccc|ccc|ccc|c|ccc|ccc} 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 & & 0 & \cdots & 0 & 0 & \cdots & 0 \\ & \vdots & & & \vdots & & & \vdots & & \cdots & & \vdots & & & \vdots & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & & 1 & \cdots & 1 & 0 & \cdots & 0 \end{array}\right]$$

$$B = \left[\begin{array}{cccccc|cccccc|c|cccccc} 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & & 0 & 0 & 1 & \cdots & 0 & 0 \\ & & & \vdots & & & & & & \vdots & & & \cdots & & & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & & 0 & 0 & 0 & \cdots & 1 & 0 \end{array}\right]$$

Dann ist das Fluss-Maximierungsproblem (P) äquivalent zu

(P) $$\text{Maximiere} \quad \hat{d}^\top \hat{x} \quad \text{unter} \quad \hat{x} \geq 0\,,\ \hat{x} \leq \hat{c}\,,\ (A-B)\hat{x} = 0\,.$$

Wir wollen das duale Problem herleiten. Dazu müssen wir das vorliegende Problem mit Hilfe von Schlupfvariablen $\hat{z} = \hat{c} - \hat{x}$ auf Standardform transformieren:

$$\text{Minimiere} \quad \begin{bmatrix} -\hat{d} \\ 0 \end{bmatrix}^\top \begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} \text{ unter } \begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} \geq 0\,, \quad \begin{bmatrix} I & I \\ A-B & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} \hat{c} \\ 0 \end{bmatrix}$$

Die Variablen des dualen Problems bezeichne ich mit $\hat{y} \in \mathbb{R}^{n^2}$ und $u \in \mathbb{R}^{n-2}$. Dann ist das duale Problem:

$$\text{Maximiere} \quad \hat{y}^\top \hat{c} \quad \text{unter} \quad \begin{bmatrix} I & A^\top - B^\top \\ I & 0 \end{bmatrix} \begin{bmatrix} \hat{y} \\ u \end{bmatrix} \leq \begin{bmatrix} -\hat{d} \\ 0 \end{bmatrix}, \quad \text{d.h.}$$

$$\text{Maximiere} \quad \hat{y}^\top \hat{c} \quad \text{unter} \quad \hat{y} + (A^\top - B^\top)u \leq -\hat{d}\,,\ \hat{y} \leq 0\,,$$

d.h. nach Ersetzung von $\hat{y}$ durch $-\hat{y}$,

(D) $$\text{Minimiere} \quad \hat{y}^\top \hat{c} \quad \text{unter} \quad (A^\top - B^\top)u + \hat{d} \leq \hat{y}\,,\ \hat{y} \geq 0\,.$$

Wir berechnen jetzt $\hat{v} = (A^\top - B^\top)u + \hat{d} \in \mathbb{R}^{n^2}$. Zunächst ist

$$A^\top = \left[\begin{array}{cccc} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ \hline 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ \hline 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{array}\right], \quad B^\top = \left[\begin{array}{cccc} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{array}\right] = \begin{bmatrix} H \\ \vdots \\ H \end{bmatrix} \text{ mit } H = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Die Komponenten des Vektors $u \in \mathbb{R}^{n-2}$ bezeichnen wir mit $u_2, \ldots, u_{n-1}$ und führen den Vektor $e := (1, \ldots, 1)^\top \in \mathbb{R}^n$ ein. Damit ist

$$\hat{v} = \begin{bmatrix} -Hu \\ u_2 e - Hu \\ \vdots \\ u_{n-1} e - Hu \\ e - Hu \end{bmatrix}.$$

Jetzt ergänzen wir u durch Setzen von $u_1 = 0$ und $u_n = 1$ zu einem Vektor im $\mathbb{R}^n$ und ergänzen H durch Einfügen einer ersten Nullspalte und einer letzten Nullspalte zu einer Matrix $\hat{H} \in \mathbb{R}^{n \times n}$. Dann ist

$$\hat{v} = \begin{bmatrix} u_1 e - \hat{H}u \\ \vdots \\ u_n e - \hat{H}u \end{bmatrix}$$

Jetzt schreiben wir den Vektor $\hat{v} \in \mathbb{R}^{n^2}$ wieder als Matrix $V \in \mathbb{R}^{n \times n}$:

$$V = \left[u_1 e - \hat{H}u \,|\, \cdots \,|\, u_n e - \hat{H}u\right].$$

Damit ist

$$V_{ij} \;=\; \begin{cases} u_j - u_i\,, & \text{für } i = 1,\dots,n-1\,, \\ u_j\,, & \text{für } i = n\,, \end{cases} \qquad \text{und } j = 1,\dots,n\,.$$

Damit können wir das duale Problem wieder in Matrizenform schreiben. Die Matrix $Y \in \mathbb{R}^{n\times n}$ entsteht natürlich aus $\hat{y}$ indem die Blöcke von $\hat{y}$ als Spalten in Y gespeichert werden. Dann ist (D) äquivalent zu:

$$\text{(D)} \qquad \text{Minimiere} \quad \sum_{i,j=1}^{n} c_{ij}\, y_{ij} \quad \text{unter} \quad y_{ij} \geq u_j - u_i \text{ und } y_{ij} \geq 0 \text{ für } i,j = 1,\dots,n\,,$$

$$\text{und } u_1 = 0,\ u_n = 1\,.$$

Man beachte, dass wir hier ausgenutzt haben, dass die Summation bzgl. i wegen $c_{nj} = 0$ nur bis $n-1$ geht (d.h. der Wert von y_{nj} ist uninteressant).

Der starke Dualitätssatz liefert die Existenz von Lösungen X^* des primalen und $(Y^*, u^*) \in \mathbb{R}^{n\times n} \times \mathbb{R}^n$ des dualen Problems mit $\max(P) = W(X^*) = \min(D)$. Wir konstruieren jetzt einen optimalen Schnitt. Dazu benötigen wir entscheidend die Komplementaritätsbedingung vom Kapitel 3. Diese besagt in unserem Fall, formuliert zunächst in Vektorform mit Schlupfvariablen: $\hat{x}_i = 0$ oder $\hat{v}_i = \hat{y}_i$ für jedes i, und $\hat{z}_i = 0$ oder $\hat{y}_i = 0$ für jedes i. Umgeschrieben in die Matrizenform bedeutet dies:

$$x^*_{ij} \;=\; 0 \quad \text{oder} \quad u^*_j - u^*_i \;=\; y^*_{ij} \quad \text{für } i = 1,\dots,n-1,\ j = 1,\dots,n, \tag{4.3a}$$

$$x^*_{ij} \;=\; c_{ij} \quad \text{oder} \quad y^*_{ij} \;=\; 0 \quad \text{für jedes } i,j = 1,\dots,n\,. \tag{4.3b}$$

Wir definieren nun den Schnitt (J^-, J^+) durch $J^- = \{i : u^*_i \leq 0\}$ und $J^+ = \{i : u^*_i > 0\}$. Dann ist $1 \in J^-$ und $n \in J^+$ und (J^-, J^+) ein Schnitt. Zur Berechnung seiner Kapazität sei $i \in J^-$ und $j \in J^+$ sowie $c_{ij} > 0$. Wäre $x^*_{ij} < c_{ij}$, so nach (4.3b) $y^*_{ij} = 0$, also $u^*_j - u^*_i \leq y^*_{ij} = 0$, ein Widerspruch zu $i \in J^-$ und $j \in J^+$. Also ist $x^*_{ij} = c_{ij}$ für alle $i \in J^-$ und $j \in J^+$ mit $c_{ij} > 0$. Daher ist

$$K(J^-, J^+) \;=\; \sum_{\substack{i\in J^- \\ j\in J^+ \\ c_{ij}>0}} c_{ij} \;=\; \sum_{\substack{i\in J^- \\ j\in J^+}} x^*_{ij} \;=\; W(X^*) \;+\; \sum_{\substack{i\in J^- \\ j\in J^+}} x^*_{ji}\,.$$

Wir zeigen schließlich, dass die letzte Summe verschwindet. Sei wieder $i \in J^-$ und $j \in J^+$. Wäre $x^*_{ji} > 0$, so wäre nach (4.3a) $u^*_i - u^*_j = y^*_{ji} \geq 0$, also $u^*_i \geq u^*_j$, ein Widerspruch zu $i \in J^-$ und $j \in J^+$. Daher ist $x^*_{ji} = 0$ für alle $i \in J^-$ und $j \in J^+$. Damit ist das folgende Max-Flow-Min-Cut-Theorem bewiesen:

Satz 4.5 *Es gilt:*

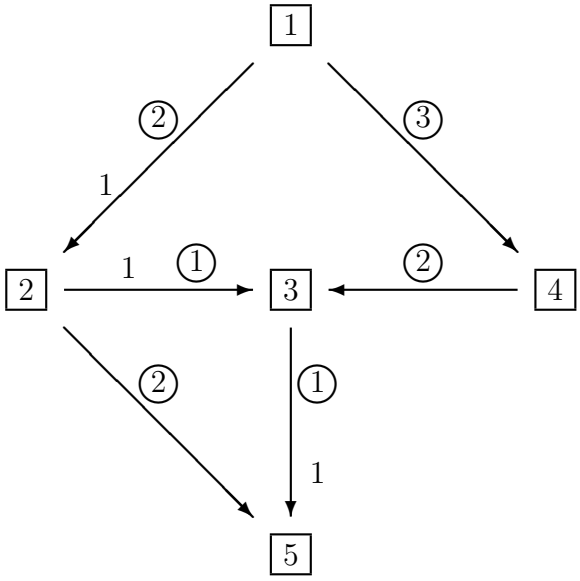
$$\max\big\{W(X) : X \textit{ Fluss}\big\} \;=\; \min\big\{K(J^-, J^+) : (J^-, J^+) \textit{ Schnitt}\big\}\,.$$

4.2 Der Algorithmus von Ford-Fulkersen

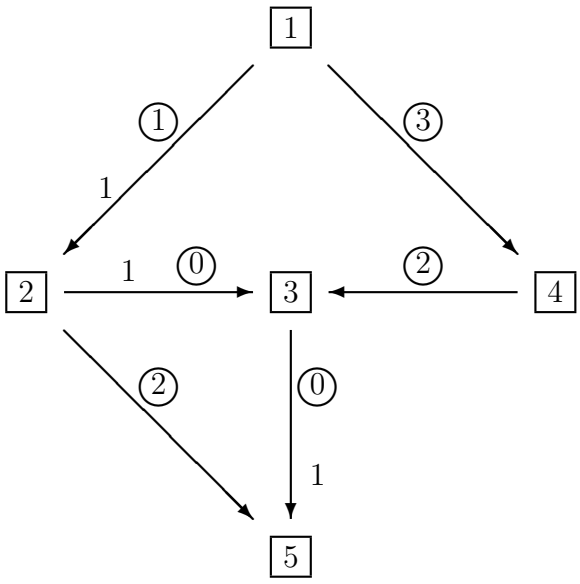
Dieser Abschnitt wird in der Vorlesung aus Zeitgründen weggelassen.

Wir beginnen mit einem kleinen Beispiel, an dem wir den Algorithmus erläutern. Die Zahlen in den Kreisen beschreiben die Kapazitäten der Kanten.

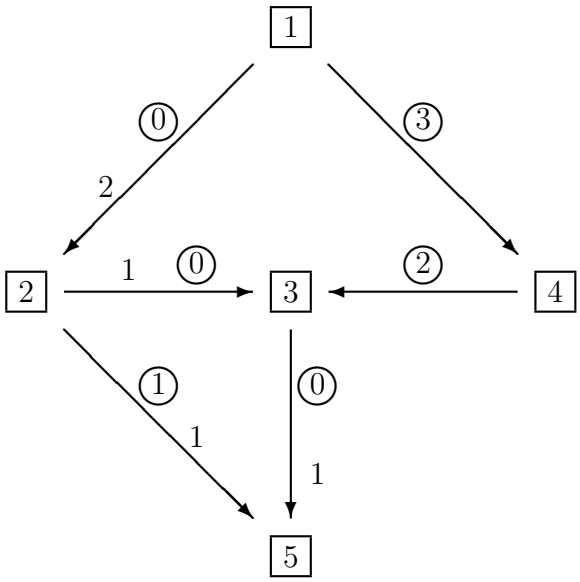
Man beginne mit einem (kleinen) Fluss.



Hier ist $\boxed{1}$=Q und $\boxed{5}$=S. Ich schicke zunächst einen Fluss mit Wert 1 von $\boxed{1}$ über $\boxed{2}$ und $\boxed{3}$ nach $\boxed{5}$ und markiere ihn, indem ich die 1 an die Kante auf der Seite der Kapazität schreibe. Diesen merke ich mir und zeichne einen neuen Graphen, in dem ich außer dem Fluss nur noch die **Überschusskapazitäten** eintrage (oberhalb bzw. rechts vom Pfeil):

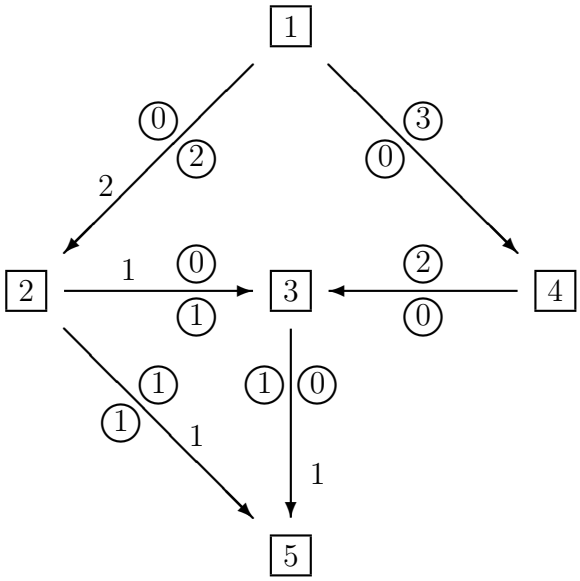


Ich versuche jetzt, einen weiteren Fluss durch das Netzwerk zu schicken, etwa der Stärke 1 von 1 über 2 nach 5. Auch diesen merke ich mir und zeichne einen neuen Graphen:

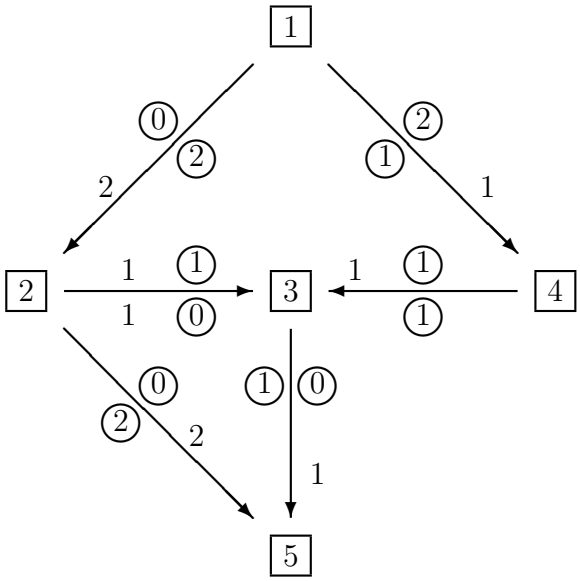


Wie geht es weiter? Ich kann versuchen, einen Fluss von 1 über 4 nach 3 zu schicken. Dann geht es aber nicht weiter. Wir haben bisher noch keine Möglichkeit, eine früher getroffene Entscheidung rückgängig zu machen. Von 2 kommt ja ein Fluss der Stärke 1. Ich kann diesen „übernehmen“, also meinen neuen Fluss von 1 über 4 und 3 nach 5 zu schicken und den alten von 1 über 2 und 3 nach 5 „umzuleiten“. Dies wird in systematischer Weise anders realisiert. Wir betrachten allgemein eine Kante von A nach B mit der ursprünglichen Kapazität 3. Ich kann dann der Kante von B nach A

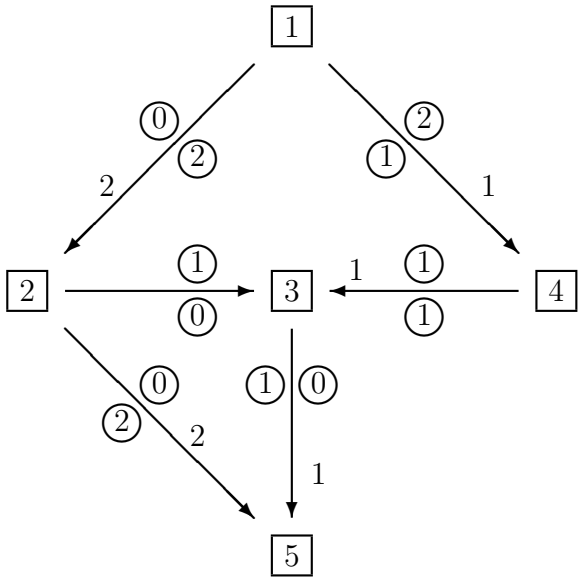
die Kapazität 0 zuordnen (von [B] nach [A] darf ja nichts fließen). Es sei jetzt ein Fluss der Stärke 2 vorhanden. Dann ist die Überschusskapazität der Kante von [A] nach [B] gerade 3-2=1. Die Überschusskapazität der entgegengesetzten Kante von [B] nach [A] ist jetzt 0+2=2, denn ich kann ja einen zusätzlichen Fluss bis zur Stärke 2 von [B] nach [A] schicken, indem ich den ursprünglichen entsprechend reduziere. Im folgenden Graphen zeichne ich die Überschusskapazitäten der entgegengesetzten Kanten unterhalb bzw. links ein:



Ich kann jetzt einen Fluss der Stärke 1 von [1] über [4] und [3] und [2] nach [5] schicken und erhalte:



Die Flüsse zwischen $\boxed{2}$ und $\boxed{3}$ heben sich auf, und wir erhalten



Dieser Fluss ist optimal, da der Schnitt durch die Kanten $\overline{25}$ und $\overline{35}$ die Kapazität 3 hat. Damit haben wir den Algorithmus von **Ford-Fulkerson** beschrieben.

Wir wollen dies jetzt formalisieren.

Definition 4.6 *Sei X ein Fluss. Ein* **ungesättigter Pfad P vom Knoten j zum Knoten** *k ist ein Indexvektor $P = (p_1, \ldots, p_m) \in \mathbb{N}^m$ mit $p_i \in \{1, \ldots, n\}$, $i = 1, \ldots, m$, $p_1 = j$, $p_m = k$, und für jedes $i = 1, \ldots, m-1$ gilt:*

- $x_{p_i p_{i+1}} < c_{p_i p_{i+1}}$, *falls* $c_{p_i p_{i+1}} > 0$, *und*
- $x_{p_{i+1} p_i} > 0$, *falls* $c_{p_i p_{i+1}} = 0$.

Jetzt können wir beweisen:

Satz 4.7 *(Ford-Fulkerson) Es gilt:*

$$\max\big\{W(X) : X \text{ Fluss}\big\} \;=\; \min\big\{K(J^-, J^+) : (J^-, J^+) \text{ Schnitt}\big\}\,.$$

Beweis: Sei X ein maximaler Fluss. Definiere die Menge

$$J^- \;:=\; \{1\} \;\cup\; \big\{k \in \{2, \ldots, n\} : \text{es gibt ungesättigten Pfad von 1 nach } k\big\}\,.$$

Wir zeigen zunächst $n \notin J^-$. Annahme, $n \in J^-$. Dann gibt es also einen ungesättigten Pfad $P = (p_1, \ldots, p_m)$ von 1 nach n. Mit I_1 und I_2 bezeichnen wir die „Vorwärts-“ und die „Rückwärtskanten“, d.h.

$$\begin{aligned} I_1 &:= \big\{i \in \{1, \ldots, m-1\} : c_{p_i p_{i+1}} > 0\big\}\,, \quad \text{und} \\ I_2 &:= \big\{i \in \{1, \ldots, m-1\} : c_{p_i p_{i+1}} = 0\big\}\,. \end{aligned}$$

Dann ist $I_1 \cup I_2 = \{1, \ldots, m-1\}$ und $I_1 \cap I_2 = \emptyset$. Setze

$$d \;:=\; \min\Big(\big\{ c_{p_i p_{i+1}} - x_{p_i p_{i+1}} : i \in I_1 \big\} \;\cup\; \big\{ x_{p_{i+1} p_i} : i \in I_2 \big\} \Big) .$$

Dann ist $d > 0$. Definiere einen neuen Fluss $\tilde{X}$ durch

$$\tilde{x}_{k\ell} \;:=\; \begin{cases} x_{p_i p_{i+1}} \,+\, d\,, & \text{falls } k = p_i \text{ und } \ell = p_{i+1} \text{ für ein } i \in I_1\,, \\ x_{p_{i+1} p_i} \,-\, d\,, & \text{falls } k = p_{i+1} \text{ und } \ell = p_i \text{ für ein } i \in I_2\,, \\ x_{k\ell}\,, & \text{sonst} \end{cases}$$

Dann ist $\tilde{X}$ wirklich ein Fluss, denn für $k \in \{2, \ldots, n-1\}$ können die folgenden Fälle auftreten:

- 1. Fall: $k \notin \{p_1, \ldots, p_m\}$. Dann ist $\tilde{x}_{k\ell} = x_{k\ell}$ für alle $\ell = 1, \ldots, n$, also

$$\sum_{\ell=1}^{n} \tilde{x}_{k\ell} \;=\; \sum_{\ell=1}^{n} x_{k\ell} \;=\; \sum_{\ell=1}^{n} x_{\ell k} \;=\; \sum_{\ell=1}^{n} \tilde{x}_{\ell k}\,.$$

- 2. Fall: $k = p_i$ für ein $i \in \{2, \ldots, m-1\}$. Es ist

$$\tilde{x}_{p_i p_{i-1}} \;+\; \tilde{x}_{p_i p_{i+1}} \;=\; x_{p_i p_{i-1}} \;+\; x_{p_i p_{i+1}} \;+\; \rho_i\,,$$

  wobei

$$\rho_i \;:=\; \begin{cases} d\,, & \text{falls } i, i-1 \in I_1\,, \\ -d\,, & \text{falls } i, i-1 \in I_2\,, \\ 0\,, & \text{sonst} \end{cases}$$

  und

$$\tilde{x}_{p_{i-1} p_i} \;+\; \tilde{x}_{p_{i+1} p_i} \;=\; x_{p_{i-1} p_i} \;+\; x_{p_{i+1} p_i} \;+\; \rho_i\,,$$

  also

$$\sum_{\ell=1}^{n} \tilde{x}_{p_i \ell} \;=\; \sum_{\ell=1}^{n} x_{p_i \ell} \;+\; \rho_i \;=\; \sum_{\ell=1}^{n} x_{\ell p_i} \;+\; \rho_i \;=\; \sum_{\ell=1}^{n} \tilde{x}_{\ell p_i}\,.$$

  Also gilt auch in diesem Fall die Gleichgewichtsbedingung.

Schließlich ist

$$\begin{aligned} W(\tilde{X}) \;&=\; \sum_{\ell=1}^{n} \tilde{x}_{1\ell} \;=\; \tilde{x}_{p_1 p_2} \;+\; \sum_{\ell \neq p_2}^{n} \tilde{x}_{1\ell} \\ &=\; x_{p_1 p_2} \;+\; d \;+\; \sum_{\ell \neq p_2}^{n} x_{1\ell} \;=\; W(X) \;+\; d \\ &>\; W(\tilde{X})\,, \end{aligned}$$

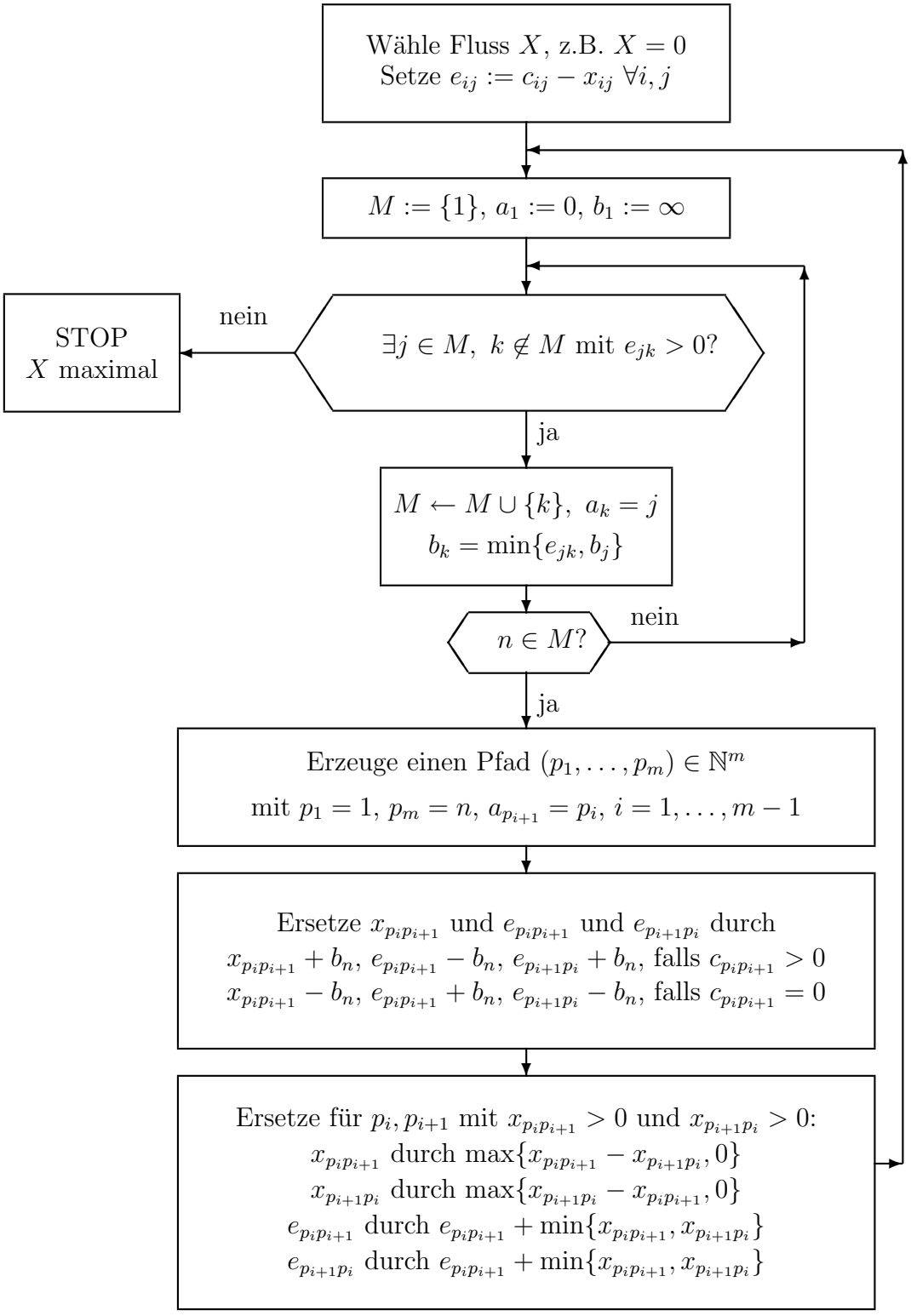
da $\tilde{x}_{p_1 p_2} = x_{p_1 p_2} + d$ wegen $1 \in I_1$ und $1 \notin I_2$. (Sonst wäre $c_{p_1 p_2} = 0$, also $x_{p_2 1} = x_{p_2 p_1} > 0$, und dies geht nicht.) Also ist X nicht maximal, ein Widerspruch. Also ist $n \notin J^-$.

Definiere $J^+ := \{1, \ldots, n\} \setminus J^-$. Dann ist (J^-, J^+) ein Schnitt. Sei $k \in J^-$ und $\ell \in J^+$. Wir behaupten, dass sowohl $x_{k\ell} = c_{k\ell}$ als auch $x_{\ell k} = 0$ ist. Wäre nämlich $x_{k\ell} < c_{k\ell}$ oder $x_{\ell k} > 0$, so können wir den Pfad $P = (p_1, \ldots, p_m)$ von 1 nach k durch $p_{m+1} := \ell$ verlängern: Im Fall $c_{k\ell} = 0$ kann nur $x_{\ell k} > 0$ sein, und im Fall $c_{k\ell} > 0$ ist $c_{\ell k} = 0$, also kann nur $x_{k\ell} < c_{k\ell}$ auftreten. Dies sind die Bedingungen für einen ungesättigten Pfad.
Damit ist nach der ersten Gleichung im schwachen Dualitätssatz:

$$W(X) \;=\; \sum_{\substack{k \in J^- \\ \ell \in J^+}} \big[x_{k\ell} \;-\; x_{\ell k}\big] \;=\; \sum_{\substack{k \in J^- \\ \ell \in J^+}} c_{k\ell} \;=\; K(J^-, J^+)\,.$$

Dies ist zu zeigen □

Wir beschreiben jetzt den am Beispiel erläuterten Algorithmus. Dazu werden den im Laufe des Verfahrens erreichten Ecken i gewisse Marken (a_i, b_i) zugordnet. a_i ist die Nummer des Vorgängers der Ecke i, und b_i der Wert des bis Ecke i durchgeschickten Flusses.


Wähle Fluss X, z.B. $X = 0$
Setze $e_{ij} := c_{ij} - x_{ij}\ \forall i, j$
$M := \{1\}$, $a_1 := 0$, $b_1 := \infty$
$\exists j \in M,\ k \notin M$ mit $e_{jk} > 0$?
nein
STOP
X maximal
ja
$M \leftarrow M \cup \{k\},\ a_k = j$
$b_k = \min\{e_{jk}, b_j\}$
$n \in M$?
nein
ja
Erzeuge einen Pfad $(p_1, \ldots, p_m) \in \mathbb{N}^m$
mit $p_1 = 1$, $p_m = n$, $a_{p_{i+1}} = p_i$, $i = 1, \ldots, m-1$
Ersetze $x_{p_ip_{i+1}}$ und $e_{p_ip_{i+1}}$ und $e_{p_{i+1}p_i}$ durch
$x_{p_ip_{i+1}} + b_n$, $e_{p_ip_{i+1}} - b_n$, $e_{p_{i+1}p_i} + b_n$, falls $c_{p_ip_{i+1}} > 0$
$x_{p_ip_{i+1}} - b_n$, $e_{p_ip_{i+1}} + b_n$, $e_{p_{i+1}p_i} - b_n$, falls $c_{p_ip_{i+1}} = 0$
Ersetze für p_i, p_{i+1} mit $x_{p_ip_{i+1}} > 0$ und $x_{p_{i+1}p_i} > 0$:
$x_{p_ip_{i+1}}$ durch $\max\{x_{p_ip_{i+1}} - x_{p_{i+1}p_i}, 0\}$
$x_{p_{i+1}p_i}$ durch $\max\{x_{p_{i+1}p_i} - x_{p_ip_{i+1}}, 0\}$
$e_{p_ip_{i+1}}$ durch $e_{p_ip_{i+1}} + \min\{x_{p_ip_{i+1}}, x_{p_{i+1}p_i}\}$
$e_{p_{i+1}p_i}$ durch $e_{p_ip_{i+1}} + \min\{x_{p_ip_{i+1}}, x_{p_{i+1}p_i}\}$

4.3 Ausflug in die Spieltheorie

In diesem Abschnitt betrachten wir Anwendungen der Optimierungstheorie in der Spieltheorie, speziell auf 2-Personen-Nullsummenspiele (Matrix-Spiele).

Spielsituation: Spieler P_1 hat Alternativen $a_1, \ldots, a_n$, Spieler P_2 die Alternativen $b_1, \ldots, b_m$. Wählt P_1 die Aktion a_j und P_2 die Aktion b_i, so ergibt ich die Auszahlung a_{ij} von P_1 an P_2. Damit ist das Spiel vollständig durch die Matrix

$$A \;=\; \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

beschrieben, daher der Name Matrix-Spiel. (Die Zahlung a_{ij} von P_1 an P_2 ist das, was P_2 erhält, daher der Name Nullsummenspiel. Es werden keine Zahlungen von außen getätigt.) $a_{ij} < 0$ bedeutet, P_2 zahlt $-a_{ij}$ an P_1.

Als **Beispiel** betrachten wir als erstes das bekannte Knobelspiel „Stein, Schere, Papier“. Wir können es durch die folgende Tabelle darstellen:

$P_2 \setminus P_1$	Stein	Schere	Papier
Stein	0	1	–1
Schere	–1	0	1
Papier	1	–1	0

Dies bedeutet offensichtlich: Spielt der Spieler P_1 etwa „Papier“, und der Spieler P_2 „Schere“ so zahlt P_1 an P_2 eine Einheit („Schere schneidet Papier“). Die Tabelle liefert also die Auszahlungsmatrix von P_1 an P_2

$$A \;=\; \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Spiele sind für ökonomische Anwendungen interessant, wenn sie vielfach in gleichartiger Weise gespielt werden. Wie soll man sich dann langfristig (im Mittel) vernünftig verhalten? Dazu wählen P_1 und P_2 Strategien

$$x \;=\; (x_1, \ldots, x_n)^\top, \quad y \;=\; (y_1, \ldots, y_m)^\top,$$

mit $x_j \geq 0$, $\sum_{j=1}^n x_j = 1$ und analog $y_i \geq 0$, $\sum_{i=1}^m y_i = 1$. Die Zahl x_j gibt die relative Häufigkeit (Wahrscheinlichkeit) dafür an, mit der die Aktion x_j eingesetzt wird. Die Strategie $x = (0, \ldots, 0, 1, 0, \ldots, 0)^\top =: e^{(j)}$ (also $j-$ter Einheitsvektor) bedeutet, dass immer a_j benutzt wird (reine Strategie, die allgemeinen Vektoren heißen entsprechend gemischte Strategien). Die Zahl

$$\Phi(x, y) \;:=\; y^\top A x \;=\; \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\, y_i\, x_j$$

gibt die erwartete (mittlere) Auszahlung an, die P_1 an P_2 zahlt, wenn die Strategien x und y benutzt werden.

Beispiel Knobeln (s.o.): Hier stellt man schnell fest (als Spieler), dass $x^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^\top$ und $y^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^\top$ „vernünftige“ Strategien sind. Es ist $\Phi(x^*, y^*) = 0$.

Wie soll sich P_1 optimal verhalten? P_1 sollte eine Strategie $x \in \mathbb{R}^n$ suchen mit $x \geq 0$, $\sum_{j=1}^n x_j = 1$ und minimalem

$$\max_{\substack{y \geq 0 \\ y^\top e = 1}} y^\top A x \,,$$

d.h. er möchte die größtmögliche Auszahlung an P_2 minimieren. Hier ist $e = (1, 1, \ldots, 1)^\top \in \mathbb{R}^m$. Die Existenz des Maximums ist natürlich gesichert, da die Restriktionenmenge kompakt und die Zielfunktion stetig ist. Umgekehrt will der Spieler P_2 die kleinstmögliche Einnahme von P_1 maximieren, also das Problem lösen:

$$\text{maximiere} \quad \min_{\substack{x \geq 0 \\ x^\top e = 1}} y^\top A x \quad \text{unter } y \geq 0 \,,\ y^\top e = 1 \,.$$

Hier ist $e = (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$. Wir haben also die folgenden beiden Optimierungsprobleme:

$$(P_1) \qquad \text{Minimiere} \quad \max_{\substack{y \geq 0 \\ y^\top e = 1}} y^\top A x \quad \text{unter } x \geq 0 \,,\ x^\top e = 1 \,,$$

$$(P_2) \qquad \text{Maximiere} \quad \min_{\substack{x \geq 0 \\ x^\top e = 1}} y^\top A x \quad \text{unter } y \geq 0 \,,\ y^\top e = 1 \,.$$

Man könnte versuchen, (P_1) als lineare Minimierungsaufgabe in der folgenden Form zu schreiben:

$$\text{Minimiere } x_0 \text{ unter} \quad x \geq 0 \,,\ x^\top e = 1 \text{ und}$$
$$(A^\top y)^\top x \ - \ x_0 \ \leq \ 0 \quad \text{für alle } y \in N \,,$$

wobei $N := \{y \in \mathbb{R}^m : y \geq 0 \,,\ y^\top e = 1\}$. In dieser Form ist dies ein *semifinites* lineares Optimierungsproblem, da es endlich viele Unbekannte, aber unendliche viele Nebenbedingungen hat. Wir führen es auf ein finites lineares Optimierungsproblem zurück und zeigen:

Lemma 4.8 *Sei $A \in \mathbb{R}^{m \times n}$.*

(a) Für festes $x \in \mathbb{R}^n$ ist

$$\max_{\substack{y \geq 0 \\ y^\top e = 1}} y^\top A x \ = \ \max_{i=1,\ldots,m} (Ax)_i \,.$$

(b) Für festes $y \in \mathbb{R}^m$ ist

$$\min_{\substack{x \geq 0 \\ x^\top e = 1}} y^\top A x \ = \ \min_{j=1,\ldots,n} (A^\top y)_j \,.$$

Beweis nur von (b): Für festes $y \in \mathbb{R}^m$ ist die linke Seite eine lineare Optimierungsaufgabe in Standardform:

$$\text{Minimiere} \quad (A^\top y)^\top x \quad \text{unter } x \geq 0\,,\ x^\top e = 1\,.$$

Wir wissen, dass der Minimalwert in einer Ecke angenommen wird. Die Ecken des Simplex $\left\{x \in \mathbb{R}^n : x \geq 0,\ e^\top x = 1\right\}$ sind aber gerade die Einheitsvektoren $\left\{e_{*j} \in \mathbb{R}^n : j = 1, \ldots, n\right\}$ (weshalb?). Wegen $(A^\top y)^\top e_{*j} = (A^\top y)_j$ ist die Behauptung (b) gezeigt. (a) verläuft analog. □

Daher können wir (P_1) und (P_2) umschreiben in

$$(P_1) \qquad \text{Minimiere} \quad \max_{i=1,\ldots,m} (Ax)_i \quad \text{unter } x \geq 0\,,\ e^\top x = 1\,,$$

$$(P_2) \qquad \text{Maximiere} \quad \min_{j=1,\ldots,n} (A^\top y)_j \quad \text{unter } y \geq 0\,,\ e^\top y = 1\,,$$

oder auch

$$(P_1^*) \qquad \text{Minimiere} \quad x_0 \quad \text{unter } x \geq 0\,,\ e^\top x = 1\,,\ Ax \leq x_0 e\,,$$

$$(P_2^*) \qquad \text{Maximiere} \quad y_0 \quad \text{unter } y \geq 0\,,\ e^\top y = 1\,,\ A^\top y \geq y_0 e\,.$$

Durch die Einführung von Schlupfvariablen bringen wir (P_1^*) in die Normalform: $z = x_0 e - Ax$ und $x_0 = x_0^+ - x_0^-$ mit $x_0^\pm \geq 0$ liefert:

$$\text{Minimiere} \quad x_0^+ - x_0^- \quad \text{unter}$$

$$\left(x_0^+, x_0^-, x, z\right) \geq 0\,, \quad \begin{bmatrix} 0 & 0 & e_n^\top & 0 \\ -e_m & e_m & A & I_m \end{bmatrix} \begin{bmatrix} x_0^+ \\ x_0^- \\ x \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0_m \end{bmatrix}.$$

Hier ist also $c = (1, -1, 0, \ldots, 0)^\top \in \mathbb{R}^{2+n+m}$. Das hierzu duale Problem ist

$$\text{Maximiere} \quad y_0 \quad \text{unter} \quad \begin{bmatrix} 0 & -e_m^\top \\ 0 & e_m^\top \\ e_n & A^\top \\ 0 & I_m \end{bmatrix} \begin{bmatrix} y_0 \\ \tilde{y} \end{bmatrix} \leq \begin{bmatrix} 1 \\ -1 \\ 0_n \\ 0_m \end{bmatrix},$$

d.h. mit $y = -\tilde{y}$:

$$\text{Maximiere} \quad y_0 \quad \text{unter } y^\top e = 1\,,\ y \geq 0\,,\ A^\top y \geq y_0 e\,,$$

und dies ist genau das Problem (P_2^*). Damit sind (P_1) und (P_2) zueinander dual, und wir haben mit dem starken Dualitätssatz:

Satz 4.9 *(Hauptsatz der Matrixspiele)*
Es gibt $x^* \in \mathbb{R}^n$ und $y^* \in \mathbb{R}^m$ mit $e^\top x^* = 1$ und $e^\top y^* = 1$ und

$$y^{*\top} A x^* \;=\; \min_{\substack{x\geq 0\\ e^\top x=1}} \max_{\substack{y\geq 0\\ e^\top y=1}} y^\top A x \;=\; \max_{\substack{y\geq 0\\ e^\top y=1}} \min_{\substack{x\geq 0\\ e^\top x=1}} y^\top A x \;=:\; v\,.$$

Diese Größe v heißt der **Wert** *des Matrixspiels. Das Matrixspiel heißt* **fair**, *wenn der Wert Null ist.*

Beweis: Die Existenz von optimalen Lösungen x^* und y^* sowie die Gleichheit der Werte liefert der starke Dualitätssatz. Es bleibt das erste Gleichheitszeichen in der Formel zu zeigen. Es ist

$$\begin{aligned} y^{*\top} A x^* &\leq \max_{\substack{y\geq 0\\ e^\top y=1}} y^\top A x^* \;=\; \max_{i=1,\ldots,m} (Ax^*)_i \;=\; v\,,\\ y^{*\top} A x^* &\geq \min_{\substack{x\geq 0\\ e^\top x=1}} y^{*\top} A x \;=\; \min_{j=1,\ldots,n} (A^\top y^*)_j \;=\; v\,. \end{aligned}$$

Damit gilt Gleichheit. □

Eine äquivalente Formulierung liefert den **Gleichgewichtssatz von Nash**, für den er 1994 den Nobelpreis erhalten hat („Beautiful Mind“): Jedes Matrixspiel A besitzt einen Gleichgewichtspunkt (x^*, y^*) in gemischten Strategien, d.h. es gilt:

$$\Phi(x^*, y) \;\leq\; \Phi(x^*, y^*) \;\leq\; \Phi(x, y^*) \quad \text{für alle Strategien } x, y\,.$$

(Eigentlich besteht seine Leistung in einer Verallgemeinerung dieser Aussage.) Zu Sattelpunktsätzen kommen wir noch im Rahmen der konvexen Optimierung.

Für eine bestimmte Klasse von Problemen können wir sofort sagen, dass sie fair sind.

Satz 4.10 *Ist $A \in \mathbb{R}^{n\times n}$ schiefsymmetrisch, d.h. $A^\top = -A$ (insbesondere muss A quadratisch sein), so ist der Wert des Spiels* 0.

Beweis: Es ist

$$v \;=\; \min_{x\in M} \max_{y\in M} y^\top A x \;=\; \max_{y\in M} \min_{x\in M} y^\top A x\,,$$

wobei $M = \{z \in \mathbb{R}^n : z \geq 0,\ e^\top z = 1\}$. Ferner ist

$$y^\top A x \;=\; x^\top A^\top y \;=\; -x^\top A y\,,$$

also

$$v \;=\; \min_{x\in M} \max_{y\in M} \big(-x^\top A y\big) \;=\; -\max_{x\in M} \min_{y\in M} x^\top A y \;=\; -v$$

wegen $\min_x \big(-f(x)\big) = -\max_x f(x)$ und $\max_x \big(-f(x)\big) = -\min_x f(x)$. Also ist $v = -v$, d.h. $v = 0$. □

Als Beispiel betrachten wir das „Stein, Schere, Papier Spiel“:

Beispiel 4.11

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

Da diese Matrix schiefsymmetrisch ist, ist das Knobelspiel fair.

Als zweites Beispiel betrachten wir das **Skinspiel**:[2] P_1 und P_2 haben je drei Karten und zwar: P_1 die Karten Pik As, Karo As, Pik 2, und P_2 die Karten Pik As, Karo As und Karo 2. Beide Spieler legen gleichzeitig eine Karte auf den Tisch. P_2 gewinnt, wenn die Karten gleiche Farbe haben, andernfalls P_1. Ein As hat den Wert 1, eine Zwei den Wert 2. Es wird durch die folgende Tabelle beschrieben:

$P_2 \setminus P_1$	♢ As	♠ As	♠ 2
♢ As	1	−1	−2
♠ As	−1	1	1
♢ 2	2	−1	−2

und hat die Auszahlungsmatrix $A = \begin{pmatrix} 1 & -1 & -2 \\ -1 & 1 & 1 \\ 2 & -1 & -2 \end{pmatrix}$

Das Spiel sieht unfair aus, da die Auszahlungsmatrix 5 negative und nur 4 positive Einträge hat. Wir werden auf dieses Skinspiel im nächsten Kapitel zurückkommen.

Besonders einfach sind die sogenannten Sattelpunktspiele.

Definition 4.12 *Ein Spiel, beschrieben durch die Matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, heißt* **Sattelpunktspiel**, *wenn*

$$\min_{j=1,\dots,n} \max_{i=1,\dots,m} a_{ij} = \max_{i=1,\dots,m} \min_{j=1,\dots,n} a_{ij}\,.$$

Weder das Knobelspiel „Stein, Schere, Papier“

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

noch das Skinspiel

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

sind Sattelpunktspiele, denn es ist für das Knobelspiel

$$\min_{j=1,\dots,3} \max_{i=1,\dots,3} a_{ij} = \min\{1,1,1\} = 1 \text{ und } \max_{i=1,\dots,3} \min_{j=1,\dots,3} a_{ij} = \max\{-1,-1,-1\} = -1\,,$$

[2]Nach Kuhn, 1957. Kuhn war Kollege von Nash.

und analog für das Skinspiel. Dagegen beschreibt etwa die Auszahlungsmatrix

$$A = \begin{pmatrix} 3 & 1 & 4 \\ 2 & 0 & 1 \\ -1 & -2 & -2 \end{pmatrix}$$

ein Sattelpunktspiel.

Satz 4.13 *Für jedes Sattelpunktspiel gibt es optimale Lösungen x^* und y^*, die aus* **reinen Strategien** *bestehen. Das bedeutet, dass x^* und y^* Einheitsvektoren sind. Genauer gilt: Ist*

$$a_{i_0 j_0} = \max_{i=1,\ldots,m} \min_{j=1,\ldots,n} a_{ij} \,,$$

so sind x^* und y^* mit $x^*_j = \delta_{jj_0}$ und $y^*_i = \delta_{ii_0}$ Lösungen.

Beweis: Nach Lemma 4.8 ist

$$\begin{aligned} \min_{\substack{x\geq 0 \\ x^\top e=1}} \max_{\substack{y\geq 0 \\ y^\top e=1}} y^\top A x &= \min_{\substack{x\geq 0 \\ x^\top e=1}} \max_{i=1,\ldots,m} (Ax)_i \leq \min_{j=1,\ldots,n} \max_{i=1,\ldots,m} a_{ij} \\ &= \max_{i=1,\ldots,m} \min_{j=1,\ldots,n} a_{ij} \leq \max_{\substack{y\geq 0 \\ y^\top e=1}} \min_{j=1,\ldots,n} (A^\top y)_j \\ &= \max_{\substack{y\geq 0 \\ y^\top e=1}} \min_{\substack{x\geq 0 \\ x^\top e=1}} y^\top A x \,. \end{aligned}$$

Der Hauptsatz impliziert überall Gleichheit. Also ist der Optimalwert

$$v = \max_{i=1,\ldots,m} \min_{j=1,\ldots,n} a_{ij} = \min_{j=1,\ldots,n} \max_{i=1,\ldots,m} a_{ij} = a_{i_0 j_0} \,.$$

Schließlich sind x^* und y^* optimal wegen $a_{i_0 j_0} = y^{*\top} A x^*$. □

5 Das Simplexverfahren für lineare Optimierungsaufgaben

Das Simplexverfahren ist zugeschnitten für Probleme in der zweiten **Normalform**, d.h.

$$(P) \qquad \text{Minimiere} \quad f(x) := c^\top x \quad \text{auf} \quad M \;=\; \{x \in \mathbb{R}^n : Ax = b\,,\; x \geq 0\}.$$

Hierbei sind $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$ mit $m \leq n$ gegeben.

Bemerkungen: (i) Jede lineare Optimierungsaufgabe kann in eine äquivalente Form der Gestalt (P) gebracht werden (Einführung von Schlupfvariablen und Ersetzung von nichtvorzeichenbeschränkten Variablen x_j durch $u_j - v_j$ mit $u_j \geq 0$, $v_j \geq 0$).
(ii) Ist Rang $A < m$, so sind die Gleichungen „redundant“, und durch Weglassen geeigneter Gleichungen wird die Matrix auf vollen Rang gebracht. Vom theoretischen Gesichtspunkt her kann man also ohne Beschränkung Rang $A = m$ annehmen.

5.1 Das Gauß-Jordan Verfahren

Das Gleichungssystem $Ax = b$ ist zu lösen, wobei wieder $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$ gegeben sind und $x \in \mathbb{R}^n$ gesucht ist. Es sei $m \leq n$. Wir fordern **nicht** unbedingt, dass Rang $A = m$ und ignorieren auch die Vorzeichenbedingung $x \geq 0$..
Für $Ax = b$ führt man die **abkürzende Schreibweise** über **Gauß-Jordan Tableaus** ein:

$$\left.\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right| \begin{array}{c} b_1 \\ \vdots \\ b_m \end{array}$$

Das **Gauß-Jordan Verfahren** besteht aus:

(i) Pivotwahl: Hier kann jedes von 0 verschiedene Element links von der $b-$Spalte genommen werden.

(ii) „Leerräumen“ der pivotisierten Spalte (ober- und unterhalb des Pivotelements) und Normierung des Pivotelements auf 1.

(iii) neuer Pivotwahl unter allen Zeilen, die noch kein Pivotelement enthalten. Sind alle Zeilen pivotisiert oder bestehen nur aus Nullen, so STOP.

(iv) weiter mit (ii).

Beispiel 5.1 Bestimme die allgemeine Lösung des Gleichungssystems

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & + & x_3 & - & x_4 & = & 1 \\ 4x_1 & + & 5x_2 & + & 5x_3 & + & 2x_4 & = & 0 \\ x_1 & - & 2x_2 & & & - & 5x_4 & = & 5 \end{array}$$

Wir erhalten die Folge von Tableaus:

$$\left(\begin{array}{cccc|c} 1 & 1 & \boxed{1} & -1 & 1 \\ 4 & 5 & 5 & 2 & 0 \\ 1 & -2 & 0 & -5 & 5 \end{array}\right) \quad 5* \text{ 1. Gl. abziehen!}$$

$$\left(\begin{array}{cccc|c} 1 & 1 & \boxed{1} & -1 & 1 \\ \boxed{-1} & 0 & 0 & 7 & -5 \\ 1 & -2 & 0 & -5 & 5 \end{array}\right) \quad \begin{array}{l} \text{2. Gl. addieren!} \\ \\ \text{2. Gl. addieren!} \end{array}$$

$$\left(\begin{array}{cccc|c} 0 & 1 & \boxed{1} & 6 & -4 \\ \boxed{-1} & 0 & 0 & 7 & -5 \\ 0 & -2 & 0 & 2 & 0 \end{array}\right) \quad \begin{array}{l} \\ *(-1) \\ :(-2) \end{array}$$

$$\left(\begin{array}{cccc|c} 0 & 1 & \boxed{1} & 6 & -4 \\ \boxed{1} & 0 & 0 & -7 & 5 \\ 0 & \boxed{1} & 0 & -1 & 0 \end{array}\right) \quad \text{3. Zeile abziehen}$$

$$\left(\begin{array}{cccc|c} 0 & 0 & \boxed{1} & 7 & -4 \\ \boxed{1} & 0 & 0 & -7 & 5 \\ 0 & \boxed{1} & 0 & -1 & 0 \end{array}\right)$$

Jetzt sind wir fertig. Ausgeschrieben heißt das Gleichungssystem:

$$\begin{array}{rcrcl} & x_3 & + & 7x_4 & = & -4 \\ x_1 & & - & 7x_4 & = & 5 \\ & x_2 & - & x_4 & = & 0 \end{array}$$

Damit kann die allgemeine Lösung angegeben werden. Als Parameter nimmt man hier die Variable x_4. Diese heißt auch **freie** Variable im Gegensatz zu den **abhängigen** Variablen x_1, x_2 und x_3. Damit ist $x = (5 + 7t, t, -4 - 7t, t)^\top \in \mathbb{R}^4$, $t \in \mathbb{R}$, die allgemeine Lösung und

$$\dim \operatorname{Kern} A \;=\; 1 \quad \text{und} \quad \operatorname{Rang} A \;=\; 3\,.$$

Allgemein ist $r = \operatorname{Rang} A$ die Anzahl der Pivotelemente und $n - r$ die Anzahl der freien Variablen.

Interessant ist die spezielle Lösung, die man erhält, wenn man alle freie Variablen auf 0 setzt. Für unser Beispiel wäre dies: $x = (5, 0, -4, 0)^\top$. Wie erhält man sie „mechanisch“?

Die Variablen x_j, die zu den unpivotisierten Spalten des Tableaus gehören, setze man auf 0. In den anderen Variablen steht die rechte Seite b „geeignet“ permutiert.
Obwohl wir mit dem Gleichungssystem fertig sind, können wir trotzdem noch weiter machen und weitere „spezielle“ Lösungen suchen, etwa:

$$\left(\begin{array}{cccc|c} 0 & 0 & 1 & \boxed{7} & -4 \\ \boxed{1} & 0 & 0 & -7 & 5 \\ 0 & \boxed{1} & 0 & -1 & 0 \end{array}\right) \qquad \left(\begin{array}{cccc|c} 0 & 0 & 1/7 & \boxed{1} & -4/7 \\ \boxed{1} & 0 & 1 & 0 & 1 \\ 0 & \boxed{1} & 1/7 & 0 & -4/7 \end{array}\right)$$

Jetzt erhalten wir als spezielle Lösung $x_3 = 0$ und $x_1 = 1$, $x_2 = -4/7$, $x_4 = -4/7$.

Das Simplexverfahren (in der Phase II) sucht diese sogenannten **Basislösungen** ab!

5.2 Idee des Simplexverfahrens am speziellen Beispiel

Maximiere $20\,x_1 + 60\,x_2$ unter den Nebenbedingungen

$$\begin{pmatrix} 1 & 1 \\ 5 & 10 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 120 \\ 700 \\ 520 \end{pmatrix}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Zunächst transformieren wir dieses Problem auf die Normalform durch Einführung von Schlupfvariablen x_3, x_4, x_5. Außerdem schreiben wir die Maximierungsaufgabe in eine Minimierungsaufgabe um. Wir müssen also $-20\,x_1 - 60\,x_2$ mimimieren unter den Nebenbedingungen:

$$x \in \mathbb{R}^5\,, \quad x \geq 0\,, \quad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 5 & 10 & 0 & 1 & 0 \\ 2 & 10 & 0 & 0 & 1 \end{pmatrix} x \;=\; \begin{pmatrix} 120 \\ 700 \\ 520 \end{pmatrix}\,.$$

Wir sehen, dass das Gleichungssystem $Ax = b$ bereits auf Gauß-Jordan Endform gebracht worden ist. Es ist $\operatorname{Rang} A = 3$. Die Variablen x_1 und x_2 sind die freien Variablen, x_3, x_4, x_5 die abhängigen. Die zugehörige Basislösung ist
$\hat{z} = (0, 0, 120, 700, 520)^\top \in \mathbb{R}^5$.
Wir schreiben die Gleichung $c^\top x = \eta$ dazu (η ist ein Parameter), also

$$\left(\begin{array}{rrrrr|c} -20 & -60 & 0 & 0 & 0 & \eta \\ \hline 1 & 1 & 1 & 0 & 0 & 120 \\ 5 & 10 & 0 & 1 & 0 & 700 \\ 2 & \boxed{10} & 0 & 0 & 1 & 520 \end{array}\right) \tag{5.1}$$

Dieses Tableau heißt **Simplextableau**. Die Basislösung $\hat{z} = (0, 0, 120, 700, 520)^\top$ löst auch die erste Gleichung für den Parameter $\eta = 0$. Die Basislösung $\hat{z}$ ist zulässig, da $b \geq 0$!

Wir gehen jetzt zu einem anderen Tableau über, indem wir die markierte 10 als Pivotelement wählen:

$$\begin{array}{rrrrr|c} -8 & 0 & 0 & 0 & 6 & \eta + 3120 \\ \hline \frac{4}{5} & 0 & 1 & 0 & -\frac{1}{10} & 68 \\ 3 & 0 & 0 & 1 & -1 & 180 \\ \frac{1}{5} & 1 & 0 & 0 & \frac{1}{10} & 52 \end{array} \tag{5.2}$$

Jetzt sind x_1, x_5 die freien Variablen, x_2, x_3, x_4 die abhängigen. Die zugehörige Basislösung ist $\tilde{z} = (0, 52, 68, 180, 0)^\top$. Wir haben Glück gehabt, denn auch $\tilde{z} \geq 0$.
$\tilde{z}$ ist Lösung des gesamten Gleichungssystems (5.2) für $\eta + 3120 = 0$, also $\eta = -3120$. Da wir nur äquivalente Umformungen gemacht haben, ist $\tilde{z}$ auch Lösung von (5.1) für $\eta = -3120$. Daher hat $\tilde{z}$ einen kleineren Zielfunktionswert als $\hat{z}$, ist also „besser“!

Bemerkungen:

(A) Was passiert, wenn man im letzten Beispiel statt a_{32} jetzt etwa a_{12} als Pivotelement nimmt? Dann erhält man

$$\begin{array}{rrrrr|c} 40 & 0 & 60 & 0 & 0 & \eta + 60 \cdot 120 \\ \hline 1 & 1 & 1 & 0 & 0 & 120 \\ -5 & 0 & -10 & 1 & 0 & -500 \\ -8 & 0 & -10 & 0 & 1 & -680 \end{array}$$

mit der Basislösung $(0, 120, 0, -500, -680)^\top$. Diese ist nicht zulässig, da einige Komponenten negativ sind. Der Simplexschritt geht offensichtlich gut, falls das Pivotelement a_{rs} positiv und b_r/a_{rs} minimal ist.

(B) Wann bekommt man rechts oben etwas $> \eta$? Natürlich genau dann, wenn $c_s < 0$ und $b_r/a_{rs} > 0$.

Wir setzen das Beispiel fort und betrachten (5.2): Wir müssen das Pivotelement in der 1. Spalte suchen (da nur dort in der ersten Zeile ein negativer Eintrag steht):

$$\begin{array}{rrrrr|c} -8 & 0 & 0 & 0 & 6 & \eta + 3120 \\ \hline \frac{4}{5} & 0 & 1 & 0 & -\frac{1}{10} & 68 \\ \boxed{3} & 0 & 0 & 1 & -1 & 180 \\ \frac{1}{5} & 1 & 0 & 0 & \frac{1}{10} & 52 \end{array}$$

$$\begin{array}{rrrrr|c} 0 & 0 & 0 & \frac{8}{3} & \frac{10}{3} & \eta + 3600 \\ \hline 0 & 0 & 1 & -\frac{4}{15} & \frac{1}{6} & 20 \\ 1 & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 60 \\ 0 & 1 & 0 & -\frac{1}{15} & \frac{1}{6} & 40 \end{array}$$

Die zugehörige Basislösung ist $x^* = (60, 40, 20, 0, 0)^\top \in M$. Diese löst auch $\hat{c}^\top x^* = \eta$ für $\eta + 3600 = 0$, d.h. $\eta = -3600$. Hier ist $\hat{c}$ der „aktuelle Kostenvektor“, d.h. die erste Zeile im aktuellen Tableau. Daher löst x^* auch das Ausgangssystem (5.1) für $\eta = -3600$. Jetzt sind wir fertig, da die aktuelle Kostenzeile nichtnegativ ist. Jeder weitere Simplexschritt würde η nur wieder vergrößern. Wir können jetzt die Schlupfvariablen x_3, x_4, x_5 wieder vergessen und haben einen optimalen Vektor $(60, 40)^\top \in \mathbb{R}^2$ als Lösung des Ausgangsproblems gefunden.

Damit ist der zweite Teil (Phase II) des Simplexverfahrens praktisch schon vollständig beschrieben worden. Phase I beschäftigt sich mit dem Problem, ein beliebiges Problem in Normalform in die Form zu bringen, mit der Phase II begonnen werden kann. In der praktischen Rechnung führt man η nicht mit und auch die Einheitsmatrix nicht. Für unsere „Handrechnungen“ ist es jedoch viel übersichtlicher!

5.3 Das Simplexverfahren

Wir benötigen zur Beschreibung einige Bezeichnungen und Definitionen. Ist $A \in \mathbb{R}^{m\times n}$ eine $(m \times n)-$Matrix, so bezeichnen wir wieder mit $a_{*j} \in \mathbb{R}^m$ die $j-$te Spalte der Matrix (als Spaltenvektor) für $j = 1, \ldots, n$.

Definition 5.2 *Sei $A \in \mathbb{R}^{m\times n}$, $\operatorname{Rang} A = m \le n$, $b \in \mathbb{R}^m$. Ein Paar $(\hat{x}, \hat{\jmath})$ mit $\hat{x} \in \mathbb{R}^n$, $\hat{\jmath} = (j_1, \ldots, j_m) \in \{1, \ldots, n\}^m$ heißt* **Basislösung** *zu $Ax = b$, falls $A\hat{x} = b$ und die Spalten $\{a_{*j_k} : k = 1, \ldots, m\}$ linear unabhängig sind, sowie $\hat{x}_\ell = 0$ für alle $\ell \notin \hat{J}$.*

Hier haben wir mit dem großen Buchstaben $\hat{J}$ die (ungeordnete) Menge der Indizes bezeichnet, also $\hat{J} = \{j_1, \ldots, j_m\}$. Diese Konvention werden wir auch in Zukunft beibehalten.

Eine Basislösung heißt **zulässig**, *falls $\hat{x} \ge 0$ gilt. Sie heißt* **nicht entartet**, *wenn $\hat{x}_{j_k} > 0$ für alle $k = 1, \ldots, m$ gilt (dann ist also $\hat{J} = \{\ell \in \{1, \ldots, n\} : \hat{x}_\ell > 0\}$).*

Für jede zulässige Basislösung $(\hat{x}, \hat{\jmath})$ ist $\hat{x}$ Ecke von M. Dies folgt aus folgendem Satz:

Satz 5.3 *Sei $M = \{x \in \mathbb{R}^n : Ax = b,\ x \ge 0\}$ und sei für $x \in M$ die Menge der aktiven Indizes definiert durch $j(x) = \{j \in \{1, \ldots, n\} : x_j > 0\}$. Es ist $x \in M$ genau eine Ecke von M, wenn die Menge $\{a_{*j} : j \in J(x)\}$ der Spalten zu den aktiven Indizes linear unabhängig sind.*

Beweis: (a) Sei $\{a_{*j} : j \in J(x)\}$ linear unabhängig und $x = \lambda y + (1 - \lambda)z$ mit $\lambda \in (0, 1)$ und $y, z \in M$.
Für $j \notin J(x)$ ist $0 = \lambda y_j + (1-\lambda)z_j \ge 0$, also $y_j = z_j = 0 = x_j$. Daher ist $\sum_{j\in J(x)} x_j\, a_{*j} = b$ und genauso $\sum_{j\in J(x)} y_j\, a_{*j} = b = \sum_{j\in J(x)} z_j\, a_{*j}$. Aus der linearen Unabhängigkeit folgt hieraus $y = z = x$. Dies bedeutet, dass x eine Ecke ist.

(b) Sei umgekehrt x eine Ecke und $z_j \in \mathbb{R}$, $j \in J(x)$, mit $\sum_{j\in J(x)} z_j\, a_{*j} = 0$. Wir ergänzen z zu einem Vektor $z \in \mathbb{R}^n$, indem wir die Komponenten $z_j = 0$ setzen für $z \notin J(x)$. Dann gilt $Az = 0$. Schreibe x als $x = \frac{1}{2}(x - \varepsilon z) + \frac{1}{2}(x + \varepsilon z)$ und wähle $\varepsilon > 0$ so klein, dass $(x \pm \varepsilon z)_j > 0$ für alle $j \in J(x)$. Dann ist $x \pm \varepsilon z \ge 0$ und $A(x \pm \varepsilon z) = Ax = b$, also $x \pm \varepsilon z \in M$. Da x eine Ecke ist, so folgt $x \pm \varepsilon z = x$, d.h. $z = 0$. Daher sind die Spalten $\{a_{*j} : j \in J(x)\}$ linear unabhängig. □

Wegen $J(\hat{x}) \subseteq \hat{J}$ sind insbesondere die Spalten $\{a_{*j} : j \in J(x)\}$ linear unabhängig. Dies bedeutet nach diesem Satz, dass $\hat{x}$ eine Ecke ist.

Wir betrachten wieder das Optimierungsproblem:

$$(P) \qquad \text{Minimiere} \quad f(x) = c^\top x \quad \text{auf} \quad M := \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}\,.$$

Das Simplexverfahren besteht aus zwei Phasen:

Phase I: Man konstruiert eine zulässige Basislösung $\left(\hat{z}, \hat{\jmath}\right)$ zu $Ax = b$, einen Vektor $\hat{c} \in \mathbb{R}^n$ und eine Darstellung $M = \{x \in \mathbb{R}^n : \hat{A}x = \hat{b},\ x \geq 0\}$ von M mit folgenden Eigenschaften:

(a) Ist $\hat{\jmath} = (j_1, \ldots, j_m)$, so ist $\hat{a}_{*j_k} = e_{*k}$ für $k = 1, \ldots, m$, wobei e_{*k} der $k-$te Einheitsvektor im $\mathbb{R}^m$ ist,

(b) $\hat{c}_{j_k} = 0$ für alle $k = 1, \ldots, m$ (also auch $\hat{c}^\top \hat{z} = 0$), und $\hat{b} \geq 0$,

(c) $f(x) = \hat{c}^\top x + f(\hat{z})$ für alle x mit $Ax = b$. Ferner sei $\hat{\gamma} = f(\hat{z})$.

Phase II: Diese setzt voraus, dass eine zulässige Basislösung $(\hat{z}, \hat{\jmath})$ zu $Ax = b$, ein Vektor $\hat{c} \in \mathbb{R}^n$ und obige Darstellung von M gefunden wurde, so dass (a), (b), (c) gelten.
Es wird versucht, eine andere zulässige Basislösung $(\tilde{z}, \tilde{\jmath})$ zu $Ax = b$, einen Vektor $\tilde{c}$ und $\tilde{A}$, $\tilde{b}$ zu finden mit den Eigenschaften (a),(b),(c), so dass $f(\tilde{z}) < f(\hat{z})$ gilt.
Dann ersetzt man $\hat{z}$ u.s.w. durch $\tilde{z}$ u.s.w. und hofft, dass eine so gefundene Folge $\left(\hat{z}^k\right)$ von Basislösungen gegen eine Optimallösung konvergiert – oder sogar abgebrochen wird mit einer optimalen Lösung oder der Information, dass $\inf(P) = -\infty$.

Das Flussdiagramm auf der folgenden Seite beschreibt die Phase II des Simplexverfahrens. Der folgende Satz liefert die theoretische Grundlage für die Durchführung.

Satz 5.4 *Sei (P), $\hat{A}$, $\hat{b}$, $\hat{c}$, $\hat{\gamma}$ sowie eine Basislösung $\left(\hat{z}, \hat{\jmath}\right)$ mit (a), (b), (c) gegeben. $\tilde{A}$, $\tilde{b}$, $\tilde{c}$, $\tilde{\gamma}$ sowie $\left(\tilde{z}, \tilde{\jmath}\right)$ seien wie im Flussdiagramm konstruiert. Dann gilt:*

(i) Ist $\hat{c} \geq 0$, so ist $c^\top x \geq c^\top \hat{z}$ für alle $x \in M$, d.h. $\hat{z}$ ist optimal.

(ii) Ist $\hat{c}_s < 0$ und $\hat{a}_{*s} \leq 0$, so ist $c^\top x$ auf M nicht beschränkt, d.h. $\inf(P) = -\infty$.

(iii) Für jedes x gilt: $\tilde{A}x = \tilde{b} \Longleftrightarrow \hat{A}x = \hat{b} \Longleftrightarrow Ax = b$.

(iv) $\hat{c}^\top x + \hat{\gamma} = \tilde{c}^\top x + \tilde{\gamma}$ für alle $x \in \mathbb{R}^n$ mit $Ax = b$.

(v) $\left(\tilde{z}, \tilde{\jmath}\right)$ ist zulässige Basislösung zu $\tilde{A}x = \tilde{b}$, für die (a), (b), (c) erfüllt sind. Es ist $\tilde{a}_{rs} = 1$ und $\tilde{\gamma} = c^\top \tilde{z} = f(\tilde{z})$. Insbesondere ist (P) äquivalent zu

$$(\hat{P}) \qquad \textit{Minimiere} \quad \hat{c}^\top x + \hat{\gamma} \quad \textit{auf} \quad M = \{x \in \mathbb{R}^n : \hat{A}x = \hat{b}\,,\ x \geq 0\}\,,$$

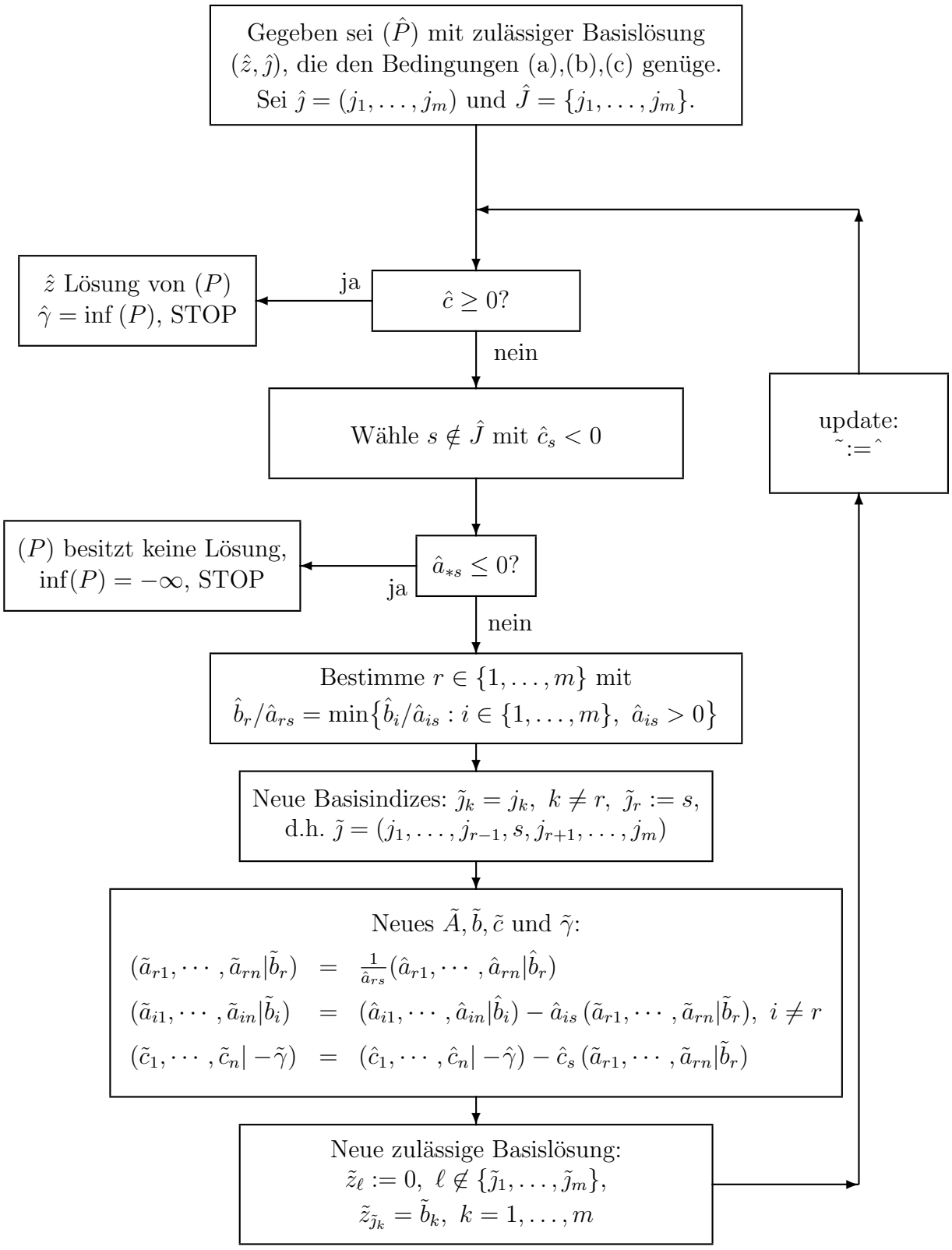
und dies äquivalent zu

$$(\tilde{P}) \qquad \textit{Minimiere} \quad \tilde{c}^\top x + \tilde{\gamma} \quad \textit{auf} \quad M = \{x \in \mathbb{R}^n : \tilde{A}x = \tilde{b}\,,\ x \geq 0\}\,.$$

mit gleichem Zielfunktionswert.

(vi) $f(\tilde{z}) = c^\top \tilde{z} = \tilde{\gamma} = \hat{c}_s \hat{b}_r / \hat{a}_{rs} + \hat{\gamma} \leq \hat{\gamma} = c^\top \hat{z} = f(\hat{z})$.

Phase II des Simplexverfahrens



Beweis: Sei wieder $\hat{\jmath} = (j_1, \ldots, j_m)$ und $\hat{J} = \{j_1, \ldots, j_m\}$.
(i) Sei $x \in M$ beliebig. Dann ist wegen (c) und (b):

$$c^\top x \;=\; f(x) \;=\; \hat{c}^\top x + f(\hat{z}) \;\geq\; f(\hat{z}) \;=\; c^\top \hat{z}\,.$$

(ii) Für $t \geq 0$ definiere man $x(t) \in \mathbb{R}^n$ durch

$$x_{j_k}(t) \;=\; \hat{b}_k - t\,\hat{a}_{ks}\;(k = 1, \ldots, m)\,, \quad x_s(t) \;=\; t\,, \quad x_\ell(t) \;=\; 0\;(\ell \notin \{j_1, \ldots, j_m, s\})\,.$$

Dann ist $x(t) \geq 0$ wegen $\hat{a}_{*s} \leq 0$. Ausserdem ist wegen $a_{*j_k} = e_{*k}$:

$$\hat{A}\,x(t) \;=\; \sum_{\ell=1}^{n} x_\ell(t)\,\hat{a}_{*\ell} \;=\; t\,\hat{a}_{*s} \;+\; \sum_{k=1}^{m}\big(\hat{b}_k - t\,\hat{a}_{ks}\big)\,e_{*k} \;=\; \hat{b}\,,$$

d.h. $x(t) \in M$ für jedes $t \geq 0$. Schließlich folgt wegen (b), (c) und $\hat{c}_s < 0$:

$$f\big(x(t)\big) \;=\; f(\hat{z}) + \hat{c}^\top x(t) \;=\; f(\hat{z}) \;+\; \sum_{k=1}^{m} \hat{c}_{j_k}\,x_{j_k}(t) \;+\; t\,\hat{c}_s \;=\; f(\hat{z}) \;+\; t\,\hat{c}_s \;\longrightarrow\; -\infty$$

für $t \to \infty$.
(iii) Dies ist trivial, da $\tilde{A}$ und $\tilde{b}$ aus $\hat{A}$ und $\hat{b}$ durch einen Gauss-Jordan Schritt entstehen.
(iv) Wir rechnen aus:

$$\tilde{c}^\top x \;+\; \tilde{\gamma} \;=\; \hat{c}^\top x \;+\; \hat{\gamma} \;-\; \hat{c}_s\left[\sum_{\ell=1}^{n} x_\ell\,\tilde{a}_{r\ell} - \tilde{b}_r\right] \;=\; \hat{c}^\top x \;+\; \hat{\gamma}\,, \quad \text{da } \tilde{A}x = \tilde{b}\,.$$

(v) Wegen $\tilde{\jmath} = \big(j_1, \ldots, j_{r-1}, s, j_{r+1}, \ldots, j_m\big)$ ist zu zeigen:

$$\tilde{a}_{*j_k} \;=\; e_{*k} \text{ für } k \neq r \quad \text{und} \quad \tilde{a}_{*s} \;=\; e_{*r}\,.$$

Sei zunächst $k \neq r$. Dann ist

$$\tilde{a}_{rj_k} \;=\; \frac{1}{\hat{a}_{rs}}\,\hat{a}_{rj_k} \;=\; 0$$

und

$$\tilde{a}_{ij_k} \;=\; \hat{a}_{ij_k} \;-\; \hat{a}_{is}\,\tilde{a}_{rj_k} \;=\; \hat{a}_{ij_k} \;=\; \delta_{ik} \quad \text{für} \quad i \neq r\,.$$

Also ist $\tilde{a}_{*j_k} = e_{*k}$ für $k \neq r$. Ferner ist $\tilde{a}_{rs} = 1$ und

$$\tilde{a}_{is} \;=\; \hat{a}_{is} \;-\; \hat{a}_{is}\,\tilde{a}_{rs} \;=\; 0 \quad \text{für} \quad i \neq r\,,$$

also $\tilde{a}_{*s} = e_{*r}$. Analog ist $\tilde{b}_r = \hat{b}_r/\hat{a}_{rs} \geq 0$, und

$$\tilde{b}_i \;=\; \hat{b}_i \;-\; \frac{\hat{a}_{is}}{\hat{a}_{rs}}\,\hat{b}_r \;=\; \hat{a}_{is}\left(\frac{\hat{b}_i}{\hat{a}_{is}} - \frac{\hat{b}_r}{\hat{a}_{rs}}\right) \geq 0\,, \quad i \neq r\,,$$

nach Wahl von r (da $\hat{a}_{is} > 0$). Also ist $(\tilde{z}, \tilde{\jmath})$ eine zulässige Basislösung, die (a) erfüllt.
Für $\ell \in \hat{J} \setminus \{j_r\}$ ist $\hat{c}_\ell = 0$ und $\tilde{a}_{r\ell} = 0$, also $\tilde{c}_\ell = 0$. Für $\ell = s$ ist $\tilde{c}_s = \hat{c}_s - \hat{c}_s\tilde{a}_{rs} = 0$, da $\tilde{a}_{rs} = 1$. Also ist auch Voraussetzung (b) erfüllt.

Um (c) für $\tilde{z}$ und $\tilde{\gamma}$ zu zeigen, benutzen wir (c) für $\hat{z}$ und $\hat{\gamma}$ und Teil (iv). Für $x = \tilde{z}$ folgt

$$f(\tilde{z}) \;=\; \hat{c}^\top \tilde{z} \;+\; f(\hat{z}) \;=\; \hat{c}^\top \tilde{z} \;+\; \hat{\gamma} \;=\; \tilde{c}^\top \tilde{z} \;+\; \tilde{\gamma} \;=\; \tilde{\gamma}\,.$$

Mit (iv) folgt dann für $x \in \mathbb{R}^n$ mit $Ax = b$

$$f(x) \;=\; \hat{c}^\top x \;+\; f(\hat{z}) \;=\; \hat{c}^\top x \;+\; \hat{\gamma} \;=\; \tilde{c}^\top x \;+\; \tilde{\gamma} \;=\; \tilde{c}^\top x \;+\; f(\tilde{z})\,.$$

(vi) Schließlich ist

$$\tilde{\gamma} \;=\; \hat{\gamma} \;+\; \hat{c}_s\,\tilde{b}_r \;=\; \hat{\gamma} \;+\; \hat{c}_s\,\frac{\hat{b}_r}{\hat{a}_{rs}} \;\leq\; \hat{\gamma}\,,$$

da $\hat{b}_r \geq 0$, $\hat{a}_{rs} > 0$ und $\hat{c}_s < 0$. Damit ist alles bewiesen! □

Beispiel 5.5 Maximiere $x_1+2x_2+4x_3$ unter den Nebenbedingungen $0 \leq x_1 \leq 2$, $x_2, x_3 \geq 0$, und

$$\begin{aligned} x_1 \;+\; x_2 \;+\; 2x_3 \;&\leq\; 4 \\ 3x_2 \;+\; 4x_3 \;&\leq\; 6 \end{aligned}$$

Nach Umschreiben auf eine Minimierungsaufgabe und Einführung von Schlupfvariablen x_4, x_5, x_6 erhalten wir die folgenden Tableaus. Hier haben wir rechts oben jeweils nur z.B. 0 statt wie früher $\gamma + 0$ geschrieben. Dort steht im Tableau also das Negative vom Wert der aktuellen Basislösung, also $-\hat{\gamma}$.

-1	-2	-4	0	0	0	0
$\boxed{1}$	0	0	1	0	0	2
1	1	2	0	1	0	4
0	3	4	0	0	1	6

$$\hat{z} \;=\; (0,0,0,2,4,6)^\top \in \mathbb{R}^6\,, \quad \hat{\gamma} = 0\,.$$

0	-2	-4	1	0	0	2
1	0	0	1	0	0	2
0	$\boxed{1}$	2	-1	1	0	2
0	3	4	0	0	1	6

$$\hat{z} \;=\; (2,0,0,0,2,6)^\top\,, \quad \hat{\gamma} = -2\,,$$

0	0	0	-1	2	0	6
1	0	0	1	0	0	2
0	1	2	-1	1	0	2
0	0	-2	$\boxed{3}$	-3	1	0

$$\hat{z} \;=\; (2,2,0,0,0,0)^\top\,, \quad \hat{\gamma} = -6\,, \quad \hat{\jmath} = (1,2,6)\,.$$

Dies ist eine **entartete Ecke**, da $\hat{b}_3 = 0$. Trotzdem geht es weiter:

$$\begin{array}{cccccc|c} 0 & 0 & -\frac{2}{3} & 0 & 1 & \frac{1}{3} & 6 \\ \hline 1 & 0 & \frac{2}{3} & 0 & 1 & -\frac{1}{3} & 2 \\ 0 & 1 & \boxed{\frac{4}{3}} & 0 & 0 & \frac{1}{3} & 2 \\ 0 & 0 & -\frac{2}{3} & 1 & -1 & \frac{1}{3} & 0 \end{array} \qquad \hat{z} \;=\; (2,2,0,0,0,0)^\top\,, \quad \hat{\jmath} = (1,2,4)\,.$$

Hier erhalten wir das gleiche $\hat{z}$, verkleinern also unseren Zielfunktionswert nicht, haben aber jetzt ein anderes $\hat{\jmath}$, nämlich $\hat{\jmath} = (1,2,4)$. Wir machen weiter:

$$\begin{array}{cccccc|c} 0 & \frac{1}{2} & 0 & 0 & 1 & \frac{1}{2} & 7 \\ \hline 1 & -\frac{1}{2} & 0 & 0 & 1 & -\frac{1}{2} & 1 \\ 0 & \frac{3}{4} & 1 & 0 & 0 & \frac{1}{4} & \frac{3}{2} \\ 0 & \frac{1}{2} & 0 & 1 & -1 & \frac{1}{2} & 1 \end{array} \qquad \hat{z} \;=\; (1,0,3/2,1,0,0)^\top\,, \quad \hat{\gamma} = -7\,.$$

Damit sind wir fertig, da alle $\hat{c}_j \geq 0$. Eine Lösung des Ausgangsproblems ist $x^* = (1,0,3/2)^\top$ mit Optimalwert 7.

Bemerkungen:

(a) Probleme der Form

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

mit $b \geq 0$ können sofort auf ein Starttableau gebracht werden durch Einführung von Schlupfvariablen.

(b) Ist $(\hat{z}, \hat{\jmath})$ nicht entartet, d.h. $\hat{z}_\ell > 0$ für alle $\ell \in \hat{J}$, so ist $\hat{b} > 0$ und daher $f(\tilde{z}) < f(\hat{z})$, d.h. es tritt eine echte Zielfunktionsverkleinerung ein, und die Basislösung $(\hat{z}, \hat{\jmath})$ kann in den folgenden Schritten nicht mehr erreicht werden. Ist keine durch das Simplexverfahren erhaltene Basislösung entartet, so muss daher das Simplexverfahren nach endlich vielen Schritten abbrechen, da es insgesamt nur endlich viele verschiedene Simplextableaus zu (P) gibt. (Jedes Simplextableau enthält die m Koordinateneinheitsvektoren des $\mathbb{R}^m$ unter den n Spalten. Für deren Stellen gibt es nur endlich viele Möglichkeiten, selbst nach Permutation der Stellen. Außerdem sind zwei Tableaus gleich, wenn die Stellen der Koordinatenvektoren übereinstimmen.) Das Simplexverfahren bricht ab mit einer Lösung oder mit der Information, dass keine solche existiert.

(c) Ist $(\hat{z}, \hat{\jmath})$ entartet, so ist $\hat{b}_r = 0$ und daher $\tilde{b} = \hat{b}$, also $\tilde{z} = \hat{z}$, aber $\tilde{\jmath} \neq \hat{\jmath}$. Bei einem Schritt des Simplexverfahrens bleibt man also in ein und derselben Lösung stehen und verändert nur die Basisdarstellung. Es kann vorkommen (allerdings nur bei konstruierten Beispielen, siehe Beispiel 5.6 unten), dass man einige Schritte lang nur die Basisdarstellung verändert und schließlich wieder zu derselben Darstellung zurückkommt. Dann spricht man von einem **Zyklus**. Ohne eine Zusatzregel, die Zyklen vermeidet, kann also nicht die Endlichkeit des Verfahrens bewiesen werden.

Das folgende Beispiel stammt von Beale (1955):

Beispiel 5.6 Minimiere $-\frac{3}{4}x_1 + 20x_2 - \frac{1}{2}x_3 + 6x_4$ unter $x_j \geq 0$, $j = 1, 2, 3, 4$ und

$$\begin{array}{rcrcrcrcl} \frac{1}{4}x_1 & - & 8x_2 & - & x_3 & + & 9x_4 & \leq & 0\,, \\ \frac{1}{2}x_1 & - & 12x_2 & - & \frac{1}{2}x_3 & + & 3x_4 & \leq & 0\,, \\ & & & & x_3 & & & \leq & 1\,. \end{array}$$

Als Auswahlregel wird vereinbart:
Pivotspalte sei die Spalte j mit minimalem $\hat{c}_j$ (und davon die mit minimalem j)
Pivotzeile sei die mit kleinstem Quotienten und, wenn es eine Auswahl gibt, die mit kleinster Zeilenzahl i.
Damit ergibt sich ein Zyklus (siehe Übung, das 7. Tableau stimmt mit dem ersten überein).

Die obige Auswahlregel ist schlecht, wie man sieht. Es gibt mehrere bessere, die Zyklen vermeiden. Eine benutzt die **lexikographische Ordnung** von Vektoren (siehe Collatz-Wetterling, S. 21, oder Werner, Numerische Mathematik 2), eine elegantere stammt von Bland (New finite pivoting rules for the simplex method, Math. Oper. Res. 2 (1977), 103–107).

Pivotregel von Bland:

(a) Pivotspalte ist die Spalte s mit $\hat{c}_s < 0$ und, wenn davon es mehrere gibt, die mit kleinster Spaltennummer, also

$$s \;=\; \min\bigl\{j : \hat{c}_j < 0\bigr\}\,.$$

(b) Pivotzeile ist die Zeile r mit kleinstem Quotienten $\hat{b}_i/\hat{a}_{is}$ und, falls es davon mehrere gibt, die mit kleinstem j_i. Hier ist wieder $\hat{\jmath} = (j_1, \ldots, j_m)$. Also ist genauer mit $I^+ := \{i : \hat{a}_{is} > 0\}$:

$$r \in I^+ \quad \text{und} \quad j_r \;=\; \min\left\{ j_k : k \in I^+\,,\ \frac{\hat{b}_k}{\hat{a}_{ks}} \leq \frac{\hat{b}_i}{\hat{a}_{is}} \text{ für alle } i \in I^+ \right\}.$$

Satz 5.7 *Das Simplexverfahren mit der Pivotregel von Bland stoppt immer und wiederholt kein Tableau.*

Beweis: Wir halten uns an das Buch: C.H. Papadimitriou, K. Steiglitz, Combinatorial Optimization. Prentice Hall, 1982, S. 53. Angenommen, es gäbe einen Zyklus

$$\hat{A}^{(\ell)},\ \hat{b}^{(\ell)},\ \hat{c}^{(\ell)},\ \hat{\gamma}^{(\ell)},\ \hat{z}^{(\ell)},\ \hat{\jmath}^{(\ell)}, \quad \ell = 1, \ldots, L\,.$$

Dann haben wir schon gesehen, dass $\hat{\gamma}^{(\ell)} = \hat{\gamma}$, $\hat{b}^{(\ell)} = \hat{b}$, $\hat{z}^{(\ell)} = \hat{z}$ konstant sind und $\hat{b}_r = 0$ für alle auftretenden Pivotzeilen r. Wir streichen jetzt alle Reihen und Spalten, die während des Zyklus kein Pivotelement enthalten. Dann bekommen wir ein neues Optimierungsproblem, das ebenfalls einen Zyklus hat. Außerdem ist dort $\hat{b}^{(\ell)} = 0$, $\ell = 1, \ldots, L$. Gewisse Variablen x_j kommen während der Zyklen in die Basislösung hinein, andere hinaus und manche bleiben vielleicht die ganze Zeit drin. Sei x_q die Variable mit

dem größten Index, die während der Zyklen hineinkommt. Irgendwann muss sie wieder hinaus: T_1 und T_2 seien die Tableaus direkt vor dem Reinkommen und Rausgehen. x_p komme für x_q hinein. Dann ist $p < q$ nach Definition von q.

Tableau T_1:

$\big(\hat{z}, \hat{\jmath}\big)$, $\hat{A}$, $\hat{c}$

$q \notin \hat{J}$, kommt aber im nächsten Tableau in die Basislösung.

	q rein $\downarrow$		
$\cdots \geq 0 \cdots$	$\hat{c}_q < 0$	$\cdots 0 \cdots$	$-\hat{\gamma}$
			0
			$\vdots$
			0

Tableau T_2:

$\big(\hat{z}, \tilde{\jmath}\big)$, $\tilde{A}$, $\tilde{c}$

$p \notin \tilde{J}$, $q \in \tilde{J}$

p kommt im nächsten Tableau für q hinein.

	p rein $\downarrow$		q raus $\uparrow$			
$\cdots$	$\tilde{c}_p < 0$	$\cdots$	0	$\cdots 0 \cdots$	$-\hat{\gamma}$	
					$\vdots$	
$\cdots$	$\tilde{a}_{rp} > 0$	$\cdots$	$\boxed{1}$	$\cdots$	0	$\leftarrow r$
					$\vdots$	
					0	

Es muss $p < q$ sein nach Wahl von x_q. Ausserdem ist wegen der Wahl der Pivotspalte q:

$$\hat{c}_\ell \;=\; \begin{cases} \geq 0 & \text{für } \ell < q, \\ < 0 & \text{für } \ell = q, \\ = 0 & \text{für } \ell > q. \end{cases}$$

(Für $\ell > q$ bleiben die x_ℓ während des ganzen Zyklus in der Basis, und daher verschwindet $\hat{c}_\ell$ für solche ℓ.)

Sei $\tilde{\jmath} = (j_1, \ldots, j_m)$. Setze

$$y_\ell \;:=\; \begin{cases} 1, & \text{falls } \ell = p, \\ -\tilde{a}_{kp}, & \text{falls } \ell = j_k\,,\ k = 1, \ldots, m, \\ 0, & \text{sonst.} \end{cases}$$

Dann ist $\tilde{A}y = 0 = \tilde{b}$, denn

$$\tilde{A}y \;=\; \sum_{\ell=1}^{n} y_\ell\, \tilde{a}_{*\ell} \;=\; \tilde{a}_{*p} \;-\; \sum_{k=1}^{m} \tilde{a}_{kp} \underbrace{\tilde{a}_{*j_k}}_{=\, e_{*k}} \;=\; \tilde{a}_{*p} - \tilde{a}_{*p} \;=\; 0\,,$$

und $\tilde{c}^\top y = \tilde{c}_p < 0$, da p die nächste Pivotspalte ist. Es ist nicht unbedingt $y \geq 0$. Trotzdem folgt mit (iv) von Satz 5.4 (beachte, dass $\hat{\gamma} = \tilde{\gamma}$):

$$\hat{c}^\top y \;=\; \tilde{c}^\top y \;=\; \tilde{c}_p < 0\,.$$

Nach Wahl der Pivotzeile ist

$$y_\ell \;=\; \begin{cases} -\tilde{a}_{rp} < 0 & \text{für } \ell = j_r = q, \\ 0, 1 \text{ oder } -\tilde{a}_{kp} \geq 0 & \text{für } \ell < q \end{cases}$$

(Wäre nämlich ein $\tilde{a}_{kp} > 0$ für ein $\ell = j_k < q = j_r$, so müsste man nach der Regel Zeile k nehmen!). Dies führt wegen

$$\hat{c}^\top y \;=\; \sum_{\ell<q} \hat{c}_\ell\, y_\ell \;+\; \hat{c}_q\, y_q \;>\; 0$$

zu einem Widerspruch. □

Wir haben gesehen, dass wir die beschriebene Phase II direkt auf Optimierungsprobleme von folgendem Typ anwenden können:

$$\text{Minimiere} \quad c^\top x \quad \text{unter} \quad Ax \leq b\,,\; x \geq 0\,,$$

wenn wir $b \geq 0$ haben. Sei aber jetzt wieder der allgemeine Fall gegeben:

$$(P) \qquad \text{Minimiere} \quad c^\top x \quad \text{auf} \quad M := \{x \in \mathbb{R}^n : Ax = b,\; x \geq 0\}\,.$$

Nach eventueller Multiplikation einer Zeile mit -1 können wir o.B.d.A. $b \geq 0$ voraussetzen.

Wir beschreiben jetzt die **Phase I** des Simplexverfahren. Es ist die Aufgabe, eine zulässige Basislösung von $Ax = b$ zu konstruieren, so dass damit die Phase II gestartet werden kann. Man betrachte dazu die Hilfsaufgabe:

$$(P_I) \qquad \text{Minimiere} \quad e^\top(b - Ax) \quad \text{unter den Nebenbedingungen} \quad Ax \leq b\,,\; x \geq 0\,,$$

wobei $e = (1, 1, \ldots, 1)^\top \in \mathbb{R}^m$.
Wir sehen: Wegen $e \geq 0$ und $b - Ax \geq 0$ für zulässige x ist $e^\top(b - Ax) \geq 0$, und $e^\top(b - Ax) = 0$ genau dann, wenn $Ax = b$. Außerdem ist (P_I) zulässig, da $x = 0$ ein zulässiger Vektor ist. (Beachte, dass $b \geq 0$ vorausgesetzt ist.)
Die Zielfunktion von (P_I) hat die Form

$$f(x) \;=\; e^\top(b - Ax) \;=\; -(A^\top e)^\top x \;+\; e^\top b \;=\; d^\top x \;+\; \alpha$$

mit $d := -A^\top e$ und $\alpha = e^\top b$. Das Minimieren von $d^\top x + \alpha$ ist natürlich äquivalent zur Minimierung von $d^\top x$ allein. Daher hat (P_I) genau die Form, auf die man Phase II direkt anwenden kann. Wegen $\inf(P_I) \geq 0$ muss das Simplexverfahren mit der Blandschen Regel mit einer optimalen Lösung von (P_I) abbrechen.

Wir starten daher mit dem Simplextableau, nehmen aber die Kostenzeile
$(c_1, c_2, \ldots, c_n, 0)$ für das ursprüngliche Problem (P) mit auf. Man beachte, dass $f(0) = \alpha = \sum_{i=1}^m b_i$ und $d_j = -\sum_{i=1}^m a_{ij}$. Daher hat die „Kostengleichung“ für das Hilfsproblem die

Form $d^\top x = \gamma - \alpha = \gamma - \sum_{i=1}^m b_i$. Wir starten also rechts oben mit $-\sum_{i=1}^m b_i$!

$-\sum_{i=1}^m a_{i1}$	$-\sum_{i=1}^m a_{i2}$	$\cdots$	$-\sum_{i=1}^m a_{in}$	0	$\cdots$	0	$-\sum_{i=1}^m b_i$
c_1	c_2	$\cdots$	c_n	0	$\cdots$	0	0
a_{11}	a_{12}	$\cdots$	a_{1n}	1	$\cdots$	0	b_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_{m1}	a_{m2}	$\cdots$	a_{mn}	0	$\cdots$	1	b_m

Da (P_I) lösbar ist, so liefert die oben beschriebene Phase II als Lösung eine zulässige Basislösung $(\hat{z}, \hat{\jmath})$. Von $\hat{z}$ sind die letzten m Komponenten Schlupfvariable. Dann sind drei Fälle möglich:

1. Fall: $\min(P_I) > 0$. Dann ist $A\hat{z} \neq b$, und $Ax = b$, $x \geq 0$, besitzt keine Lösung, d.h. $M = \emptyset$. Wir können also mit der Phase I testen, ob ein Gleichungssystem $Ax = b$ eine **nichtnegative Lösung** x besitzt.

2. Fall: $\min(P_I) = 0$ und $\hat{J} \subseteq \{1, \ldots, n\}$. Dann ist $A\hat{z} = b$, $\hat{z} \geq 0$, also $\hat{z} \in M$, und ist Basislösung von $Ax = b$ mit $\hat{a}_{*j_k} = e_{*k}$, $k = 1, \ldots, m$, falls $\hat{\jmath} = (j_1, \ldots, j_m)$. Streicht man die erste Zeile des Endtableaus von Phase I und die letzten m Spalten (vor der b-Spalte), so kann man gleich mit Phase II weitermachen. Dies ist der angenehme Fall!

3. Fall: $\min(P_I) = 0$, aber $\hat{J} \cap \{n+1, \ldots, n+m\} \neq \emptyset$. In diesem Fall sind Schlupfvariable in der Basislösung, etwa $\hat{z}_{j_s}$ mit $j_s > n$. In diesem Fall ist die Lösung sicher entartet, da wegen $\min(P_I) = 0$ die Schlupfvariable den Wert 0 haben müssen. Wir eliminieren diese Variable, indem wir sie mit einem $\ell \in \{1, \ldots, n\} \setminus \hat{J}$ mit $\hat{a}_{s\ell} \neq 0$ austauschen (Simplexschritt). Dadurch ändert sich die $b-$Spalte nicht. Findet man kein ℓ mit $\hat{a}_{s\ell} \neq 0$, so ist s-te Zeile Nullzeile und kann weggelassen werden. Dies kann aber nur passieren, wenn der Rang von A echt kleiner als m ist und kann daher zur Prüfung der Rangvoraussetzung dienen. Diese Austauschschritte werden solange durchgeführt, bis alle Schlupfvariablen aus der Basis eliminiert sind!

Beispiele 5.8 (A) Minimiere $3x_1 + x_2 - x_3$ unter den Nebenbedingungen

$$\begin{aligned} -5x_1 + 2x_2 - 3x_3 &= 2\,, \\ x_1 + x_2 + 2x_3 &= 1\,, \end{aligned} \qquad x_j \geq 0\,,\ j = 1, 2, 3\,.$$

Tableaus für Phase I:

4	-3	1	0	0	-3
3	1	-1	0	0	0
-5	2	-3	1	0	2
1	$\boxed{1}$	2	0	1	1

7	0	7	0	3	0
2	0	-3	0	-1	-1
-7	0	-7	1	-2	0
1	1	2	0	1	1

(Nach der Blandschen Regel müssten wir im ersten Tableau eigentlich $\hat{a}_{12} = 2$ als Pivotelement nehmen. Mit $\hat{a}_{22} = 1$ lässt sich aber besser rechnen.) Damit ist die Phase I abgeschlossen, $\hat{z} = (0,1,0,0,0)^\top$, $\hat{\jmath} = (4,2)$, also $\hat{J} \not\subseteq \{1,2,3\}$. Daher liegt der 3. Fall vor, und wir müssen $\hat{z}_4$ eliminieren. Wir wählen $\hat{a}_{11} = -7$ als Pivotelement und erhalten:

0	0	0	1	1	0
0	0	−5	2/7	−11/7	−1
1	0	1	−1/7	2/7	0
0	1	1	1/7	5/7	1

Jetzt ist $\hat{z} = (0,1,0,0,0)^\top$, $\hat{J} = \{1,2\} \subseteq \{1,2,3\}$.

Damit erhalten wir das Starttableau für Phase II:

0	0	−5	−1
1	0	$\boxed{1}$	0
0	1	1	1

und dann

5	0	0	−1
1	0	1	0
−1	1	0	1

$x^* = (0,1,0)^\top$ ist optimal mit Minimalwert 1.

(B) Minimiere $-x_1 + 2x_2 + 3x_3$ unter den Nebenbedingungen

$$(P) \qquad \begin{aligned} x_1 + 3x_2 - 2x_3 &\le 10, \\ 3x_1 - 2x_2 - 3x_3 &\le -5, \qquad x_1 \ge 0, \quad x_2 \ge 0. \\ 4x_1 - 2x_2 &= 4, \end{aligned}$$

Um dieses Problem auf Normalform zu transformieren, ersetzen wir x_3 durch $x_3 - x_4$ mit $x_3 \ge 0$, $x_4 \ge 0$ und führen die Schlupfvariablen x_5, x_6 für die beiden Ungleichungen ein. Multiplikation der 2. Gleichung mit -1 liefert die Aufgabe:

Minimiere $-x_1 + 2x_2 + 3x_3 - 3x_4$ unter den Nebenbedingungen

$$(P_N) \qquad \begin{aligned} x_1 + 3x_2 - 2x_3 + 2x_4 + x_5 &= 10 \\ -3x_1 + 2x_2 + 3x_3 - 3x_4 - x_6 &= 5 \\ 4x_1 - 2x_2 &= 4 \end{aligned}$$

Das Tableau für Phase II sieht so aus (Versuch!):

−1	2	3	−3	0	0	0
1	3	−2	2	1	0	10
−3	2	3	−3	0	−1	5
4	−2	0	0	0	0	4

Den ersten Einheitsvektor haben wir in der 5. Spalte. Er gehört zur ersten Ungleichung. Die beiden anderen Gleichungen liefern uns keine Einheitsvektoren. Daher muss auf diese die Phase I angewandt werden. Es seien x_7 und x_8 die Schlupfvariablen in Phase I für diese beiden Gleichungen. Wir erhalten die Tableaus:

$$\begin{array}{rrrrrrr:rr|r}
-1 & 0 & -3 & 3 & 0 & 1 & 0 & 0 & -9 \\
\hline
-1 & 2 & 3 & -3 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 3 & -2 & 2 & 1 & 0 & 0 & 0 & 10 \\
-3 & 2 & \boxed{3} & -3 & 0 & -1 & 1 & 0 & 5 \\
4 & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 4
\end{array}$$

$$\begin{array}{rrrrrrr:rr|r}
-4 & 2 & 0 & 0 & 0 & 0 & 1 & 0 & -4 \\
\hline
2 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & -5 \\
\hline
-1 & 13/3 & 0 & 0 & 1 & -2/3 & 2/3 & 0 & 40/3 \\
-1 & 2/3 & 1 & -1 & 0 & -1/3 & 1/3 & 0 & 5/3 \\
\boxed{4} & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 4
\end{array}$$

$$\begin{array}{rrrrrrr:rr|r}
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
\hline
0 & 1 & 0 & 0 & 0 & 1 & -1 & -1/2 & -7 \\
\hline
0 & 23/6 & 0 & 0 & 1 & -2/3 & 2/3 & 1/4 & 43/3 \\
0 & 1/6 & 1 & -1 & 0 & -1/3 & 1/3 & 1/4 & 8/3 \\
1 & -1/2 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1
\end{array}$$

und $\hat{z} = (1, 0, 8/3, 0, 43/3, 0, 0, 0)^\top \in \mathbb{R}^8$ und $\hat{j} = (5, 3, 1)$.

Damit ist die Phase I abgeschlossen, und es liegt der 2. Fall vor. Wir erhalten das Starttableu für Phase II durch Streichen der erste Zeile und der letzten zwei Spalten vor der $b-$Spalte (nur diese sind die Schlupfvariablen für Phase I):

$$\begin{array}{rrrrrr|r}
0 & 1 & 0 & 0 & 0 & 1 & -7 \\
\hline
0 & 23/6 & 0 & 0 & 1 & -2/3 & 43/3 \\
0 & 1/6 & 1 & -1 & 0 & -1/3 & 8/3 \\
1 & -1/2 & 0 & 0 & 0 & 0 & 1
\end{array}$$

und $\hat{z} = (1, 0, 8/3, 0, 43/3, 0)^\top \in \mathbb{R}^6$. Diese Basislösung $\hat{z}$ ist schon optimal, da die Kostenzeile nichtnegativ ist. Damit ist $(1, 0, 8/3, 0, 43/3, 0)^\top \in \mathbb{R}^6$ Lösung von (P_N), und daher ist $x^* = (1, 0, 8/3)^\top \in \mathbb{R}^3$ Lösung von (P) mit Minimalwert $f(x^*) = 7$.

Abschließend wollen wir noch kurz auf das **revidierte Simplexverfahren** eingehen, das meistens programmiert wird. Dazu zeigen wir zunächst wichtige Beziehungen zwischen den aktuellen Tableaus und den Ausgangsdaten des Problems.
Wir wollen eine praktische Notation einführen. Sei $m \leq n$. Für jeden Multiindex $\jmath = (j_1, \ldots, j_m) \in \{1, \ldots, n\}^m$, jeden Vektor $x \in \mathbb{R}^n$ und jede Matrix $A \in \mathbb{R}^{m \times n}$ definieren wir den „Untervektor“ $x(\jmath) \in \mathbb{R}^m$ und die Untermatrix $A(\jmath) \in \mathbb{R}^{m \times m}$ durch

$$x(\jmath) \;:=\; \big(x_{j_1}, \ldots, x_{j_m}\big)^\top \quad \text{und} \quad A(\jmath) \;:=\; \big[a_{*j_1}, \ldots, a_{*j_m}\big]\,.$$

Ist $k \in \{1, \ldots, n\}^{n-m}$ ein zweiter Multiindex mit $\{j_1, \ldots, j_m\} \cup \{k_1, \ldots, k_{n-m}\} = \{1, \ldots, n\}$, so sind alle j_i und k_ℓ paarweise verschieden und

$$c^\top x \;=\; c(\jmath)^\top x(\jmath) \;+\; c(k)^\top x(k) \quad \text{sowie} \quad Ax \;=\; A(\jmath)x(\jmath) \;+\; A(k)x(k)\,.$$

$\big(x(\jmath), x(k)\big)$ enthalten also alle Komponenten von x, nur permutiert. Wir nennen ein solches k ein Komplement von $\jmath$.
Schauen wir uns ein Simplextableau mit aktueller Basislösung $(\hat{z}, \hat{\jmath})$ an, so haben wir offenbar $\hat{A}(\hat{\jmath})\hat{z}(\hat{\jmath}) = \hat{b}$.

Lemma 5.9 *Sei $\big(\hat{z}, \jmath\big)$ aktuelle Basislösung des Simplexverfahrens. (Wir schreiben der Kürze $\jmath$ für $\hat{\jmath}$.) Sei k ein Komplement von $\jmath$, d.h. $k \in \{1, \ldots, n\}^{n-m}$ mit $\{j_1, \ldots, j_m\} \cup \{k_1, \ldots, k_{n-m}\} = \{1, \ldots, n\}$.*[3] *Dann gilt:*

(a) $A(\jmath) \in \mathbb{R}^{m \times m}$ ist regulär und $\hat{z}(\jmath) = A(\jmath)^{-1}b$ sowie $\hat{z}(k) = 0$.

(b) $\hat{c}(k) \;=\; c(k) - \big(A(\jmath)^{-1}A(k)\big)^\top c(\jmath)$ und $\hat{a}_{*s} = A(\jmath)^{-1}a_{*s}$.

Beweis: (a) Die Matrix $A(\jmath)$ ist regulär, da sie nach äquivalenten Gauss-Jordan-Umformungen in die Einheitsmatrix gebracht werden kann. Die weitere Aussage ist klar, da wegen $\hat{A}(\jmath)\hat{z}(\jmath) = \hat{b}$ auch $A(\jmath)\hat{z}(\jmath) = b$ gilt.
(b) Sei $x \in \mathbb{R}^n$ mit $Ax = b$, d.h. $A(\jmath)x(\jmath) + A(k)x(k) = b$, also $x(\jmath) + A(\jmath)^{-1}A(k)x(k) = A(\jmath)^{-1}b = \hat{z}(\jmath)$, also

$$c^\top x \;=\; c(k)^\top x(k) + c(\jmath)^\top \big[\hat{z}(\jmath) - A(\jmath)^{-1}A(k)x(k)\big] \;=\; c^\top \hat{z} + \big[c(k) - \big(A(\jmath)^{-1}A(k)\big)^\top c(\jmath)\big]^\top x(k)$$

und daher

$$\big[c(k) - \big(A(\jmath)^{-1}A(k)\big)^\top c(\jmath)\big]^\top x(k) \;=\; c^\top x - c^\top \hat{z}$$

Wegen Eigenschaft (c) des Algorithmus ist

$$c^\top x - c^\top \hat{z} \;=\; \hat{c}^\top x \;=\; \hat{c}(k)^\top x(k)\,.$$

Hier haben wir $\hat{c}(\jmath) = 0$ benutzt. Daher folgt

$$\big[c(k) - \big(A(\jmath)^{-1}A(k)\big)^\top c(\jmath)\big]^\top x(k) \;=\; \hat{c}(k)^\top x(k)\,.$$

Dies gilt für alle $x(k) \in \mathbb{R}^{n-m}$, denn dann setze $x(\jmath) = A(\jmath)^{-1}[b - A(k)x(k)]$. Hieraus folgt die Behauptung.
Genauso geht es mit der zweiten Gleichung: Sei $w := A(\jmath)^{-1}a_{*s}$, also $A(\jmath)w = a_{*s}$, also $\hat{A}(\jmath)w = \hat{a}_{*s}$, also $w = \hat{a}_{*s}$, da $\hat{A}(\jmath) = I$. □

Damit berechnet das Simplexverfahren (Phase II) folgende Schritte, wobei wir mit einer Basislösung $(\hat{z}, \jmath)$ starten, die aus Phase I stammt. Es sei k ein Komplement von $\jmath$.

[3] k enthält also die Indizes der freien Variablen. Sie können z.B. der Größe nach geordnet sein.

1. Starte mit $\hat{z}(\jmath) = A(\jmath)^{-1}b$ und $\hat{z}(k) = 0$, sowie Kosten $\gamma = c(\jmath)^\top A(\jmath)^{-1}b$. Speichere die Matrix $A(\jmath)^{-1} \in \mathbb{R}^{m\times m}$.

2. Setze $y := A(\jmath)^{-\top}c(\jmath)$. Teste, ob $A(k)^\top y \le c(k)$. Falls ja, so STOP. (Der Vektor $\hat{z}$ ist optimal.) Andernfalls fahre mit Schritt 3 fort.

3. Bestimme Index $s \in \{k_1, \ldots, k_{n-m}\}$ mit $\big(A(k)^\top y\big)_s > c_s$ (Blandsche Regel).

4. Berechne $w = A(\jmath)^{-1}a_{*s}$. Falls $w \le 0$, so STOP. (Es ist $\inf(P) = -\infty$.) Andernfalls fahre mit Schritt 5 fort.

5. Bestimme $r \in \{1, \ldots, m\}$ mit $\hat{z}_r/w_r = \min\big\{\hat{z}_i/w_i : w_i > 0\big\}$ (Blandsche Regel).

6. Setze $\tilde{\jmath} = (j_1, \ldots, j_{r-1}, s, j_{r+1}, \ldots, j_m)$, bestimme Komplement $\tilde{k}$ von $\tilde{\jmath}$ und setze

$$\begin{aligned} A(\tilde{\jmath})^{-1} &= \left(I - \frac{(w - e_{*r})e_{*r}^\top}{w_r}\right) A(\jmath)^{-1}\,, \\ \tilde{z}(\tilde{\jmath}) &= A(\tilde{\jmath})^{-1}b\,, \quad \tilde{z}(\tilde{k}) = 0\,, \\ \tilde{\gamma} &= \gamma \;+\; \frac{\hat{z}_r}{w_r}\big[c_s - (A(k)^\top y)_s\big] \end{aligned}$$

7. Update $\hat{} = \tilde{}$ und weiter mit Schritt 2.

Die Form der Inversen $A(\tilde{\jmath})^{-1}$ erhält man aus $A(\jmath)^{-1}$ mit der Darstellung

$$A(\tilde{\jmath}) \;=\; A(\jmath) \;+\; \big(a_{*s} - a_{*r}\big)\, e_{*r}^\top \;=\; A(\jmath)\,\big[I \;+\; (w - e_{*r})\, e_{*r}^\top\big]$$

und der Sherman-Morrison-Formel:

Lemma 5.10 *Es seien $u, v \in \mathbb{R}^m$ mit $u^\top v \neq -1$. Dann ist die Matrix $I + u\,v^\top \in \mathbb{R}^{m\times m}$ regulär und*

$$\big(I + u\,v^\top\big)^{-1} \;=\; I \;-\; \frac{u\,v^\top}{1 + u^\top v}\,.$$

Der Beweis verbleibt als Übung!

Wir sehen, dass man nur $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ fest gespeichert halten muß, und $A(\jmath)^{-1} \in \mathbb{R}^{m\times m}$, $y \in \mathbb{R}^m$, $w \in \mathbb{R}^m$ in jedem Schritt neu berechnen und speichern muß. Dieses Version nennt man **revidiertes Simplexverfahren**!

6 Konvexe Optimierung

In diesem Kapitel behandeln wir Optimierungsprobleme mit konvexer Zielfunktion und konvexen Nebenbedingungen. Der erste Abschnitt befasst sich mit Charakterisierungen konvexer Funktionen.

6.1 Konvexe Funktionen

Definition 6.1 *Sei $D \subseteq \mathbb{R}^n$ konvex.*

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

$$f\big(\lambda x + (1-\lambda)y\big) \;\leq\; \lambda f(x) + (1-\lambda) f(y)\,.$$

b) f heißt **strikt konvex**, *wenn für alle $x, y \in D$, $x \neq y$ und alle $\lambda \in (0,1)$ gilt:*

$$f\big(\lambda x + (1-\lambda)y\big) \;<\; \lambda f(x) + (1-\lambda) f(y)\,.$$

c) f heißt **gleichmäßig konvex**, *wenn es $c_0 > 0$ gibt, so dass für alle $x, y \in D$, $\lambda \in [0,1]$ gilt*

$$f\big(\lambda x + (1-\lambda)y\big) \;+\; c_0 \lambda(1-\lambda)\|x-y\|^2 \;\leq\; \lambda f(x) + (1-\lambda) f(y)\,.$$

Natürlich gilt: f gleichmäßig konvex $\Rightarrow$ f strikt konvex $\Rightarrow$ f konvex.

Beispiele 6.2 *a) $f(x) = c^\top x$, $x \in \mathbb{R}^n$, ist konvex, aber nicht strikt konvex.*

b) $f(x) = 1/(1+x)$, $x \geq 0$, ist strikt konvex, aber nicht gleichmäßig konvex, (Aufgabe)

c) Sei $Q \in \mathbb{R}^{n\times n}$ symmetrisch und positiv semi-definit (bzw. positiv definit), sowie $c \in \mathbb{R}^n$. Dann ist $f(x) = x^\top Q x + c^\top x$ konvex (bzw. gleichmäßig konvex).

Als nächstes wollen wir eine Charakterisierung der Konvexität über Ableitungen beweisen. Dazu erinnern wir an den **Gradienten** $\nabla f(x) \in \mathbb{R}^n$ und die **Hesse-Matrix** $\nabla^2 f(x) \in \mathbb{R}^{n\times n}$ der (hinreichend glatten) Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in $x \in \mathbb{R}^n$:

$$\nabla f(x) \;=\; \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^\top \quad \text{und} \quad \nabla^2 f(x) \;=\; \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,n}$$

Satz 6.3 *Sei $D \subseteq \mathbb{R}^n$ offen und konvex, $f : D \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gilt*

(a) f ist genau dann gleichmäßig konvex auf D, wenn eine der folgenden äquivalenten Bedingungen gilt:

$(*)$ *Es gibt $c_0 > 0$ mit* $\quad f(x) - f(y) \;\geq\; \nabla f(y)^\top (x-y) + c_0 \|x-y\|^2 \quad \forall x, y \in D$

$(**)$ *Es gibt $c_0 > 0$ mit* $\quad \big[\nabla f(x) - \nabla f(y)\big]^\top (x-y) \;\geq\; 2c_0 \|x-y\|^2 \quad \forall x, y \in D$

$(***)$ *Es gibt $c_0 > 0$ mit* $\quad z^\top \nabla^2 f(x) z \;\geq\; 2c_0 \|z\|^2 \quad \forall x \in D\,,\; z \in \mathbb{R}^n$

(b) f ist genau dann strikt konvex, falls in (a) überall $c_0 = 0$ gesetzt und „$\geq$“ durch „$>$“ für $x \neq y$ und für $z \neq 0$ ersetzt wird.

(c) f ist genau dann konvex, falls in (a) überall $c_0 = 0$ gesetzt wird.

Beweis nur von (a):

(i) Sei f gleichmäßig konvex mit Konstante $c_0 > 0$, also

$$f\big(y + \lambda(x-y)\big) \;\leq\; f(y) + \lambda\big[f(x) - f(y)\big] - c_0\lambda(1-\lambda)\|x-y\|^2$$

für alle $x, y \in D$, und $\lambda \in [0,1]$. Dann ist

$$\frac{f\big(y + \lambda(x-y)\big) - f(y)}{\lambda} \;\leq\; f(x) - f(y) - c_0(1-\lambda)\|x-y\|^2 .$$

Für $\lambda \to 0$ folgt $\nabla f(y)^\top (x-y) \leq f(x) - f(y) - c_0\|x-y\|^2$.

(ii) Es gebe $c_0 > 0$ mit $f(x) - f(y) \geq \nabla f(y)^\top (x-y) + c_0\|x-y\|^2$ für alle $x, y \in D$. Vertauschung von x und y und Addition liefert

$$0 \;\geq\; \big[\nabla f(x) - \nabla f(y)\big]^\top (y-x) + 2c_0\|x-y\|^2 , \quad \text{d.h}$$

$$\big[\nabla f(x) - \nabla f(y)\big]^\top (x-y) \;\geq\; 2c_0\|x-y\|^2 .$$

(iii) Es gebe $c_0 > 0$ mit $(**)$. Sei $z \in \mathbb{R}^n$ und $t \neq 0$ so klein, dass $x := y + tz \in D$. Dann ist wegen $(**)$:

$$2c_0 t^2\|z\|^2 \;\leq\; t\,\big[\nabla f(y+tz) - \nabla f(y)\big]^\top z, \quad \text{also}$$

$$2c_0\|z\|^2 \;\leq\; \frac{1}{t}\big[\nabla f(y+tz)^\top z - \nabla f(y)^\top z\big] \;=\; \sum_{j=1}^n z_j \left(\frac{\partial f(y+tz)}{\partial x_j} - \frac{\partial f(y)}{\partial x_j}\right)\frac{1}{t} .$$

Für $t \to 0$ folgt

$$2c_0\|z\|^2 \;\leq\; \sum_{j=1}^n z_j\, \nabla \frac{\partial f(y)}{\partial x_j}^\top z \;=\; z^\top \nabla^2 f(y) z .$$

(iv) Es gelte schließlich $(***)$. Wir setzen

$$\psi(\lambda) \;=\; f(x) + \lambda\big[f(y) - f(x)\big] - f\big(x + \lambda(y-x)\big) - c_0\lambda(1-\lambda)\|y-x\|^2$$

für $\lambda \in [0,1]$. Dann ist $\psi(0) = \psi(1) = 0$ und

$$\begin{aligned}
\psi'(\lambda) &= f(y) - f(x) - \nabla f\big(x + \lambda(y-x)\big)^\top (y-x) - c_0(1-2\lambda)\|y-x\|^2 , \\
\psi''(\lambda) &= -(y-x)^\top \nabla^2 f\big(x + \lambda(y-x)\big)(y-x) + 2c_0\|y-x\|^2 \;\leq\; 0 .
\end{aligned}$$

Also ist ψ' monoton nicht steigend. Mit $\psi(0) = \psi(1) = 0$ folgt $\psi(\lambda) \geq 0$ für alle $\lambda \in [0,1]$. (Widerspruchsbeweis: Andernfalls existiert $\hat\lambda \in (0,1)$ mit $\psi(\hat\lambda) < 0$ und $\psi'(\hat\lambda) = 0$. Dann ist auch $\psi'(\lambda) \leq 0$ für $\lambda \in [\hat\lambda, 1]$, also auch $\psi(1) < 0$, ein Widerspruch.) Daher ist f gleichmäßig konvex. □

Bemerkung: Die Charakterisierungen von $(*)$ und $(**)$ gelten auch für einmal stetig differenzierbare Funktionen.

6.2 Existenz und Eindeutigkeit

Bevor wir auf die Dualitätstheorie eingehen, möchten wir einiges zur Existenz beitragen. Wir erinnern uns: Für **lineare** Optimierungsprobleme reicht schon für Existenz aus, dass $M \neq \emptyset$ und $\inf(P) > -\infty$ (d.h. $\inf\limits_{x \in M} f(x) > -\infty$). Für allgemeine konvexe Probleme reicht dies nicht aus. Betrachte etwa

$$f(x) \;=\; \frac{1}{x} \quad \text{auf} \quad M = [1, \infty)\,.$$

Dann ist $\inf\limits_{x \in M} f(x) = 0$, aber $f(x) > 0$ für alle $x \in M$. Man braucht also mehr Voraussetzungen. Wir betrachten wieder das allgemeine konvexe Problem:

$$(P) \qquad \text{Minimiere} \quad f(x) \quad \text{auf} \quad M = \{x \in K : g(x) \leq 0,\ Ax = b\},$$

wobei $K \subseteq \mathbb{R}^n$ eine konvexe Menge ist, $f : K \to \mathbb{R}$, $g : K \to \mathbb{R}^p$ konvexe Funktionen, $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$.
Zunächst bemerken wir, daß jedes **lokale** Minimum x^* schon ein **globales** Minimum ist:

Lemma 6.4 *Sei $M \subseteq \mathbb{R}^n$ konvex, $f : M \to \mathbb{R}$ konvex und x^* ein lokales Minimum von f auf M, d.h. es existiert $\varepsilon > 0$ mit $f(x^*) \leq f(x)$ für alle $x \in M$ mit $\|x - x^*\| \leq \varepsilon$. Dann ist x^* sogar globales Minimum, d.h. $f(x^*) \leq f(x)$ für alle $x \in M$.*

Beweis: Sei $x \in M$ beliebig. Wähle dann $\lambda \in (0,1)$ so klein, dass $\lambda\|x - x^*\| \leq \varepsilon$. Dann ist $f(x^*) \leq f(x^* + \lambda(x - x^*)) \leq f(x^*) + \lambda\big[f(x) - f(x^*)\big]$, also $f(x) - f(x^*) \geq 0$. Daher ist x^* globales Minimum. □

Wir erinnern an die Menge

$$\Lambda \;=\; \Big\{\big(f(x) + r, g(x) + z, Ax - b\big) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^m : r \geq 0,\ z \geq 0,\ x \in K\Big\}\,.$$

Dann können wir zeigen:

Satz 6.5 *Es sei $M \neq \emptyset$, $\inf(P) > -\infty$ und Λ sei abgeschlossen. Dann besitzt (P) eine Lösung.*

Bemerkung: In unserem Beispiel ist

$$\Lambda \;=\; \left\{\frac{1}{x} + r : x \geq 1\,,\ r \geq 0\right\} \;=\; (0, \infty)\,,$$

und dies Intervall ist nicht abgeschlossen!

Beweis: Wir zeigen, daß $\big(\inf(P), 0, 0\big) \in \Lambda$. Nach Definition von $\inf(P)$ existiert eine Folge (x_k) in M mit $f(x_k) \to \inf(P)$. Es ist $\big(f(x_k), 0, 0\big) \in \Lambda$, da

$$\begin{aligned} f(x_k) &= f(x_k) + r_k \quad \text{mit} \quad r_k = 0\,, \\ 0 &= g(x_k) + z_k \quad \text{mit} \quad z_k = -g(x_k) \geq 0\,, \\ 0 &= Ax_k - b\,. \end{aligned}$$

Ferner konvergiert $\big(f(x_k), 0, 0\big)$ gegen $\big(\inf(P), 0, 0\big)$. Wegen der Abgeschlossenheit von Λ ist auch $\big(\inf(P), 0, 0\big) \in \Lambda$. Also existiert $x^* \in K$, $r \geq 0$, $z \geq 0$ mit $\inf(P) = f(x^*) + r$ sowie

$g(x^*)+z=0$ und $Ax^*-b=0$. Also ist $x^* \in M$. Wegen $f(x^*) \geq \inf(P) = f(x^*)+r \geq f(x^*)$ muss $r=0$ sein, also $f(x^*)=\inf(P)$. Daher ist x^* optimal. □

Wir sehen uns noch einmal das lineare Problem an:

$$\text{Minimiere} \quad c^\top x \quad \text{auf} \quad M = \{x : x \geq 0,\ Ax = b\}\,.$$

Dann ist

$$\Lambda \;=\; \big\{(c^\top x + r, Ax - b) \in \mathbb{R} \times \mathbb{R}^m : x \geq 0,\ r \geq 0\big\}\,.$$

Wir schreiben Λ in der Form

$$\Lambda \;=\; \left\{ \begin{bmatrix} c^\top & 1 \\ A & 0_m \end{bmatrix} \begin{bmatrix} x \\ r \end{bmatrix} : \begin{bmatrix} x \\ r \end{bmatrix} \geq 0 \right\} \;-\; \begin{bmatrix} 0 \\ b \end{bmatrix}\,.$$

Also ist die Menge wieder ein endlich erzeuter Kegel und daher nach dem Satz von Weyl (Satz 2.6) abgeschlossen. Die zusätzliche Voraussetzung der Abgeschlossenheit von Λ ist bei linearen Problemen also immer erfüllt.

Für lineare Probleme können wir keine Eindeutigkeit erwarten (weshalb nicht?), daher auch nicht für konvexe Probleme. Wir haben aber:

Satz 6.6 *Ist $M \neq \emptyset$ und f* **strikt konvex***, so besitzt (P) höchstens eine Lösung.*

Beweis: Wären $x_1^*, x_2^* \in M$ optimal und $x_1^* \neq x_2^*$, so wäre $f(x_1^*) = f(x_2^*) = \inf(P)$, also $\inf(P) = \frac{1}{2}f(x_1^*) + \frac{1}{2}f(x_2^*) > f\left(\frac{1}{2}x_1^* + \frac{1}{2}x_2^*\right)$ und $x := \frac{1}{2}x_1^* + \frac{1}{2}x_2^* \in M$. Dies ist ein Widerspruch zur Optimalität von x_1^* und x_2^*. □

Schießlich zeigen wir noch:

Satz 6.7 *Ist $M \neq \emptyset$ und f stetig und* **gleichmäßig konvex***, so ist (P) eindeutig lösbar.*

Beweis: Wegen des letzten Satzes ist nur die Existenz zu zeigen, da jede gleichmäßig konvexe Funktion strikt konvex ist.
Halte $\hat{x} \in M$ fest und definiere $\hat{M} := \{x \in M : f(x) \leq f(\hat{x})\}$. Jedes Minimum von f auf $\hat{M}$ ist auch Minimum von f auf M, denn für $x \in M \setminus \hat{M}$ gilt ja $f(x) > f(\hat{x}) \geq \min\big\{f(y) : y \in \hat{M}\big\}$. Wir zeigen, dass $\hat{M}$ beschränkt ist. Dann ist $\hat{M}$ kompakt, und f besitzt auf $\hat{M}$ sicher ein Minimum.
Wähle $\delta > 0$ mit $|f(x) - f(\hat{x})| \leq 1$ für $\|x - \hat{x}\| \leq \delta$. Dann gilt $f(x) \geq f(\hat{x}) - 1 =: \gamma$ für alle $x \in \hat{M}$, $\|x - \hat{x}\| \leq \delta$. Sei jetzt $x \in \hat{M}$, $x \neq \hat{x}$, beliebig, setze $\lambda := \delta / \|x - \hat{x}\|$.
1. Fall: $\lambda \geq 1/2$. Dann ist $\|x - \hat{x}\| \leq 2\,\delta$.
2. Fall: $\lambda \leq 1/2$. Dann ist $1 - \lambda \geq 1/2$, also

$$\begin{aligned} f(\hat{x}) \;&\geq\; \lambda f(x) + (1-\lambda) f(\hat{x}) \;\geq\; f\big(\lambda x + (1-\lambda)\hat{x}\big) + c\,\lambda(1-\lambda)\|x - \hat{x}\|^2 \\ &\geq\; \gamma \;+\; c\,\lambda(1-\lambda)\|x - \hat{x}\|^2 \;=\; \gamma \;+\; c\,\delta\,(1-\lambda)\|x - \hat{x}\| \\ &\geq\; \gamma \;+\; \frac{c\delta}{2}\,\|x - \hat{x}\|\,, \end{aligned}$$

da $\|\big(\lambda x + (1-\lambda)\hat{x}\big) - \hat{x}\| \;=\; \lambda\|x - \hat{x}\| \;\leq\; \delta$.
Also ist zusammen

$$\|x - \hat{x}\| \;\leq\; \max\left\{2\delta\,,\; \frac{2\big[f(\hat{x}) - \gamma\big]}{c\,\delta}\right\} \quad \text{für alle } x \in \hat{M}\,.$$

□

6.3 Das duale Problem

Sei jetzt eine konvexe Menge $K \subseteq \mathbb{R}^n$, konvexe Funktionen $f : K \longrightarrow \mathbb{R}$ und $g_i : K \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$, sowie eine Matrix $A \in \mathbb{R}^{m \times n}$ und ein Vektor $b \in \mathbb{R}^m$ gegeben. In diesem Kapitel behandeln wir Probleme der Form:

$$(P) \qquad \text{Minimiere} \quad f(x) \quad \text{auf} \quad M \;:=\; \big\{x \in \mathbb{R}^n : x \in K,\; g(x) \leq 0,\; Ax = b\big\}\,.$$

Die Menge M ist konvex (klar?).

Beispiele 6.8 *und Spezialfälle:*

(a) $f(x) = c^\top x$ mit $c \in \mathbb{R}^n$, $K = \{x \in \mathbb{R}^n : x \geq 0\}$, und Funktionen g_i treten nicht auf. Dies liefert das lineare Optimierungsproblem in der zweiten Normalform (P_2) von Seite 5 im Skript. Ist $K = \mathbb{R}^n$, $g(x) = Ax - b$, so erhalten wir die erste Normalform (P_1).

(b) Ist $c \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$ symmetrisch und positiv semidefinit, so setzen wir

$$f(x) \;:=\; x^\top Q x + c^\top x\,, \quad x \in \mathbb{R}^n\,.$$

Dann ist f konvex. Am einfachsten sieht man dies, wenn man die Charakterisierung der Konvexität über die zweite Ableitung benutzt: Eine zweimal stetig differenzierbare Funktion $f : \mathbb{R}^n \supseteq K \to \mathbb{R}$ ist genau dann konvex auf K, wenn die Hesse-Matrix $\nabla^2 f(x)$ in allen Punkten $x \in K$ positiv semidefinit ist. Für unser Beispiel ist $\nabla^2 f(x) = 2Q$ für jedes $x \in \mathbb{R}^n$, und diese Matrix ist nach Voraussetzung positiv semidefinit.

Wir formen (P) so um, dass wir es graphisch veranschaulichen und dann genau wie bei linearen Problemen ein duales Problem formulieren können. Betrachte dazu die Menge

$$\Lambda \;:=\; \big\{(f(x) + r, g(x) + z, Ax - b) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^m : r \geq 0,\; z \geq 0_p,\; x \in K\big\}$$

und $\Lambda_0 := \Lambda \cap \big(\mathbb{R} \times \{0_p\} \times \{0_m\}\big)$, sowie das Problem:

$$(\tilde{P}) \qquad \text{Minimiere} \quad \varphi(\beta, u, v) := \beta \quad \text{unter} \quad (\beta, u, v) \in \Lambda_0\,.$$

Die Menge Λ ist konvex. Dies sieht man so: Es seien $\lambda \in [0, 1]$, $r_j \geq 0$, $z_j \geq 0_p$ und $x_j \in K$ gegeben, $j = 1, 2$. Dann sind $\big(f(x_j) + r_j, g(x_j) + z_j, Ax_j - b\big) \in \Lambda$ für $j = 1, 2$. Setze dann

$$\begin{aligned}
x_0 \;&:=\; \lambda x_1 \;+\; (1-\lambda)x_2 \;\in\; K\,,\\
r_0 \;&:=\; \lambda f(x_1) + (1-\lambda) f(x_2) \;-\; f(x_0) \;+\; \lambda r_1 + (1-\lambda) r_2 \;\geq\; 0\,,\\
z_0 \;&:=\; \lambda g(x_1) + (1-\lambda) g(x_2) \;-\; g(x_0) \;+\; \lambda z_1 + (1-\lambda) z_2 \;\geq\; 0\,.
\end{aligned}$$

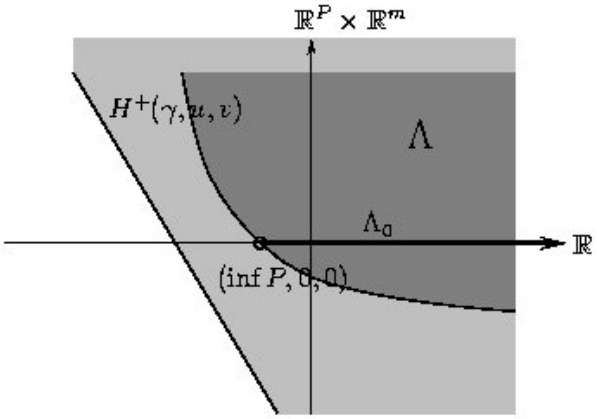
Dann ist

$$\begin{aligned}
\lambda[f(x_1) + r_1] \;+\; (1-\lambda)[f(x_2) + r_2] \;&=\; f(x_0) + r_0\,,\\
\lambda[g(x_1) + z_1] \;+\; (1-\lambda)[g(x_2) + r_2] \;&=\; g(x_0) + z_0\,,\\
\lambda[Ax_1 - b] \;+\; (1-\lambda)[Ax_2 - b] \;&=\; Ax_0 - b\,.
\end{aligned}$$

Dies beweist die Konvexität Es ist

$$\begin{aligned}(\beta,u,v)\in\Lambda_0 &\iff \exists r\geq 0\,,\ z\geq 0\,,\ x\in K \text{ mit } \beta=f(x)+r\,,\ 0=g(x)+z\,,\ 0=Ax-b\,.\\ &\iff \exists x\in K\,,\ \beta\geq f(x) \text{ mit } g(x)\leq 0 \text{ und } Ax=b\\ &\iff \exists x\in M\,,\ \beta\geq f(x)\,.\end{aligned}$$

Also ist $(\tilde{P})$ äquivalent zu (P).



Das **duale Problem** soll darin bestehen, unter allen Hyperebenen in $\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m$, die nicht parallel zu $\mathbb{R}\times\{0\}\times\{0\}$ sind und Λ im nichtnegativen Halbraum enthalten, diejenige zu finden, deren Schnitt mit $\mathbb{R}\times\{0\}\times\{0\}$ maximal ist. Solche Hyperebenen werden beschrieben durch

$$H(\gamma,u,v) \;:=\; \big\{(t,w,z)\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m : t+u^\top w+v^\top z=\gamma\big\}$$

mit zugehörigem Halbraum

$$H^+(\gamma,u,v) \;:=\; \big\{(t,w,z)\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m : t+u^\top w+v^\top z\geq\gamma\big\}\,.$$

Das duale Problem kann also formuliert werden als:

$$(\tilde{D})\qquad \text{Maximiere}\quad \psi(\gamma,u,v):=\gamma\quad\text{unter}\quad \Lambda\subseteq H^+(\gamma,u,v)\,.$$

Es ist

$$\begin{aligned}\Lambda\subseteq H^+(\gamma,u,v) &\iff f(x)+r+u^\top(g(x)+z)+v^\top(Ax-b)\ \geq\ \gamma\\ &\qquad \text{für alle } x\in K\,,\ z\geq 0\,,\ r\geq 0\\ &\iff \big[f(x)+u^\top g(x)+v^\top(Ax-b)\big]+u^\top z\ \geq\ \gamma\\ &\qquad \text{für alle } x\in K\,,\ z\geq 0\,. \qquad (*)\end{aligned}$$

Setzen wir speziell $z=0$, so erhalten wir

$$f(x)+u^\top g(x)+v^\top(Ax-b)\ \geq\ \gamma\quad\text{für alle } x\in K\,.$$

Wir zeigen, dass diese Bedingung zusammen mit $u \geq 0$ sogar äquivalent zu $(*)$ ist. Dazu müssen wir offenbar zeigen, dass $u \geq 0$ aus der Formel $(*)$ folgt. Wäre dies nicht der Fall, so würde ein k existieren mit $u_k < 0$. Halten wir dann ein $x \in K$ fest und setzen

$$z_j \;=\; \begin{cases} 0, & j \neq k\,, \\ \alpha, & j = k \end{cases} \quad \text{für } \alpha > 0\,,$$

so ist $\alpha\, u_k \geq \gamma - [f(x) + u^\top g(x) + v^\top (Ax - b)]$. Für $\alpha \to \infty$ konvergiert die linke Seite gegen $-\infty$, ein Widerspruch.
Also ist $u \geq 0$ und daher

$$\begin{aligned} \Lambda \subseteq H^+(\gamma, u, v) \;\iff\; & u \;\geq 0 \quad \text{und} \\ & \big[f(x) + u^\top g(x) + v^\top (Ax - b)\big] \;\geq\; \gamma \text{ für alle } x \in K\,. \end{aligned}$$

Wir definieren die Funktion

$$F(u, v) \;:=\; \inf_{x \in K} \big[f(x) + u^\top g(x) + v^\top (Ax - b)\big]$$

und haben dann, dass $(\tilde{D})$ äquivalent ist zu:

(D) Maximiere $F(u, v)$ auf $N := \big\{(u, v) \in \mathbb{R}^p \times \mathbb{R}^m : u \geq 0\,,\; F(u, v) > -\infty\big\}$

Wie sieht (D) in den Spezialfällen aus?

(a) Sei zunächst $f(x) = c^\top x$ und $M = \{x \in \mathbb{R}^n : x \geq 0\,,\; Ax = b\}$. Dann hat (D) die Form

$$(D) \quad \text{Maximiere} \quad F(v) \;=\; \inf_{x \geq 0} \big[c^\top x + v^\top (Ax - b)\big] \;=\; \inf_{x \geq 0} \big[(c + A^\top v)^\top x\big] \;-\; v^\top b$$

auf $N = \{v \in \mathbb{R}^m : F(v) > -\infty\}$. Es ist genau dann $F(v) > -\infty$, wenn $c + A^\top v \geq 0$ (klar?), und in diesem Fall ist $F(v) = (-v)^\top b$. Setzen wir $y = -v$, so erhalten wir das Problem, $y^\top b$ zu maximieren unter der Nebenbedingung $A^\top y \leq c$. Dies ist das duale Problem, das wir schon kennen!

(b) Sei jetzt $f(x) = c^\top x$ mit $M = \{x \in \mathbb{R}^n : Ax - b \leq 0\}$. Es verbleibt als Aufgabe, die Äquivalenz von (D) mit dem dualen Problem dieses linearen Problems zu zeigen.

(c) Jetzt betrachten wir das quadratische Problem:

$$(P) \qquad \text{Minimiere} \quad x^\top Q x + c^\top x \quad \text{unter} \quad Ax = b\,,\; x \geq 0\,,$$

mit positiv semidefiniter Matrix $Q \in \mathbb{R}^{n \times n}$. Wir setzen $K = \mathbb{R}^n$ und $g(x) = -x$ und erhalten

$$\begin{aligned} F(u, v) \;&=\; \inf_{x \in \mathbb{R}^n} \big[x^\top Q x + c^\top x - u^\top x + v^\top (Ax - b)\big] \\ &=\; \inf_{x \in \mathbb{R}^n} \big[x^\top Q x + (c - u + A^\top v)^\top x\big] \;-\; v^\top b\,, \end{aligned}$$

sowie

$$N \;=\; \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^m : u \geq 0\,,\; F(u, v) > -\infty\big\}\,.$$

Wir benötigen

Lemma 6.9 *Sei Q symmetrisch und positiv semidefinit, $d \in \mathbb{R}^n$. Dann ist*

$$\inf_{x\in\mathbb{R}^n}[x^\top Qx + d^\top x] \;=\; \begin{cases} -z^\top Qz, & \textit{falls es } z \textit{ gibt mit } 2Qz + d = 0\,, \\ -\infty, & \textit{sonst, d.h. falls } d \notin W(Q)\,. \end{cases}$$

Beweis: (i) Es gebe z mit $2Qz + d = 0$. Die binomische Formel liefert:

$$\begin{aligned}[x^\top Qx + d^\top x] - [z^\top Qz + d^\top z] \;&=\; 2\,z^\top Q(x-z) + d^\top(x-z) + \underbrace{(x-z)^\top Q(x-z)}_{\geq\, 0} \\ &\geq\; (2Qz+d)^\top(x-z) \;=\; 0\,.\end{aligned}$$

Also ist $x^\top Qx + d^\top x \geq z^\top Qz + d^\top z = -z^\top Qz$.
(ii) Es gebe kein z mit $2Qz + d = 0$. Dies bedeutet, dass $d \notin W(Q)$. Zerlege $d = \hat{d} + \tilde{d}$ mit $\hat{d} \in W(Q)$ und $\tilde{d} \perp W(Q)$. Setze $x_t := -t\tilde{d}$ für $t \geq 0$. Es ist $\tilde{d} \neq 0$, da $d \notin W(Q)$. Dann ist, da $Q\tilde{d} \in W(Q)$,

$$x_t^\top Qx_t + d^\top x_t \;=\; t^2\,\tilde{d}^\top Q\tilde{d} - t\,d^\top\tilde{d} \;=\; -t\|\tilde{d}\|^2 \;\longrightarrow\; -\infty\,, \quad t \to \infty.$$

Also ist $\inf\limits_{x\in\mathbb{R}^n}[x^\top Qx + d^\top x] = -\infty$. □

Wenden wir dieses Lemma auf

$$F(u,v) \;=\; \inf_{x\in\mathbb{R}^n}[x^\top Qx + (c - u + A^\top v)^\top x] \;- v^\top b$$

an, so ist

$$F(u,v) > -\infty \iff \text{es gibt } z \in \mathbb{R}^n \text{ mit } 2Qz + c - u + A^\top v = 0\,,$$

und in diesem Fall ist

$$F(u,v) \;=\; -z^\top Qz \;-\; v^\top b\,.$$

Also ist (D) äquivalent zu:

$$(D)\quad \text{Maximiere} \quad -z^\top Qz - b^\top v \text{ auf } \tilde{N} \;:=\; \big\{(z,v) \in \mathbb{R}^n \times \mathbb{R}^m : 2Qz + c + A^\top v \geq 0\big\}\,.$$

Ersetzen wir v durch $-y$, so erhalten wir die endgültige Form:

$$(D)\quad \text{Maximiere} \quad b^\top y - z^\top Qz \quad \text{auf} \quad N \;=\; \big\{(z,y) \in \mathbb{R}^n \times \mathbb{R}^m : 2Qz + c - A^\top y \geq 0\big\}\,.$$

Wir erkennen wieder, dass im Fall $Q = O$ (Nullmatrix) dieses duale Problem genau die Form des dualen Problems für lineare Optimierungsaufgaben besitzt.

Beispiel 6.10 Gegeben seien $a_{*j} \in \mathbb{R}^n$, $j = 1, \ldots, m$. Gesucht ist der Umkreis mit dem kleinsten Radius. Dies formulieren wir so:

$$(P_1) \qquad \text{Finde minimales } r, \text{ so daß} \quad \|a_{*j} - x\| \;\leq\; r \text{ für alle } j\,.$$

Die Nebenbedingung ist äquivalent zu $\|a_{*j}\|^2 - 2\,a_{*j}^\top x + \|x\|^2 \leq r^2$.

Die Zielfunktion ist in diesem Beispiel linear, nämlich $f(x,r) = r$, die Nebenbedingung ist quadratisch, aber nicht konvex:

$$g_j(x,r) \;:=\; \|x\|^2 \;-\; 2\,a_{*j}^\top x \;-\; r^2 \;+\; \|a_{*j}\|^2\,, \quad j = 1,\ldots,m\,.$$

Wir transformieren das Problem auf eins mit quadratischer Zielfunktion und linearen Nebenbedingungen:
Setze $x_0 := \|x\|^2 - r^2$. Dann ist $r^2 = \|x\|^2 - x_0$, also haben wir:
(P_2)
Minimiere $\quad f(x_0,x) = \|x\|^2 - x_0 \quad$ unter $\quad x_0 - 2\,a_{*j}^\top x \;\le\; -\|a_{*j}\|^2 \quad$ für $\quad j = 1,\ldots,m\,.$

(P_2) ist ein quadratisches Optimierungsproblem mit linearen Ungleichungen als Nebenbedingungen. (Stellen Sie die Matrix Q und den Vektor c auf!) Wir haben:

Lemma 6.11 *Die Probleme (P_1) und (P_2) sind äquivalent in dem folgende Sinn:*

(a) Ist (x,r) zulässig (bzw. optimal) für (P_1), so ist (x,x_0) mit $x_0 = \|x\|^2 - r^2$ zulässig (bzw. optimal) für (P_2).

(b) Ist (x,x_0) zulässig (bzw. optimal) für (P_2), so ist $\|x\|^2 \ge x_0$ und (x,r) mit $r = \sqrt{\|x\|^2 - x_0}$ ist zulässig (bzw. optimal) für (P_1).

Beweis: (a) ist klar, für (b) ist nur $\|x\|^2 \ge x_0$ zu zeigen:
Aus $x_0 - 2\,a_{*j}^\top x \le -\|a_{*j}\|^2$ folgt $\|a_{*j}\|^2 - 2\,a_{*j}^\top x \le -x_0$, d.h. mit quadratischer Ergänzung $0 \le \|a_{*j} - x\|^2 \le \|x\|^2 - x_0$. □

Auf dieses Beispiel kommen wir später noch zurück.

6.4 Die Dualitätssätze

In diesem Abschnitt sei $K = \mathbb{R}^n$. Damit haben wir das Optimierungsproblem:

$$(P) \qquad \text{Minimiere} \quad f(x) \quad \text{auf} \quad M \;:=\; \{x \in \mathbb{R}^n : g(x) \le 0,\; Ax = b\}\,,$$

und

$$(D) \qquad \text{Maximiere} \quad F(u,v) \;:=\; \inf_{x\in\mathbb{R}^n}\big[f(x) + u^\top g(x) + v^\top(Ax-b)\big] \quad \text{auf}$$

$$N \;=\; \big\{(u,v) \in \mathbb{R}^p \times \mathbb{R}^m : u \ge 0,\; F(u,v) > -\infty\big\}\,.$$

Schon hier führen wir die **Lagrangefunktion** $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ ein:

$$L(x,u,v) \;:=\; f(x) \;+\; u^\top g(x) \;+\; v^\top(Ax-b)\,, \quad x \in \mathbb{R}^n\,,\; u \in \mathbb{R}^p\,,\; v \in \mathbb{R}^m\,.$$

Dann ist $F(u,v) = \inf\{L(x,u,v) : x \in \mathbb{R}^n\}$. Die dualen Variablen u und v heißen auch **Lagrangesche Multiplikatoren**. Wir hängen einfach an die Zielfunktion f die Nebenbedingungen mit den Multiplikatoren an. Darauf werden wir noch genauer im nächsten Kapitel eingehen.

Einfach ist wieder der Nachweis des schwachen Dualitätssatzes:

Satz 6.12 *(schwacher Dualitätssatz)*
Gegeben seien (P) und (D). Dann gilt:

$$f(x) \;\geq\; F(u,v) \quad \textit{für alle } x \in M \textit{ und } (u,v) \in N\,.$$

Ist $f(x^*) = F(u^*,v^*)$ für ein $x^* \in M$ und ein Paar $(u^*,v^*) \in N$, so sind x^* und (u^*,v^*) optimal für (P) bzw. (D).

Der **Beweis** ist trivial wegen der Ungleichungskette

$$F(u,v) \;\leq\; f(x) \;+\; \underbrace{u^\top g(x)}_{\leq 0} \;+\; v^\top(\underbrace{Ax-b}_{=0}) \;\leq\; f(x)\,.$$

□

Für den starken Dualitätssatz benötigen wir zwei Zusatzvoraussetzungen. Diese fasst man zusammen und nennt sie (verallgemeinerte) **Slaterbedingung:**

(SB) (i) $\operatorname{Rang} A = m \leq n$, und

(ii) es gibt $\hat{x} \in \mathbb{R}^n$ mit $A\hat{x} = b$ und $g_i(\hat{x}) < 0$ für alle $i = 1,\ldots,p$.

Satz 6.13 *(starker Dualitätssatz)*
f, g seien stetig, das Problem (P) sei lösbar durch $x^* \in M$ und die Slaterbedingung (SB) gelte. Dann gibt es auch eine Lösung $(u^*,v^*) \in N$ von (D) und $f(x^*) = F(u^*,v^*)$. Ferner gilt $u_j^*\, g_j(x^*) = 0$ für alle $j = 1,\ldots,p$.

Beweis: Ähnlich wie oben betrachten wir die Mengen

$$\Lambda_\ell \;:=\; \big\{(f(x)+r, g(x)+z, Ax-b) \in \mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m : r \geq 0,\; z \geq 0,\; \|x\| \leq \ell\big\}\,.$$

Diese Mengen sind nicht leer, konvex und abgeschlossen. Die Konvexität von Λ_ℓ haben wir oben bei der Konvexität von Λ schon gezeigt. (Man setze dort nämlich nur $K = \big\{x \in \mathbb{R}^n : \|x\| \leq \ell\big\}$.) Um die Abgeschlossenheit von Λ_ℓ zu zeigen, betrachten wir irgendeine Folge $\big(f(x_k)+r_k, g(x_k)+z_k, Ax_k-b\big)$ in Λ_ℓ, die gegen ein $(\alpha,u,v) \in \mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m$ konvergiert. Dann ist $\|x_k\| \leq \ell$, $r_k \geq 0$ und $z_k \geq 0$ für alle $k \in \mathbb{N}$. Es gibt daher eine konvergente Teilfolge $x_k \to x$ und $\|x\| \leq \ell$. Wegen $f(x_k)+r_k \to \alpha$ und $g(x_k)+z_k \to u$ konvergieren also auch $r_k \to r$ und $z_k \to z$ mit $r \geq 0$ und $z \geq 0$, also $\alpha = f(x)+r$, $u = g(x)+z$ und auch $Ax-b = v$. Damit ist die Abgeschlossenheit von Λ_ℓ gezeigt.

Es ist $\big(f(x^*) - 1/\ell, 0, 0\big) \notin \Lambda_\ell$, denn sonst würde es $x \in \mathbb{R}^n$, $r \geq 0$ und $z \geq 0$ geben mit $f(x^*) - 1/\ell = f(x) + r$, $0 = g(x) + z$ und $0 = Ax - b$. Diese würde $x \in M$ und $f(x) = f(x^*) - 1/\ell - r < f(x^*)$ bedeuten, ein Widerspruch zur Optimalität von x^*.

Also liegt gehört der Punkt $\big(f(x^*)-1/\ell, 0, 0\big)$ nicht zur konvexen, abgeschlossenene Menge Λ_ℓ. Daher kann ich den Trennungssatz im Raum $\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m$ anwenden: Für jedes $\ell \in \mathbb{N}$ existiert $(\lambda_\ell, u_\ell, v_\ell) \in \mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^m$ und $\gamma_\ell \in \mathbb{R}$ mit $\lambda_\ell^2 + \|u_\ell\|^2 + \|v_\ell\|^2 = 1$ und

$$\lambda_\ell\big[f(x)+r\big] \;+\; u_\ell^\top\big[g(x)+z\big]) \;+\; v_\ell^\top[Ax-b] \;\geq\; \gamma_\ell \;\geq\; \lambda_\ell\left[f(x^*) - \frac{1}{\ell}\right] \qquad (*)$$

für alle x mit $\|x\| \le \ell$ und alle $r \ge 0$, $z \ge 0$. Dies gilt für jedes ℓ. Es gibt konvergente Teilfolgen $\lambda_\ell \to \lambda$, $u_\ell \to u$ und $v_\ell \to v$ wegen $\lambda_\ell^2 + \|u_\ell\|^2 + \|v_\ell\|^2 = 1$. Wegen $(*)$ ist dann auch (γ_ℓ) beschränkt, also $\gamma_\ell \to \gamma$ (für eine weitere Teilfolge). Für festes $x \in \mathbb{R}^n$, $r \ge 0$, $z \ge 0$, können wir ℓ gegen ∞ gehen lassen, erhalten $\|x\| \le \ell$ und dann aus $(*)$:

$$\lambda\big[f(x)+r\big] \;+\; u^\top\big[g(x)+z\big] \;+\; v^\top[Ax-b] \;\ge\; \gamma \;\ge\; \lambda f(x^*)\,. \qquad (**)$$

Dies gilt für alle $x \in \mathbb{R}^n$, $r \ge 0$ und $z \ge 0$. Außerdem ist $\lambda^2 + \|u\|^2 + \|v\|^2 = 1$. Es ist $\lambda \ge 0$ und $u \ge 0$ (bekannte Schlussweise). Wir zeigen sogar $\lambda > 0$. Annahme $\lambda = 0$. Dann setzen wir $z = 0$ und $x = \hat{x}$ (aus der Slaterbedingung) und erhalten $\underbrace{u^\top}_{\ge 0}\underbrace{g(\hat{x})}_{<0} + v^\top(\underbrace{A\hat{x}-b}_{=0}) \ge \gamma \ge 0$, also $u = 0$. Damit hätten wir $v^\top(Ax-b) \ge 0$ für alle $x \in \mathbb{R}^n$, also $(A^\top v)^\top x \ge v^\top b$ für alle $x \in \mathbb{R}^n$. Hieraus folgt $A^\top v = 0$ und, da $A^\top$ injektiv wegen der Slaterbedingung, $v = 0$. Dies widerspricht $\lambda^2 + \|u\|^2 + \|v\|^2 = 1$. Also ist $\lambda > 0$. Setze schließlich $u^* := u/\lambda \ge 0$ und $v^* := v/\lambda$. Dann folgt aus $(**)$ für $r = 0$ und $z = 0$:

$$f(x) \;+\; {u^*}^\top g(x) \;+\; {v^*}^\top(Ax-b) \;\ge\; f(x^*) \quad \text{für alle} \quad x \in \mathbb{R}^n\,. \qquad (***)$$

Nimmt man das Infimum über x, so erhält man $F(u^*, v^*) \ge f(x^*)$, also Gleichheit nach dem schwachen Dualitätssatz. Nimmt man $x = x^*$ in $(***)$, so erhält man ${u^*}^\top g(x^*) = 0$, also $u_j^* g_j(x^*) = 0$ für alle $j = 1, \ldots, p$. □

Wir können nun einen weiteren Hauptsatz beweisen.

Satz 6.14 *(Sattelpunktsatz)*
Es sei das konvexe Optimierungsproblem (P) gegeben mit $K = \mathbb{R}^n$. Die Slaterbedingung (SB) sei erfüllt. Definiere wieder die **Lagrangefunktion** *durch*

$$L(x,u,v) \;:=\; f(x) \;+\; u^\top g(x) \;+\; v^\top(Ax-b)\,, \quad x \in \mathbb{R}^n\,,\; u \in \mathbb{R}^p\,,\; v \in \mathbb{R}^m\,.$$

(a) Sei $x^* \in M$ optimal für (P). Dann existieren $u^* \in \mathbb{R}^p$, $v^* \in \mathbb{R}^n$ mit ${u^*}^\top g(x^*) = 0$ und

$$L(x^*,u,v) \;\le\; L(x^*,u^*,v^*) \;\le\; L(x,u^*,v^*) \qquad (6.1)$$

für alle $x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$ mit $u \ge 0$ und alle $v \in \mathbb{R}^m$.

Dies bedeutet, dass L in (x^*,u^*,v^*) einen Sattelpunkt besitzt: x^* ist Minimum von $L(\cdot,u^*,v^*)$ auf $\mathbb{R}^n$, und (u^*,v^*) ist Maximum von $L(x^*,\cdot,\cdot)$ auf $\mathbb{R}^p_{\ge 0} \times \mathbb{R}^m$.

(b) Falls es $(x^*,u^*,v^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ mit (6.1) gibt, so ist x^* Lösung von (P) und (u^*,v^*) Lösung von (D).

Beweis: (a) Sei u^*, v^* wie im starken Dualitätssatz. Für $x \in \mathbb{R}^n$ $u \in \mathbb{R}^p$ und $v \in \mathbb{R}^m$ ist dann

$$\begin{aligned} L(x^*,u,v) \;&=\; f(x^*) \;+\; \underbrace{u^\top}_{\ge 0}\underbrace{g(x^*)}_{\le 0} + v^\top(\underbrace{Ax^*-b}_{=0}) \;\le\; f(x^*) \;=\; L(x^*,u^*,v^*) \\ &\le\; f(x) \;+\; {u^*}^\top g(x) \;+\; {v^*}^\top(Ax-b) \;=\; L(x,u^*,v^*)\,. \end{aligned}$$

(b) Aus der Ungleichung

$$L(x^*, u, v) \;=\; f(x^*) \;+\; u^\top g(x^*) \;+\; v^\top(Ax^* - b) \;\le\; L(x^*, u^*, v^*)$$

für alle $u \ge 0$ und $v \in \mathbb{R}^m$ folgt $Ax^* = b$ und $g(x^*) \le 0$, also $x^* \in M$. Für $u = 0$ folgt weiter $f(x^*) \le f(x^*) + u^{*\top} g(x^*) \le f(x^*)$, also $u^{*\top} g(x^*) = 0$. Aus der rechten Ungleichung folgt

$$L(x^*, u^*, v^*) \;=\; f(x^*) \;\le\; L(x, u^*, v^*) \;=\; f(x) \;+\; u^{*\top} g(x) \;+\; v^{*\top}(Ax - b)$$

für alle $x \in \mathbb{R}^n$, also

$$f(x^*) \;\le\; F(u^*, v^*) \;=\; \inf_{x \in \mathbb{R}^n} \left[f(x) + u^{*\top} g(x) + v^{*\top}(Ax - b) \right].$$

Der schwache Dualitätssatz liefert wieder die Behauptung. □

Schon im Vorgriff auf das nächste Kapitel können wir aus dem Sattelpunktsatz eine Lagrangesche Multiplikatorenregel ableiten. Dazu benötigen wir die folgende notwendige Optimalitätsbedingung, die für $n = 1$ schon aus der Schule bekannt ist:

Lemma 6.15 *Sei $U \subseteq \mathbb{R}^n$ offen (!), $f : U \to \mathbb{R}$ differenzierbar, und $x^* \in U$ ein lokales Minimum von f auf U. Dann ist $\nabla f(x^*) = 0$.*

Beweis: Sei $h \in \mathbb{R}^n$ ein beliebiger, festgehaltener Einheitsvektor. Setze $\psi(t) := f(x^* + th)$, $|t| \le \varepsilon$, wobei $\varepsilon > 0$ so klein gewählt ist, dass zum einen $x^* + th \in U$ und zum anderen $f(x^*) \le f(x^* + th)$ für alle $|t| \le \varepsilon$ gilt. Daher ist $t = 0$ Minimum von ψ auf dem Intervall $[-\varepsilon, \varepsilon] \subseteq \mathbb{R}$. Also ist $\psi'(0) = 0$, also (Kettenregel!) $\nabla f(x^*)^\top h = 0$. Da dies für alle Einheitsvektoren h gilt, so folgt $\nabla f(x^*) = 0$. □

Angewandt auf die Lagrangefunktion haben wir:

Korollar 6.16 *Es seien $f : \mathbb{R}^n \to \mathbb{R}$, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, p$, konvex und differenzierbar, und die Slaterbedingung (SB) sei erfüllt. Es ist $x^* \in M$ genau dann eine Lösung von (P), wenn es $u^* \in \mathbb{R}^p$ mit $u^* \ge 0$ und $v^* \in \mathbb{R}^m$ gibt mit*

$$\nabla f(x^*) \;+\; \sum_{j=1}^{p} u_j^* \nabla g(x^*) \;+\; A^\top v^* \;=\; 0 \tag{6.2a}$$

und

$$u_j^* g_j(x^*) \;=\; 0 \quad \textit{für alle } j = 1, \ldots, p\,. \tag{6.2b}$$

Beweis: Ist $x^* \in M$ eine Lösung, so besagt der Sattelpunktsatz, dass x^* ein Minimum von $L(\cdot, u^*, v^*)$ auf $\mathbb{R}^n$ ist. Also ist nach dem vorigen Lemma $\nabla_x L(x^*, u^*, v^*) = 0$, d.h. (6.2a).
Es gebe jetzt umgekehrt ein $u^* \in \mathbb{R}^p$ mit $u^* \ge 0$ und ein $v^* \in \mathbb{R}^m$ mit (6.2a) und (6.2b). Dies impliziert $\nabla_x L(x^*, u^*, v^*) = 0$. Nach Satz 6.3 gilt, da $L(\cdot, u^*, v^*)$ konvex ist:

$$\begin{aligned}
L(x, u^*, v^*) \;&\ge\; L(x^*, u^*, v^*) \;+\; \underbrace{\nabla_x L(x^*, u^*, v^*)^\top}_{=0}(x - x^*) \\
&=\; L(x^*, u^*, v^*) \;=\; f(x^*) \\
&\ge\; f(x^*) \;+\; \underbrace{u^\top}_{\ge 0}\, \underbrace{g(x^*)}_{\le 0} \;+\; v^\top \underbrace{(Ax^* - b)}_{=0} \\
&=\; L(x^*, u, v) \quad \text{für alle } x \in \mathbb{R}^n\,,\; u \ge 0\,,\; v \in \mathbb{R}^m\,.
\end{aligned}$$

Mit dem Sattelpunktsatz folgt die Behauptung. □

Beispiel 6.17 400 m^3 Kies sollen von einem Ort zu einem anderen transportiert werden. Dies geschehe durch mehrmaligen Transport in einer oben offenen Box der Länge t_1, der Breite t_2 und der Höhe t_3. Der Boden und die beiden Längsseiten müssen aus einem Material hergestellt werden, das zwar nichts kostet, von dem aber nur 4 m^2 zur Verfügung steht. Das Material für die beiden Querseiten kostet 200 DM pro m^2. Ein Transport der Box kostet 1 DM. Wie hat man die Box zu konstruieren? **Lösung:** Damit in die zu konstruierende Box überhaupt etwas hineingetan werden kann, hat man die Nebenbedingungen $t_1, t_2, t_3 > 0$. Dadurch, dass von dem Material für den Boden und für die beiden Längsseiten nur 4 m^2 zur Verfügung stehen, hat man noch die Restriktion $t_1t_2 + 2t_1t_3 \leq 4$. Die Herstellung einer Box der Länge t_1, der Breite t_2 und der Höhe t_3 (in Metern) kostet $400\, t_2\, t_3$ (in DM), als Transportkosten hat man ferner $400\, t_1^{-1}t_2^{-1}t_3^{-1}$ (in DM), denn so oft muss man fahren! Insgesamt ist also das Optimierungsproblem zu lösen:

$$\text{Minimiere} \quad t_1^{-1}t_2^{-1}t_3^{-1} + t_2\, t_3 \quad \text{unter den Nebenbedingungen}$$
$$\frac{1}{4}t_1t_2 + \frac{1}{2}t_1t_3 \leq 1 \quad \text{und} \quad t_1, t_2, t_3 > 0\,.$$

Probleme solcher Form nennt man auch **geometrische** Probleme. Diese kann man umformen, indem man $t_j = \exp(x_j)$ setzt. Dann erhalten wir

$$\text{Minimiere} \quad e^{-x_1-x_2-x_3} + e^{x_2+x_3} \quad \text{unter} \quad e^{x_1+x_2} + 2\, e^{x_1+x_3} \;\leq\; 4\,.$$

Die Zielfunktion und die Restriktionsfunktion

$$f(x) \;:=\; e^{-x_1-x_2-x_3} + e^{x_2+x_3} \quad \text{und} \quad g(x) \;:=\; e^{x_1+x_2} + 2\, e^{x_1+x_3} - 4$$

sind beide konvex, da $x \mapsto \exp(a^\top x)$ konvex ist für jedes $a \in \mathbb{R}^3$!

Die Lagrangefunktion hat die Form:

$$L(x,u) \;=\; e^{-x_1-x_2-x_3} \;+\; e^{x_2+x_3} \;+\; u\left(e^{x_1+x_2} + 2\, e^{x_1+x_3} - 4\right)\,.$$

Wir müssen jetzt also die Bedingungen $\nabla_x L(x,u) = 0$ und $u\, g(x) = 0$ ausschlachten. Wir rechnen diese Bedingungen aus:

$$-e^{-x_1-x_2-x_3} \;+\; u\left(e^{x_1+x_2} + 2\, e^{x_1+x_3}\right) \;=\; 0 \tag{6.3a}$$
$$-e^{-x_1-x_2-x_3} \;+\; e^{x_2+x_3} \;+\; u\, e^{x_1+x_2} \;=\; 0 \tag{6.3b}$$
$$-e^{-x_1-x_2-x_3} \;+\; e^{x_2+x_3} \;+\; 2u\, e^{x_1+x_3} \;=\; 0 \tag{6.3c}$$
$$u\left(e^{x_1+x_2} + 2\, e^{x_1+x_3} - 4\right) \;=\; 0\,. \tag{6.3d}$$

Aus (6.3a) folgt $u \neq 0$, also reduzieren sich (6.3d) und (6.3a) auf

$$e^{x_1+x_2} + 2\, e^{x_1+x_3} \;=\; 4 \quad \text{und} \quad 4u \;=\; e^{-x_1-x_2-x_3}\,. \tag{6.3e}$$

Subtraktion von (6.3b) und (6.3c) liefert $e^{x_2} = 2\, e^{x_3}$. Dies wieder eingesetzt in die erste Gleichung von (6.3e) liefert $\boxed{e^{x_1+x_2} = 2.}$ Eingesetzt in (6.3a) liefert $4u = u\big(2 + 2e^{x_1+x_3}\big)$, d.h. $e^{x_1+x_3} = 1$, d.h. $x_3 = -x_1$. Damit liefert die zweite Gleichung von (6.3e) $\boxed{4u = e^{-x_2},}$ und (6.3b) hat die Form $4u = e^{x_2-x_1} + 2u$, d.h. $\boxed{2u = e^{x_2-x_1}.}$ Die eingerahmten Formeln liefern schnell $u = 1/4$, $x_1 = -x_3 = \ln 2$ und $x_2 = 0$, d.h. $(t_1, t_2, t_3) = (2, 1, 1/2)$.

7 Das quadratische Problem

In diesem Kapitel betrachten wir das folgende quadratische Optimierungsproblem:

$$(P) \quad \text{Minimiere} \quad f(x) \;:=\; x^\top Q x \;+\; c^\top x \quad \text{auf} \quad M \;=\; \left\{x \in \mathbb{R}^n : x \geq 0,\; Ax = b\right\},$$

wobei $Q \in \mathbb{R}^{n\times n}$ eine beliebige quadratische Matrix ist, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$.
Bemerkungen: (a) Wegen $x^\top Q x = x^\top Q^\top x$ ist

$$f(x) \;=\; x^\top Q x + c^\top x \;=\; \frac{1}{2}\, x^\top (Q + Q^\top) x + c^\top x \;=\; x^\top Q_{sym} x + c^\top x\,,$$

und $Q_{sym} = \frac{1}{2}(Q+Q^\top)$ ist offenbar symmetrisch. Also können wir uns ohne Einschränkung auf **symmetrische** Matrizen Q beschränken.
(b) Ist Q symmetrisch und positiv semidefinit, so ist die Zielfunktion konvex, und wir haben einen Spezialfall eines konvexen Problems vor uns. Für diesen gelten natürlich alle Aussagen für den konvexen Fall. Es gilt aber mehr bzw. unter schwächeren Voraussetzungen, wie wir gleich sehen werden.
(c) Restriktionenmengen der Form $\left\{x \in \mathbb{R}^n : Ax \leq b\right\}$ können wieder durch Aufspaltung $x = x^+ - x^-$ mit $x^\pm \geq 0$ und Einführung von Schlupfvariablen in die Form von M gebracht werden.

7.1 Ein Existenzsatz und der Satz von Kuhn-Tucker

Der folgende Existenzsatz kommt ohne eine Forderung an die Menge Λ aus und gilt auch ohne die Voraussetzung der Definitheit. Er entspricht genau dem Existenzsatz 3.1 der linearen Optimierung.

Satz 7.1 *Sei (P) zulässig und $\inf(P) > -\infty$. Dann besitzt (P) eine Lösung.*

Beweis: Da $M \neq \emptyset$, so existiert $\rho_0 > 0$ mit $M_\rho := \left\{x \in M : \|x\| \leq \rho\right\} \neq \emptyset$ für alle $\rho \geq \rho_0$. Die Probleme

$$(P_\rho) \qquad \text{Minimiere} \quad f(x) \quad \text{auf} \quad M_\rho$$

sind für $\rho \geq \rho_0$ alle lösbar, da M_ρ kompakt ist. Es ist $\min(P_\rho)$ monoton nichtsteigend und $\lim\limits_{\rho\to\infty} \min(P_\rho) = \inf(P)$ (man versuche dies zu beweisen). Sei $x^\rho \in M_\rho$ die Lösung mit minimaler euklidischer Norm von (P_ρ) (diese existiert, da die Menge der optimalen Lösungen kompakt ist. Man beweise dies.) Wir zeigen jetzt:

(1) Es gibt $\rho^* \geq \rho_0$ mit $\|x^\rho\| < \rho$ für alle $\rho \geq \rho^*$

(2) $\min(P_\rho) = \inf(P)$ für ein $\rho \geq \rho^*$, d.h. x^ρ ist Lösung von (P).

Beweis von (1): Andernfalls gäbe es Folge $\rho_k \to \infty$ mit $\|x^{\rho_k}\| = \rho_k$ für jedes k. Setze $y^k := x^{\rho_k}/\rho_k$ für $k \in \mathbb{N}$. Dann gibt es wegen $\|y^k\| = 1$ eine konvergente Teilfolge von (y^k), o.B.d.A: $y^k \to y$, $k \to \infty$, und $y \geq 0$, $\|y\| = 1$, $Ay = 0$, da $Ay^k = \frac{1}{\rho_k} b \to 0$. Außerdem folgt $y^\top Q y = 0$, denn:

$$-\infty \;<\; \inf(P) \;\leq\; f(x^{\rho_k}) \;=\; \frac{1}{2}\rho_k^2\, (y^k)^\top Q y^k \;+\; \rho_k\, c^\top y^k \;\leq\; f(x^{\rho_0})$$

für $k \in \mathbb{N}$ (da $M_{\rho_0} \subseteq M_{\rho_k}$ und x^{ρ_k} optimal für (P_{ρ_k}) ist). Division durch ρ_k^2 liefert $y^\top Q y = 0$. Daher ist für jedes $k \in \mathbb{N}$ und $t > 0$ zunächst $x^{\rho_k} + ty \in M$ und dann

$$f(x^{\rho_k} + ty) \;=\; f(x^{\rho_k}) \;+\; t\left[2\, y^\top Q x^{\rho_k} + c^\top y\right].$$

Es ist $2\, y^\top Q x^{\rho_k} + c^\top y \geq 0$ für alle k (sonst wäre $f(x^{\rho_k} + ty) \to -\infty$, $t \to +\infty$, ein Widerspruch). Wir zeigen jetzt, dass auch folgendes gilt:

(i) $x^{\rho_k} - ty \in M$,

(ii) $\|x^{\rho_k} - ty\| < \|x^{\rho_k}\|$,

(iii) $f(x^{\rho_k} - ty) = f(x^{\rho_k})$

für große k und kleine $t > 0$. Dies wäre dann ein Widerspruch zur Wahl von x^{ρ_k}.

Zu (i): Es ist $A(x^{\rho_k} - ty) = b$ wegen $Ay = 0$. Sei $J = \big\{j \in \{1, \ldots, n\} : y_j = 0\big\}$. Für $j \in J$ ist $(x^{\rho_k} - ty)_j \geq 0$ für alle $k \in \mathbb{N}$ und $t \geq 0$. Sei $\varepsilon := \min\{y_j : j \notin J\} \in (0, 1]$. Wähle k so groß, dass $y_j^k \geq \frac{\varepsilon}{2}$ für alle $j \notin J$. Dann gilt

$$(x^{\rho_k} - ty)_j \;=\; \rho_k\, y_j^k - t y_j \;\geq\; \rho_k \frac{\varepsilon}{2} - t \;\geq\; 0 \quad \text{für alle } t \in \Big(0, \frac{\varepsilon}{2}\rho_k\Big).$$

Zu (ii): Wähle außerdem k so, dass $y^\top y^k \geq \frac{1}{2}$. Dann ist

$$\|x^{\rho_k} - ty\|^2 \;=\; \|x^{\rho_k}\|^2 - t\big[2\rho_k\, y^\top y^k - t\|y\|^2\big] \;<\; \|x^{\rho_k}\|^2 \quad \text{für kleine } t > 0\,.$$

Zu (iii): Da $x^{\rho_k} - ty \in M_{\rho_k}$, so für diese k und $t > 0$:

$$f(x^{\rho_k}) \;\leq\; f(x^{\rho_k} - ty) \;=\; f(x^{\rho_k}) \;-\; t\left[2\, y^\top Q x^{\rho_k} + c^\top y\right] \;\leq\; f(x^{\rho_k})\,,$$

also Gleichheit!
Damit sind (i), (ii), (iii) bewiesen und damit die Behauptung (1).

Beweis von (2): Angenommen, $\inf(P) < \min(P_\rho)$ für alle $\rho \geq \rho^*$. Wegen $\min(P_\rho) \searrow \inf(P)$, $\rho \to \infty$), existiert $\rho_2 > \rho_1 \geq \rho^*$ mit $f(x^{\rho_1}) > f(x^{\rho_2}) > \inf(P)$ und $\|x^{\rho_j}\| < \rho_j$ für $j = 1, 2$. Dann ist $\rho_1 < \|x^{\rho_2}\| < \rho_2$, denn sonst wäre $x^{\rho_2} \in M_{\rho_1}$ mit kleinerem Funktionswert. Setze $\rho := \|x^{\rho_2}\| \in (\rho_1, \rho_2)$. Dann ist $\|x^\rho\| < \rho$ und $f(x^{\rho_2}) \leq f(x^\rho)$, da $x^\rho \in M_{\rho_2}$.
1. Fall: $f(x^{\rho_2}) = f(x^\rho)$, d.h. x^ρ ist optimal für (P_{ρ_2}). Dies ist ein Widerspruch wegen $\|x^\rho\| < \rho = \|x^{\rho_2}\|$.
2. Fall: $f(x^{\rho_2}) < f(x^\rho)$. Es ist $x^{\rho_2} \in M_\rho$, also ein Widerspruch zur Optimalität von x^ρ.
Damit ist der Beweis schließlich beendet. □

Wir kommen jetzt zu dem Satz von **Kuhn-Tucker** für quadratische Probleme. Bei linearen und konvexen Problemen ist dieser Teil der starken Dualitätssätze und kann auch als **Lagrangesche Multiplikatorenregel** bezeichnet werden. Wir formulieren diese noch einmal als eigenständige Sätze.

Satz 7.2 *Sei $x^* \in M$ Lösung des linearen Problems*

$$(LP) \qquad \textit{Minimiere} \quad c^\top x \quad \textit{auf } M = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}\,.$$

Dann existiert ein **Lagrange Multiplikator**, *d.h. $y^* \in \mathbb{R}^m$ mit $c + A^\top y^* \geq 0$. Außerdem gilt die Gleichgewichtsbedingung $(c + A^\top y^*)^\top x^* = 0$.*

Beweis: Aus der Existenz von x^* folgt insbesondere, dass (LP) zulässig ist und dass $\inf(LP) > -\infty$. Aus Teil (ii) des starken Dualitätssatzes 3.5 folgt die Zulässigkeit des dualen Problems, also aus Teil (i) auch die Existenz $\hat{y}$ einer Lösung des dualen Problems. Die Zulässigkeit liefert $c + A^\top \hat{y} \geq 0$. Die Folgerung aus Satz 3.5 liefert die Gleichgewichtsbedingung $(A^\top \hat{y} - c)^\top x^* = 0$. Mit $y^* = -\hat{y}$ folgt die Behauptung. □

Wir erinnern an das Korollar 6.16:

Sei $x^* \in M$ Lösung des konvexen Problems

$$(KP) \qquad \text{Minimiere} \quad f(x) \quad \text{auf } M = \{x \in \mathbb{R}^n : g(x) \leq 0,\ Ax = b\}\,,$$

und die Slaterbedingung (SB) sei erfüllt. Es ist $x^* \in M$ genau dann eine Lösung von (P), wenn es $u^* \in \mathbb{R}^p$ mit $u^* \geq 0$ und $v^* \in \mathbb{R}^m$ gibt mit

$$\nabla f(x^*) \ + \ \sum_{j=1}^{p} u_j^* \nabla g_j(x^*) \ + \ A^\top v^* \ = \ 0\,, \quad \text{d.h.} \quad \nabla f(x^*) \ + \ g'(x^*)^\top u^* \ + \ A^\top v^* \ = \ 0$$

und $g(x^*)^\top u^* = 0$. Hier ist $g'(x^*) \in \mathbb{R}^{p \times n}$ die Funktionalmatrix von g an der Stelle x^*.

Für quadratische Probleme kommt er ohne die Slaterbedingung aus.

Satz 7.3 *(Kuhn-Tucker)*
Gegeben sei das quadratische Optimierungsproblem (P).

(a) Sei $x^* \in M$ Lösung von (P). Dann gibt es $y^* \in \mathbb{R}^m$ mit

(i) $2Qx^* + c + A^\top y^* \geq 0$ und
(ii) $(2Qx^* + c + A^\top y^*)^\top x^* = 0$.

(b) Sei Q symmetrisch und positiv semidefinit, $x^* \in M$, und es gebe $y^* \in \mathbb{R}^m$ mit (i), (ii). Dann ist x^* Lösung von (P).

Bemerkung: (x^*, y^*) ist dann jeweils die Lösung von (D), denn aus (ii) folgt

$$2\,x^{*\top} Q x^* \ + \ c^\top x^* \ = \ x^{*\top} A^\top y^* \ = \ b^\top y^*$$

also

$$f(x^*) \ = \ x^{*\top} Q x^* \ + \ c^\top x^* \ = \ b^\top y^* - x^{*\top} Q x^*$$

und $(x^*, y^*) \in N$.

Beweis: für beliebige $x, x^* \in \mathbb{R}^n$ gilt die „binomische Formel“ in der Form

$$\begin{aligned} &[x^\top Q x + c^\top x] - [x^{*\top} Q x^* + c^\top x^*] \\ &\quad = \ 2x^{*\top} Q(x - x^*) \ + \ c^\top (x - x^*) \ + \ (x - x^*)^\top Q (x - x^*)\,, \end{aligned}$$

also

$$f(x) \ = \ f(x^*) \ + \ [2Qx^* + c]^\top (x - x^*) \ + \ (x - x^*)^\top Q (x - x^*)\,. \qquad (*)$$

(a) Wir gehen so vor wie auch später im allgemeinen nichtlinearen Fall und **linearisieren** das Problem. Für $x \sim x^*$ ist der letzte Teil „quadratisch klein“. Betrachte daher die linearisierte Aufgabe:

$$(LP) \qquad \text{Minimiere} \quad [2Qx^* + c]^\top h \quad \text{unter} \quad Ah = 0 \text{ und } h_j \geq 0 \text{ für alle } j \in J\,,$$

wobei $J = \big\{j \in \{1,\ldots,n\} : x_j^* = 0\big\}$ die Menge der aktiven Indizes ist. Wir zeigen, dass $h^* = 0$ optimal für (LP) ist. Natürlich ist h^* zulässig. Sei $h \in \mathbb{R}^n$ mit $Ah = 0$ und $h_j \geq 0$ für alle $j \in J$ und $[2Qx^* + c]h < 0$. Für kleine $t > 0$ ist $x^* + th \in M$ (unterscheide $j \in J$ und $j \notin J$!)und wegen $(*)$ für $x = x^* + th$ ist

$$f(x^* + th) \;= f(x^*) \;+\; t\,\big[\underbrace{(2Qx^* + c)^\top h}_{<0} \;+\, t\,h^\top Q h\big]\,.$$

Für hinreichend kleine t ist $f(x^* + th) < f(x^*)$, ein Widerspruch zur Optimalität von x^*. Also ist $h^* = 0$ optimal für (LP). Nun besagt der starke Dualitätssatz der linearen Theorie, dass es es eine Lösung $y^* \in \mathbb{R}^m$ des dualen Problems zu (LP) gibt. Wie erhalten wir das duale Problem? Ist $J = \{j_1,\ldots,j_q\}$, so setzen wir $\jmath = (j_1,\ldots,j_q)$ und wählen ein Komplement $k \in \{1,\ldots,m\}^{n-q}$. Wir zerlegen wieder A, h, c u.s.w. bezüglich $\jmath$ und k. Insbesondere zerlegen wir $h(k) = h^+(k) - h^-(k)$ mit $h^\pm(k) \geq 0$. Das Problem (LP) hat dann die Form

$$\text{Minimiere} \quad \big[\,(2Qx^* + c)(\jmath)^\top \,|\, (2Qx^* + c)(k)^\top \,|\, -(2Qx^* + c)(k)^\top\,\big] \begin{bmatrix} h(\jmath) \\ h^+(k) \\ h^-(k) \end{bmatrix}$$

$$\text{unter} \quad \begin{bmatrix} h(\jmath) \\ h^+(k) \\ h^-(k) \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{und} \quad \big[\,A(\jmath) \,|\, A(k) \,|\, -A(k)\big] \begin{bmatrix} h(\jmath) \\ h^+(k) \\ h^-(k) \end{bmatrix} \;=\; 0\,.$$

Der Satz 7.2 von Kuhn-Tucker der linearen Theorie liefert dann die Existenz von $\hat{y} \in \mathbb{R}^m$ mit

$$\begin{bmatrix} A(\jmath)^\top \\ A(k)^\top \\ -A(k)^\top \end{bmatrix} \hat{y} \;\leq\; \begin{bmatrix} (2Qx^* + c)(\jmath) \\ (2Qx^* + c)(k) \\ -(2Qx^* + c)(k) \end{bmatrix},$$

d.h.

$$A(\jmath)^\top \hat{y} \;\leq\; 2(Qx^*)(\jmath) \;+\; c(\jmath) \quad \text{und} \quad A(k)^\top \hat{y} \;=\; 2(Qx^*)(k) \;+\; c(k)$$

sowie der Gleichgewichtsbedingung $\big[2Qx^* + c - A^\top \hat{y}\big]^\top x^* = 0$. Mit $y^* = -\hat{y}$ folgt die Behauptung!

(b) Aus $(*)$ folgt für beliebiges $x \in M$ (mit (i) und $x \geq 0$ und (ii)):

$$\begin{aligned} f(x) \;&\geq\; f(x^*) \;+\; [2Qx^* + c]^\top x \;-\; [2Qx^* + c]^\top x^* \\ &\geq\; f(x^*) \;-\; [A^\top y^*]^\top x \;+\; [A^\top y^*]^\top x^* \\ &=\; f(x^*) - \big[A(x - x^*)\big]^\top y^* \;=\; f(x^*)\,. \end{aligned}$$

Damit ist der Satz bewiesen. □

Beispiel 7.4 Unter einem **linearen Komplementaritätsproblem** versteht man bei gegebenen $c \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n\times n}$ die folgende Aufgabe:

$$(P) \qquad \text{Bestimme } x \in \mathbb{R}^n \text{ mit } \quad x \geq 0\,, \quad c + Qx \geq 0\,, \quad \text{und} \quad x^\top (x + Qx) = 0\,.$$

Wir zeigen: Ist $Q \in \mathbb{R}^{n\times n}$ symmetrisch und positiv semidefinit und existiert ein $\hat{x} \in \mathbb{R}^n$ mit $c + Q\hat{x} \geq 0$, so hat das lineare Komplementaritätsproblem (P) eine Lösung.

Beweis: Man betrachte das quadratische Problem

$$(P_0) \qquad \text{Minimiere} \quad f(x) := c^\top x + \frac{1}{2} x^\top Q x \quad \text{unter} \quad x \geq 0\,.$$

Nach Satz 7.1 besitzt (P_0) eine Lösung, falls (P_0) zulässig (dies ist trivial) und $\inf(P_0) > -\infty$ ist. Wir müssen also diese letzte Bedingung zeigen.
Sei $x \geq 0$ beliebig und $\hat{x} \in \mathbb{R}^n$ ein nach Voraussetzung existierendes Element mit $c + Q\hat{x} \geq 0$. Dann ist

$$\begin{aligned} f(x) - f(\hat{x}) &= c^\top x \;+\; \frac{1}{2} x^\top Q x \;-\; c^\top \hat{x} \;-\; \frac{1}{2} \hat{x}^\top Q \hat{x} \\ &= \underbrace{(c + Q\hat{x})^\top}_{\geq 0} (x - \hat{x}) \;+\; \frac{1}{2} \underbrace{(x - \hat{x})^\top Q (x - \hat{x})}_{\geq 0} \\ &\geq -(c + Q\hat{x})^\top \hat{x}\,. \end{aligned}$$

Daher ist $\inf(P) \geq f(\hat{x}) - (c + Q\hat{x})^\top \hat{x} > -\infty$, und (P_0) besitzt Lösung x^*. Der Satz von Kuhn-Tucker liefert also

$$c + Qx^* \;\geq\; 0 \quad \text{und} \quad x^\top (c + Qx^*) \;=\; 0\,.$$

□

Als Abschluss dieses Abschnitts wollen wir noch die entsprechenden Sätze für das quadratische Optimierungsproblem in zweiter Standardform erwähnen. Sei also jetzt (P_2) gegeben als:

$$(P_2) \qquad \text{Minimiere} \quad f(x) \;:=\; x^\top Q x \;+\; c^\top x \quad \text{auf} \quad M \;=\; \left\{ x \in \mathbb{R}^n : Ax \leq b \right\},$$

mit den gleichen Voraussetzungen wie für (P). Der Existenzsatz 7.1 ist wörtlich zu übernehmen. Der Satz von Kuhn-Tucker hat die Form:

Satz 7.5 *(Kuhn-Tucker)*
Gegeben sei das quadratische Optimierungsproblem in der Form (P_2).

(a) Sei $x^* \in M$ Lösung von (P_2). Dann gibt es $u^* \in \mathbb{R}^m$ mit $u^* \geq 0$ und

(i) $2Qx^* + c + A^\top u^* = 0$ sowie
(ii) $(b - Ax^*)^\top u^* = 0$.

(b) Sei Q symmetrisch und positiv semidefinit, $x^* \in M$, und es gebe $y^* \in \mathbb{R}^m$ mit $y^* \geq 0$ und (i), (ii). Dann ist x^* Lösung von (P_2).

7.2 Das Verfahren von Goldfarb-Idnani

Wir betrachten jetzt das Problem

$$(P) \qquad \text{Minimiere} \quad f(x) \;:=\; x^\top Q x + c^\top x \quad \text{auf } M \;=\; \big\{x \in \mathbb{R}^n : Ax \le b\big\}$$

und machen die Voraussetzungen: $Q \in \mathbb{R}^{n\times n}$ sei symmetrisch und positiv definit (d.h. es gibt $\gamma > 0$ mit $x^\top Q x \ge \gamma \|x\|^2$ für alle $x \in \mathbb{R}^n$), $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$.

a_{i*} seien jetzt wieder die Zeilenvektoren von A, also $A = \begin{bmatrix} a_{1*} \\ \vdots \\ a_{m*} \end{bmatrix}$.

Dann ist $A^\top y = \sum_{i=1}^m y_i a_{i*}^\top$.

Das Verfahren von Goldfarb-Idnani ist ein **duales** Verfahren, d.h. es erzeugt dual-zulässige Paare (x^k, u^k), wobei x^k aber nicht primal-zulässig sind (d.h. es gilt nicht $Ax^k \le b$). Nach Aufgabe 29 werden dual-zulässige Paare (x^k, u^k) beschrieben durch $u^k \ge 0$ und $2Qx^k + c + A^\top u^k = 0$.

Man startet mit dem unrestringierten Problem, löst dieses und nimmt nach und nach weitere Restriktionen hinzu. Für die genauere Beschreibung benötigen wir die folgenden Hilfsprobleme:

Sei $I \subseteq \{1, \dots, m\}$ eine beliebige Teilmenge. Betrachte dann

$$(P_I) \qquad \text{Minimiere } f(x) \;=\; x^\top Q x + c^\top x \quad \text{auf } M_I := \big\{x \in \mathbb{R}^n : (Ax)_i \le b_i \;\forall i \in I\big\}\,.$$

Die Probleme (P) und (P_I) sind alle eindeutig lösbar nach Satz 6.7. Wegen der Sätze von Kuhn und Tucker sind $x^* \in M$ und $\overline{x} \in M_I$ die optimalen Lösungen von (P) bzw. (P_I) genau dann, wenn es $u^* \in \mathbb{R}^m$ bzw. $\overline{u} \in \mathbb{R}^{|I|}$ gibt mit

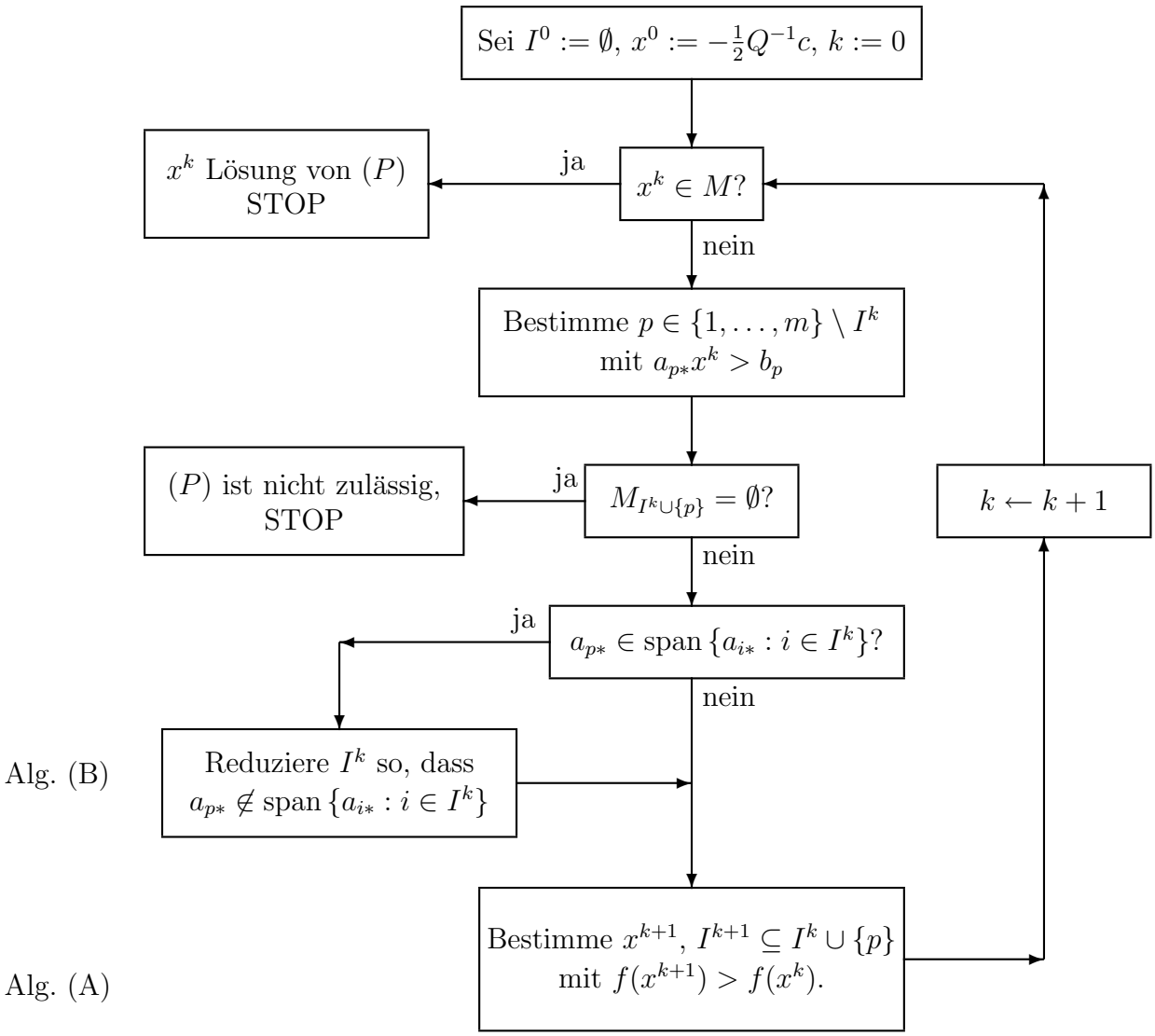
$$u^* \ge 0\,, \quad c + 2Qx^* + A^\top y^* = 0\,, \quad u_i^*\big[(Ax^*)_i - b_i\big] = 0 \text{ für alle } i = 1, \dots, m\,,$$

$$\text{bzw.} \qquad \overline{u} \ge 0\,, \quad c + 2Q\overline{x} + \sum_{i\in I} \overline{u}_i a_{i*}^\top = 0\,, \quad \overline{u}_i\big[(A\overline{x})_i - b_i\big] = 0 \text{ für alle } i \in I\,.$$

Wir nennen u^* und $\overline{u}$ in Zukunft wieder **Multiplikatoren**.

Das Verfahren startet mit $I^0 = \emptyset$ und erzeugt eine Folge (I^k) so, dass $\min(P_{I^k}) < \min(P_{I^{k+1}})$ (aber nicht notwendig $I^k \subseteq I^{k+1}$). Es hält den zugehörigen Multiplikator nichtnegativ. Wegen der Vergrößerung des Minimalwertes kann kein Problem zweimal angelaufen werden, und das Verfahren muss daher nach endlich vielen Schritten stoppen.

Algorithmus von Goldfarb-Idnani:



Bevor wir mit der Beschreibung von Algorithmus (A) beginnen, formulieren wir die folgende **Voraussetzung (V)**, die nach der Bestimmung von p vorliegt:

(V): $x^k \in M_{I^k}$ sei optimal für (P_{I^k}) mit zugehörigem Multiplikator $u^k \geq 0$. Ferner sei die p-te Restriktion verletzt, d.h. mit $a := a_{p*}^\top$ und $\beta := b_p$ gelte $a^\top x^k > \beta$. Schließlich seien die Zeilen $\{a_{i*} : i \in I^k\}$ linear unabhängig.

Im Fall $k = 0$ ist diese Voraussetzung sicher erfüllt.

Für den Algorithmus (A) sei zusätzlich noch

$$(\mathbf{V_A}) \qquad a \notin \text{span}\{a_{i*}^\top : i \in I^k\}$$

vorausgesetzt. Betrachte jetzt den Algorithmus (A):

(A0) Setze $\theta := 0$, $f := f(x^k)$, $x := x^k$, $I := I^k$, $u := u^k \in \mathbb{R}^{|I^k|}$.

(A1) Schreibe I in der Form $I = \{i_1, \ldots, i_q\}$ mit $q = |I|$ und definiere $A_I \in \mathbb{R}^{|I| \times n}$ durch

$$A_I^\top \;:=\; \left[a_{i_1*}^\top, \ldots, a_{i_q*}^\top\right] \in \mathbb{R}^{n \times |I|}\,.$$

(A2) Berechne $d := (A_I Q^{-1} A_I^\top)^{-1} A_I Q^{-1} a$ (falls $I \neq \emptyset$, sonst $d := 0$), $z := \frac{1}{2} Q^{-1}(a - A_I^\top d)$ und $t_1 := (a^\top x - \beta)/(a^\top z)$.

(A3a) Falls $u - t_1 d \geq 0$ oder $I = \emptyset$, so setze

$$\begin{array}{rclcrcl} x^{k+1} & := & x - t_1 z\,, & \qquad & u^{k+1} & := & \begin{bmatrix} u - t_1 d \\ \theta + t_1 \end{bmatrix} \in \mathbb{R}^{|I|+1}, \\ I^{k+1} & := & I \cup \{p\}\,, & & f^{k+1} & := & f + t_1(\theta + t_1/2)\, a^\top z, \end{array}$$

und weiter im äußeren Algorithmus mit $\boxed{k \leftarrow k+1}$.

(A3b) Falls $u - t_1 d \not\geq 0$ und $I \neq \emptyset$, so setze $t_2 := \min\{u_i/d_i : d_i > 0, i \in I\} = u_\ell/d_\ell$ mit $d_\ell > 0$ und $\ell \in I$, und

$$\begin{array}{rclcrcl} x^- & := & x - t_2 z\,, & & I^- & := & I \setminus \{\ell\}\,, \\ u_i^- & := & u_i - t_2 d_i \text{ für } i \in I^-\,, & & f^- & := & f + t_2(\theta + t_2/2)\, a^\top z\,, \\ \theta^- & := & \theta + t_2\,. & & & & \end{array}$$

(A4) Update $(x, u, f, \theta, I) := (x^-, u^-, f^-, \theta^-, I^-)$ und weiter mit (A1).

Lemma 7.6 *Es gelte die Voraussetzungen (V) und (V_A). Dann ist der Algorithmus (A) wohldefiniert und liefert in endlich vielen Schritten die Lösung x^{k+1} von ($P_{I^{k+1}}$) mit Multiplikator $u^{k+1} \geq 0$. Ferner ist $f(x^{k+1}) = f^{k+1} > f(x^k)$, und* $\operatorname{Rang} A_{I^{k+1}} = |I^{k+1}|$.

Beweis: Der Algorithmus (A) ist ebenfalls iterativ aber endlich, da man im Schritt (A3b) von I nur endlich oft ein Element entfernen kann. Wir nehmen an, dass[4] in einem Schritt des Algorithmus folgendes gelte:
(∗)
$$A_I x \leq b_I\,, \quad a^\top x > \beta\,, \quad c + 2Qx + A_I^\top u + \theta a = 0\,, \quad u \geq 0\,, \quad \theta \geq 0\,, \quad f = f(x)\,.$$

Beim Start von Algorithmus (A) ist dies nach den Voraussetzungen (V) für $\theta = 0$ erfüllt! Außerdem ist a von den Zeilen $\{a_{i*}^\top : i \in I\}$ linear unabhängig.

Es ist $z \neq 0$, da sonst $a = A_I^\top d$, und würde ja bedeuten, dass a doch von den Zeilen $\{a_{i*}^\top : i \in I\}$ linear abhängig wäre. Außerdem ist im Fall $I \neq \emptyset$: $A_I z = \frac{1}{2}\left[A_I Q^{-1} a - A_I Q^{-1} A_I^\top d\right] = 0$, und daher $z^\top(a - 2Qz) = z^\top A_I^\top d = 0$, also ist $z^\top a = 2z^\top Q z > 0$. Ist $I = \emptyset$, so ist $d = 0$, d.h. $2Qz = a$ und ebenfalls $z^\top a = 2z^\top Q z > 0$. Daher sind die Schritte (A1) und (A2) wohldefiniert und $t_1 > 0$.

[4] mit den Bezeichnungen im Algorithmus

Setze $x(t) := x - tz$. Dann ist $A_I x(t) = A_I x - tA_I z \leq b_I$, und $a^\top x(t) = a^\top x - ta^\top z = \beta + (t_1 - t)a^\top z$. Ferner ist $c + 2Qx(t) = c + 2Qx - 2tQz = -A_I^\top u - \theta a - t[a - A_I^\top d] = -A_I^\top(u - td) - (\theta + t)a$ und (nachrechnen!)

$$f(x - tz) \;=\; f(x) - t(c + 2Qx)^\top z + t^2 \underbrace{z^\top Q z}_{=\frac{1}{2}z^\top a} \;=\; f(x) \;+\; tz^\top(A_I^\top u + \theta a) \;+\; \frac{t^2}{2}\, z^\top a$$

$$=\; f(x) + t(\theta + t/2)\, z^\top a\,.$$

Fall (A3a): Ist $u - t_1 d \geq 0$, so ist (benutze obige Formeln für $t = t_1$): $x - t_1 z \in M_I$, $a^\top(x - t_1 z) = \beta$, $t_1 + \theta > 0$ und $c + 2Q(x - t_1 z) + A_I^\top(u - t_1 d) - (\theta + t_1)a = 0$, also ist $\left(x - t_1 z, \begin{bmatrix} u - t_1 d \\ t_1 + \theta \end{bmatrix}\right)$ ein Kuhn-Tucker Paar für $(P_{I\cup\{p\}})$ d.h. mit $x^{k+1} = x - t_1 z$ ist x^{k+1} Lösung von $(P_{I^{k+1}})$ mit Multiplikator $\begin{bmatrix} u - t_1 d \\ t_1 + \theta \end{bmatrix} \geq 0$. Außerdem ist $f(x^{k+1}) = f(x^k) + t_1(\theta + t_1/2)z^\top a > f(x^k)$.

Fall (A3b): Ist $u_i - t_1 d_i < 0$ für ein $i \in I$, so muss $d_i > 0$ sein, d.h. t_2 ist wohldefiniert und $0 \leq t_2 < t_1$. Es ist dann $u - t_2 d \geq 0$ und $u_\ell - t_2 d_\ell = 0$ nach Definition von ℓ. Außerdem ist $(x - t_2 z)^\top a = x^\top a - t_2 z^\top a > x^\top a - t_1 z^\top a = \beta$. Mit den obigen Formeln für $t = t_2$ sind x^-, u^-, f^-, θ^- und I^- wohldefiniert und genügen wieder der Ausgangssituation $(*)$. □

Der **Algorithmus (B)** behandelt den Fall, dass a_{p*} von den Zeilen der Matrix A_I linear abhängig ist. Wir setzen jetzt also zusätzlich zu (V) voraus, dass $a^\top = a_{p*}$ von den Zeilen der Matrix A_I abhängig ist. Getestet wird dies im Schritt (A2) des Algorithmus (A), siehe den Anfang des Beweises von Lemma 7.6. Ist nämlich $z = 0$, so ist $a = A_I^\top d$ mit $d = (A_I Q^{-1} A_I^\top)^{-1} A_I Q^{-1} a$. Wir betrachten also jetzt den folgenden Algorithmus, wobei wieder $a := a_{p*}^\top$, $\beta := b_p$ und $a^\top x^k > \beta$ gelte:

(B0) Setze $x := x^k$, $f := f^k$, $I := \{i \in I^k : (Ax)_i = b_i\}$, $u := u_I^k$.

(B1) Berechne $d := (A_I Q^{-1} A_I^\top)^{-1} A_I Q^{-1} a$.

(B2) Ist $d \leq 0$, so ist $(P_{I\cup\{p\}})$ und damit auch (P) nicht zulässig, STOP. Andernfalls:

(B3) Bestimme $\ell \in I$ mit $d_\ell > 0$ und

$$t_2 \;:=\; \min\left\{\frac{u_i}{d_i} : i \in I, d_i > 0\right\} \;=\; \frac{u_\ell}{d_\ell} \;\geq\; 0\,.$$

Setze $x^- := x$, $\quad I^- := I \setminus \{\ell\}$, $\quad u_i^- := u_i - t_2 d_i$, $i \in I^-$, $\quad \theta := t_2$.

(B4) Update: $(x, u, I) := (x^-, u^-, I^-)$ und weiter mit (A1).

Lemma 7.7 *Unter den obigen Voraussetzungen gilt:*
(a) Ist $d \leq 0$, so ist $M_{I\cup\{p\}} = \emptyset$.
(b) Ist $d_i > 0$ für ein i, so liefert der Algorithmus (B) ein 5-Tupel (x, u, f, θ, I), welches die Voraussetzungen $(*)$ erfüllt mit $\theta = t_2$.

Beweis: Es ist $a = A_I^\top d$, denn: Wegen $a \in \operatorname{span}\{a_{i*}^\top : i \in I\}$ existiert $\hat{d}$ mit $a = A_I^\top \hat{d}$. Dann ist aber $A_I Q^{-1} A_I^\top d = A_I Q^{-1} a = A_I Q^{-1} A_I^\top \hat{d}$, also $d = \hat{d}$.

(a) Angenommen, es gäbe $z \in \mathbb{R}^n$ mit $x + z \in M_{I \cup \{p\}}$. Dann muss $A_I z \leq 0$ wegen $A_I x = b_I$ gelten und außerdem $a^\top z \leq \beta - a^\top x < 0$. Andererseits ist aber $a = A_I^\top d$, also $z^\top a = z^\top A_I^\top d = (A_I z)^\top d \geq 0$, und dies ist ein Widerspruch!

(b) Es ist $A_{I^-} x^- = b_{I^-}$ und $c + Qx^- + A_{I^-}^\top u^- + \theta a = c + Qx + A_I^\top (u - t_2 d) + \theta a = \theta a - t_2 A_I^\top d = (\theta - t_2) a = 0$. Außerdem ist $u_i^- \geq 0$, $\forall i \in I^-$ und $a^\top x^- > \beta$ sowie $\theta \geq 0$. Schließlich ist $a = \sum_{i \in I^-} d_i a_{i*}^\top + d_\ell a_{\ell*}^\top$ mit $d_\ell \neq 0$. Wäre $a \in \operatorname{span}\{a_{i*}^\top : i \in I^-\}$, so würde auch $a_{\ell*}^\top \in \operatorname{span}\{a_{i*}^\top : i \in I^-\}$ sein, d.h. $\operatorname{Rang} A_I \leq |I| - 1$ gelten, ein Widerspruch! $\square$

Bemerkungen:
(a) z ist die orthogonale Projektion von $Q^{-1}a$ auf Kern A unter dem Skalarprodukt $\langle u, v \rangle := u^\top Q v$ für $u, v \in \mathbb{R}^n$. Dies gibt die Möglichkeit, stabile und effiziente Verfahren zur Berechnung von z zu benutzen.
(b) Wir haben uns in diesem Abschnitt auf die Originalarbeit (D. Goldfarb, A. Idnani: A Numerical Stable Dual Method for Solving Strictly Convex Quadratic Programs. Math. Progr. 27 (1983), 1–33.) gehalten.

8 Differenzierbare Optimierungsprobleme

Jetzt wollen wir Probleme studieren der Form:

$$(P) \qquad \text{Minimiere} \quad f(x) \quad \text{auf } M \;:=\; \big\{x \in \mathbb{R}^n : g(x) \le 0\,,\; h(x) = 0\big\}\,,$$

wobei $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar seien.
Ziel dieses Kapitels ist der Beweis der Lagrangeschen Multiplikatorenregel. Wir gehen in drei Schritten vor: Sei $x^* \in M$ eine (lokale) Lösung von (P).

(A) Wir linearisieren (P) um x^*. Dies führt auf eine *lineare Optimierungsaufgabe* (LP) vom Typ, wie wir sie im ersten Kapitel behandelt haben.

(B) Wir zeigen, dass 0 eine Lösung von (LP) ist.

(C) Wir wenden den Satz von Kuhn-Tucker auf (LP) mit Lösung 0 an.

Die Lösung (u^*, v^*) des dualen Problems zu (LP) enthält dann gerade die Lagrangeschen Multiplikatoren. Schritt (B) ist der schwierigste und benötigt eine Form des Satzes über die implizite Funktion. Darauf gehen wir im ersten Abschnitt ein.

8.1 Der Satz von Lyusternik

Wir erinnern an die Definition der Differenzierbarkeit von Funktionen $g : \mathbb{R}^n \to \mathbb{R}^p$ im Punkt $x \in \mathbb{R}^n$: Es gibt eine Matrix $g'(x) \in \mathbb{R}^{p\times n}$ mit

$$g(x+z) \;=\; g(x) \;+\; g'(x)z \;+\; \tilde{g}(x,z) \quad \text{für alle } z \in \mathbb{R}^n \quad \text{und } \lim_{z\to 0} \frac{\tilde{g}(x,z)}{\|z\|} \;=\; 0\,.$$

Hierfür benutzt man auch das Landausymbol und schreibt

$$g(x+z) \;=\; g(x) \;+\; g'(x)z \;+\; o(\|z\|)\,, \quad z \to 0\,.$$

Eine solche Darstellung gilt analog auch für f und h. Die Matrizen $f'(x) \in \mathbb{R}^{1\times n}$, $g'(x) \in \mathbb{R}^{p\times n}$, $h'(x) \in \mathbb{R}^{m\times n}$ heißen die **Funktionalmatrizen** und bestehen aus den partiellen Ableitungen, etwa

$$g'(x) \;=\; \left(\frac{\partial g_i(x)}{\partial x_j}\right)_{\substack{i=1,\dots,p\\ j=1,\dots,n}} \;\in\; \mathbb{R}^{p\times n}\,.$$

Es ist $f'(x) = \nabla f(x)^\top$, wobei $\nabla f(x) \in \mathbb{R}^n$ der Gradient (=Spaltenvektor!) von f bei x ist.

Ist $x^* \in M$ (optimal oder nicht), so können wir das zu (P) um x^* linearisierte Problem aufstellen. Man ersetze $f(x)$ durch $f(x^*) + f'(x^*)z$ usw. und erhält:

$$(LP) \qquad \text{Minimiere} \quad f'(x^*)z \quad \text{unter } z \in \mathbb{R}^n\,,\; g(x^*) + g'(x^*)z \le 0\,,\; h'(x^*)z = 0\,.$$

Um zu zeigen, dass 0 Lösung von (LP) ist, falls x^* Lösung von (P) ist, müssen wir zu jedem für (LP) zulässigen z in der Lage sein, ein dicht bei x^* gelegenes $x \in M$ zu konstruieren. Für die Gleichungsnebenbedingungen ist dies nicht so einfach. Wir benötigen eine Anwendung des folgenden Satzes über implizite Funktionen:

Satz 8.1 *(über implizite Funktionen)*
Sei $H : \mathbb{R}^m \times \mathbb{R} \longrightarrow \mathbb{R}^m$ *eine stetig differenzierbare Funktion,* $H(\hat{y}, \hat{t}) = 0$ *und*

$$\frac{\partial H(\hat{y}, \hat{t})}{\partial y} \;:=\; \left(\frac{\partial H_i(\hat{y}, \hat{t})}{\partial y_j} \right)_{i,j=1,\ldots,m} \;\in\; \mathbb{R}^{m \times m}$$

sei regulär. Dann existiert $\delta > 0$ *und eine stetig differenzierbare Funktion* $\gamma : \big(\hat{t} - \delta, \hat{t} + \delta\big) \to \mathbb{R}^m$ *mit* $\gamma(\hat{t}) = \hat{y}$ *und* $H\big(\gamma(t), t\big) = 0$ *für alle* $t \in \big(\hat{t} - \delta, \hat{t} + \delta\big)$.

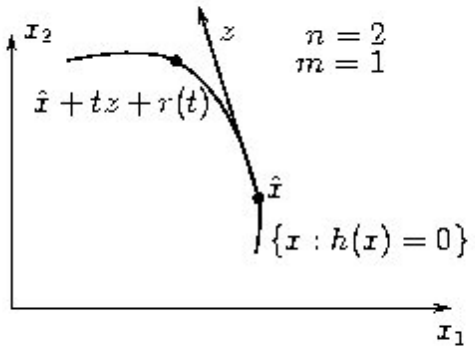
Die Ableitung γ' erhält man über die Kettenregel:

$$\frac{\partial H(\hat{y}, \hat{t})}{\partial y}\, \gamma'(\hat{t}) \;+\; \frac{\partial H(\hat{y}, \hat{t})}{\partial t} \;=\; 0\,, \quad \text{also} \quad \gamma'(\hat{t}) \;=\; -\left[\frac{\partial H(\hat{y}, \hat{t})}{\partial y} \right]^{-1} \frac{\partial H(\hat{y}, \hat{t})}{\partial t}\,.$$

Damit können wir zeigen:

Satz 8.2 *(Lyusternik)*
Sei $h : \mathbb{R}^n \to \mathbb{R}^m$ *stetig differenzierbar und* $h(\hat{x}) = 0$. *Die Funktionalmatrix* $h'(\hat{x}) \in \mathbb{R}^{m \times n}$ *habe den Rang* m *(also insbesondere* $m \le n$*). Sei ferner* $z \in \mathbb{R}^n$ *mit* $h'(\hat{x})z = 0$. *Dann existiert* $\delta > 0$ *und eine stetig differenzierbare Funktion* $r : (-\delta, \delta) \to \mathbb{R}^n$ *mit* $r(0) = 0$ *und* $r'(0) = \lim_{t \to 0} \frac{r(t)}{t} = 0$ *und* $h\big(\hat{x} + tz + r(t)\big) = 0$ *für alle* $t \in (-\delta, \delta)$.

Beweis: Es sei $V := \big(\operatorname{Kern} h'(\hat{x})\big)^{\perp} \subseteq \mathbb{R}^n$ das orthogonale Komplement des Kerns der Matrix $h'(\hat{x}) \in \mathbb{R}^{m \times n}$. Es ist $\dim V = m$, da $m = \operatorname{Rang} h'(\hat{x})$ (Dimensionssatz für Matrizen!). Sei $\big\{v^1, \ldots, v^m\big\} \subseteq V$ eine Basis von V und $B = [v^1 \cdots v^m] \in \mathbb{R}^{n \times m}$ die von ihnen gebildete Matrix. Definiere jetzt die Funktion $H : \mathbb{R}^m \times \mathbb{R} \longrightarrow \mathbb{R}^m$ durch



$$H(y, t) \;:=\; h(\hat{x} + tz + By)\,, \quad y \in \mathbb{R}^m\,,\; t \in \mathbb{R}\,.$$

Dann ist $H(0, 0) = h(\hat{x}) = 0$ und

$$\frac{\partial H(y, t)}{\partial y} \;=\; \underbrace{h'(\hat{x} + tz + By)}_{\in\, \mathbb{R}^{m \times n}} \underbrace{B}_{\in \mathbb{R}^{n \times m}} \;\in\; \mathbb{R}^{m \times m}\,,$$

also $\partial H(0,0)/\partial y = h'(\hat{x})\,B$. Diese quadratische Matrix ist regulär: $h'(\hat{x})\,By = 0$ impliziert $By \in \operatorname{Kern} h'(\hat{x})$. Es ist aber auch $By \in V = \big(\operatorname{Kern} h'(\hat{x})\big)^{\perp}$, also $By = 0$ und daher $y = 0$ (da $\operatorname{Rang} B = m$). Daher ist der Satz über implizite Funktionen anwendbar. Es existiert also $\delta > 0$ und eine stetig differenzierbare Funktion $\gamma : (-\delta, \delta) \to \mathbb{R}^m$ mit $\gamma(0) = 0$ und $h\big(\hat{x} + tz + B\gamma(t)\big) = 0$ für alle $t \in (-\delta, \delta)$. Setze $r(t) := B\gamma(t) \in \mathbb{R}^n$ für $t \in (-\delta, \delta)$. Dann ist r stetig differenzierbar, $r(0) = 0$ und

$$r'(0) \;=\; -B\left[h'(\hat{x})\,B\right]^{-1} \underbrace{h'(\hat{x})z}_{=0} \;=\; 0\,,$$

da $\partial H(y,t)/\partial t = h'(\hat{x} + tz + By)z$. □

8.2 Die Lagrangesche Multiplikatorenregel

Nach diesen Vorbereitungen können wir jetzt die Lagrangesche Multplikatorenregel beweisen:

Satz 8.3 *(Lagrangesche Multplikatorenregel)*
Sei $K \subseteq \mathbb{R}^n$ offen, $f : K \longrightarrow \mathbb{R}$, $g : K \longrightarrow \mathbb{R}^p$, $h : K \longrightarrow \mathbb{R}^m$ stetig differenzierbar und x^* ein (lokales) Minimum von f auf der Menge

$$M \;=\; \big\{x \in K : g(x) \le 0\,,\; h(x) = 0\big\}\,.$$

Es gelte die **constraint qualification**

(CQ1)
$$\begin{aligned} &(i) \quad \mathit{Rang}\, h'(x^*) \;=\; m \;\le\; n\,, \\ &(ii) \quad \text{es gibt } \hat{z} \in \mathbb{R}^n \text{ mit } g(x^*) \;+\; g'(x^*)\hat{z} \;<\; 0 \text{ und } h'(x^*)\hat{z} \;=\; 0\,. \end{aligned}$$

Dann gibt es $u^* \in \mathbb{R}^p$, $u^* \ge 0$, und $v^* \in \mathbb{R}^m$ mit

$$\nabla f(x^*) \;+\; \sum_{j=1}^{p} u_j^*\, \nabla g_j(x^*) \;+\; \sum_{j=1}^{m} v_j^*\, \nabla h_j(x^*) \;=\; 0\,.$$

Ferner ist $u_j^* g_j(x^*) = 0$ für alle $j = 1, \ldots, p$. Mit der Lagrangefunktion

$$L(x,u,v) \;=\; f(x) \;+\; u^{\top} g(x) \;+\; v^{\top} h(x)\,, \quad (x,u,v) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m\,,$$

können wir dies auch so ausdrücken:

$$\nabla_x L(x^*, u^*, v^*) \;=\; 0\,.$$

Beweis: Wir betrachten das linearisierte Problem:

(LP) Minimiere $f(x^*) + \nabla f(x^*)^{\top} z$ unter den Nebenbedingungen

$$g(x^*) \;+\; g'(x^*)z \;\le\; 0 \text{ und } h'(x^*)z \;=\; 0\,.$$

Wir zeigen, dass $z^* = 0$ eine Lösung dieses linearen Optimierungsproblems ist. Annahme, es gebe ein $z \in \mathbb{R}^n$ mit $g(x^*) + g'(x^*)z \le 0$ und $h'(x^*)z = 0$ und $\nabla f(x^*)^{\top} z < 0$. Setze $z_\varepsilon := (1-\varepsilon)z + \varepsilon\hat{z}$. Dann ist $h'(x^*)z_\varepsilon = 0$ und

$$g(x^*) + g'(x^*)z_\varepsilon \;=\; (1-\varepsilon)\big[g(x^*) + g'(x^*)z\big] \;+\; \varepsilon\big[g(x^*) + g'(x^*)\hat{z}\big] \;<\; 0$$

für alle $\varepsilon \in (0,1)$. Außerdem ist

$$\nabla f(x^*)^\top z_\varepsilon \;=\; (1-\varepsilon)\,\underbrace{\nabla f(x^*)^\top z}_{<0} \;+\; \varepsilon\,\nabla f(x^*)^\top \hat{z}\,.$$

Wähle $\varepsilon > 0$ so klein, dass $\nabla f(x^*)^\top z_\varepsilon < 0$, und halte im folgenden dieses ε fest.
Der Satz von Lyusternik liefert die Existenz einer Abbildung $r : (-\delta, \delta) \longrightarrow \mathbb{R}^n$ mit

$$h\big(x^* + t z_\varepsilon + r(t)\big) \;=\; 0 \text{ für alle } |t| < \delta \text{ und } r(0) = 0 \text{ und } r'(0) = \lim_{t\to 0} \frac{r(t)}{t} = 0\,.$$

Wähle δ so klein, dass $x_t := x^* + t z_\varepsilon + r(t) \in K$ für $|t| < \delta$.
Um $x_t \in M$ zu zeigen, müssen wir nur noch $g(x_t) \le 0$ beweisen. Wegen der Differenzierbarkeit von g existiert $\tilde{g}$ mit

$$g(x_t) \;=\; g(x^*) \;+\; g'(x^*)\big[t\,z_\varepsilon + r(t)\big] \;+\; \tilde{g}(t) \quad \text{und} \quad \lim_{t\to 0} \frac{\tilde{g}(t)}{\|t z_\varepsilon + r(t)\|} \;=\; 0\,.$$

Wir schreiben daher

$$g(x_t) \;=\; t\left[\underbrace{\big(g(x^*) + g'(x^*) z_\varepsilon\big)}_{<0} \;+\; \left(g'(x^*)\frac{r(t)}{t} + \frac{\tilde{g}(t)}{t}\right)\right] \;+\; (1-t)\,\underbrace{g(x^*)}_{\le 0}\,.$$

Für hinreichend kleine $t > 0$ ist also $x_t \in M$. Ganz analog ist

$$f(x_t) \;=\; t\left[\underbrace{\nabla f(x^*)^\top z_\varepsilon}_{<0} \;+\; \left(\nabla f(x^*)^\top \frac{r(t)}{t} + \frac{\tilde{f}(t)}{t}\right)\right] \;+\; f(x^*) \;<\; f(x^*)$$

für hinreichend kleine $t > 0$. Dies ist ein Widerspruch zur Optimalität von x^*.

Damit haben wir gezeigt, dass $z^* = 0$ eine Lösung des linearen Optimierungsproblems (LP) ist. Wir schreiben dies in der Standardform als:

$$\text{Minimiere} \quad \nabla f(x^*)^\top z \quad \text{unter} \quad \begin{bmatrix} g'(x^*) \\ h'(x^*) \\ -h'(x^*) \end{bmatrix} z \;\le\; \begin{bmatrix} -g(x^*) \\ 0 \\ 0 \end{bmatrix}.$$

Der Satz von Kuhn-Tucker für diese Standardform (z.B. Satz 7.5 für $Q = 0$) liefert die Existenz von $u^* \in \mathbb{R}^p$, $y^+, y^- \in \mathbb{R}^m$ mit $u^* \ge 0$, $y^+ \ge 0$, $y^- \ge 0$ und

$$\nabla f'(x^*) \;+\; g'(x^*)^\top u^* \;+\; h'(x^*)^\top y^+ \;-\; h'(x^*)^\top y^- \;=\; 0\,,$$

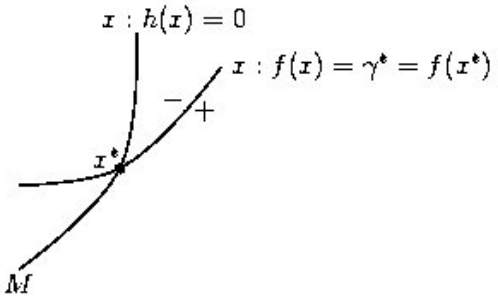
also mit $v^* = y^+ - y^-$:

$$\nabla f'(x^*) \;+\; g'(x^*)^\top u \;+\; h'(x^*)^\top v \;=\; 0\,.$$

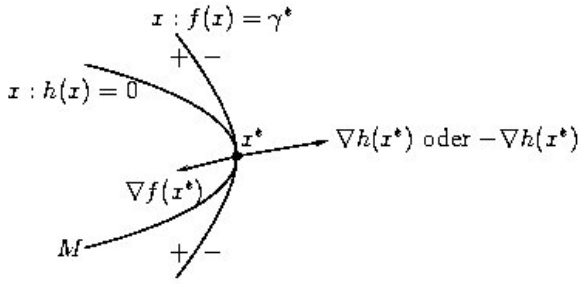
Dies ist gerade die Behauptung. □

Wir wollen jetzt die Lagrangesche Multiplikatonregel geometrisch veranschaulichen, zunächst für eine Gleichungs-, dann für eine Ungleichungsrestriktion. Sei also zunächst $n = 2$,

$m = 1$, und Ungleichungsrestriktionen treten nicht auf. Wir haben also das Problem, f unter $h(x) = 0$ zu minimieren, wobei $f, h : \mathbb{R}^2 \to \mathbb{R}$. Sei x^* eine (lokale) Lösung. Wir zeichnen die Niveaulinien $f(x) = c^* := f(x^*)$ und $h(x) = 0$ auf:

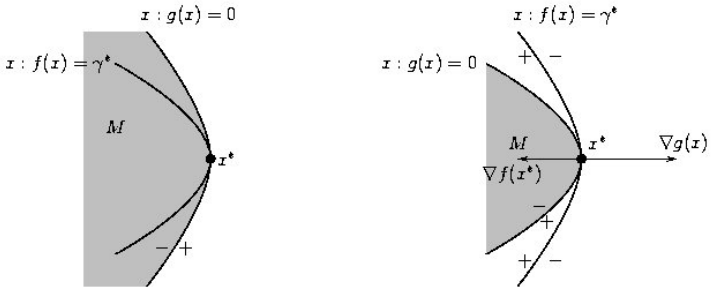


Das Minus- bzw. Pluszeichen gibt an, wo $f(x) < c^*$ bzw. $f(x) > c^*$ ist. So wie in diesem Bild kann es nicht aussehen, da man sicher kleinere zulässige Funktionswerte bekommt, etwa für $\hat{x}$. Offenbar müssen sich im Optimum die Höhenlinien tangential berühren, also so:



Der Gradient $\nabla f(x^*)$ steht bekanntlich senkrecht auf der Höhenlinie und zeigt in Richtung der größten Steigung, also in die „+“–Menge. Wir erkennen, dass $\nabla f(x^*)$ und $\nabla h(x^*)$ kollinear sein müssen. Dies bedeutet, dass es $\lambda \in \mathbb{R}$ geben muss mit $\nabla f(x^*) = \lambda \nabla h(x^*)$. Dies ist die Aussage der Multiplikatorenregel.

Jetzt betrachten wir den Fall einer Ungleichung, also das Problem, f unter $g(x) \leq 0$ zu minimieren, wobei jetzt $f, g : \mathbb{R}^2 \to \mathbb{R}$. Wiederum zeichen wir die Niveaulinien $f(x) = c^* := f(x^*)$ und $g(x) = 0$ auf und markieren die „+“– und „−“–Mengen.



Wir erkennen, dass $\nabla f(x^*)$ und $\nabla g(x^*)$ wieder kollinear sind, aber in entgegengesetzte Richtungen zeigen. Das heißt, es gibt $\lambda \geq 0$ mit $\nabla f(x^*) = -\lambda \nabla g(x^*)$, und dies ist wieder die Aussage der Multiplikatorenregel.

Beispiele 8.4 (a) Es seien reelle Zahlen $t_1, \ldots, t_n > 0$ mit $\sum_{j=1}^n t_j = 1$ gegeben. Dann gilt die **Ungleichung vom geometrisch-arithmetrischen Mittel**, d.h.

$$\prod_{j=1}^n x_j^{t_j} \;\leq\; \sum_{j=1}^n t_j\, x_j \quad \text{für alle } x_j \geq 0\,.$$

Beweis: Wir betrachten die Maximierungsaufgabe:

$$\text{Maximiere} \quad \prod_{j=1}^n x_j^{t_j} \quad \text{unter } x_j \geq 0 \text{ und } \sum_{j=1}^n t_j\, x_j \;=\; 1\,.$$

Es gibt eine Lösung x^*, da die zulässige Menge kompakt und die Zielfunktion stetig ist. Es ist sicher $x_j^* > 0$ für alle $j = 1, \ldots, n$, da der Zielfunktionswert sonst 0 und damit nicht maximal wäre. Damit verschwinden alle Lagrangemultiplikatoren zu den Ungleichungen $x \geq 0$. Die constraint qualification ist erfüllt, denn mit

$$h(x) \;=\; \sum_{j=1}^n t_j\, x_j \;-\; 1 \quad \text{ist} \quad \nabla h(x^*) = (t_1, \ldots, t_n)^\top \neq 0\,.$$

Die Zielfunktion logarithmieren wir, betrachten also das Problem,

$$f(x) \;:=\; -\sum_{j=1}^n t_j \ln x_j$$

zu minimieren unter $x > 0$ und $h(x) = 0$. Anwendung der Multiplikatorenregel liefert die Existenz von $v \in \mathbb{R}$ mit $\nabla f(x^*) + v \nabla h(x^*) = 0$, also

$$-\frac{t_j}{x_j^*} \;+\; v\, t_j \;=\; 0 \quad \text{für alle } j = 1, \ldots, n\,.$$

Hieraus folgt $x_j^* = 1/v$ für alle j. Wegen $\sum_{j=1}^n t_j\, x_j^* = 1$ folgt $v = 1$, also $x_j^* = 1$ für alle j. Daher ist

$$\prod_{j=1}^n x_j^{t_j} \;\leq\; 1 \quad \text{für alle } x \geq 0 \text{ mit} \quad \sum_{j=1}^n t_j\, x_j \;=\; 1\,,$$

und Gleichheit tritt genau für $x = (1, \ldots, 1)^\top$ ein.
Ist nun $x \geq 0$ beliebig, so setzen wir $\hat{x}_j = x_j \left[\sum_{k=1}^n t_k\, x_k\right]^{-1}$. Dann ist $\hat{x}$ zulässig für das Optimierungsproblem, also $\prod_{j=1}^n \hat{x}_j^{t_j} \leq 1$, d.h. $\prod_{j=1}^n x_j^{t_j} \leq \sum_{j=1}^n t_j\, x_j$. □

(b) Sei $A \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix. Betrachte das Problem:

$$\text{Maximiere} \quad x^\top A\, x \quad \text{unter} \quad g(x) \;:=\; x^\top x \;-\; 1 \;\leq\; 0\,.$$

Es existiert eine Lösung x^*, da die Einheitskugel kompakt und die Zielfunktion stetig ist. Es ist mit $f(x) = -x^\top A x$:

$$\nabla f(x) \;=\; -2Ax \quad \text{und} \quad \nabla g(x) \;=\; 2x\,.$$

Die constraint qualification lautet: Es gibt $\hat{x} \in \mathbb{R}^n$ mit $\|x^*\|^2 + 2\,\hat{x}^\top x^* < 1$. Dies ist z.B. für $\hat{x} = -\frac{1}{2}\,x^*$ erfüllt.
Also existiert $\lambda^* \geq 0$ mit $\nabla f(x^*) + \lambda^* \nabla g(x^*) = 0$, d.h. $-Ax^* + \lambda^* x^* = 0$. Daher ist entweder $x^* = 0$ oder λ^* ein Eigenwert von A mit zugehörigem Eigenvektor x^*. Damit haben wir gezeigt:

$$\max_{\|x\| \leq 1} x^\top A x \;=\; \max\bigl\{\lambda : \lambda \text{ ist Eigenwert von } A\bigr\} \;\cup\; \{0\}\,.$$

(c) Man bestimme den kürzesten Abstand zwischen der Ellipse $x^2 + y^2/4 = 1$ und der Kurve $y = 3\cos x$. Der Abstand zwischen den Punkten (x, y) und (w, z) ist $\sqrt{(x-w)^2 + (y-z)^2}$. Also haben wir $f(x, y, w, z) := (x-w)^2 + (y-z)^2$ zu minimieren unter den Nebenbedingungen

$$\begin{aligned} h_1(x, y, w, z) \;&:=\; x^2 \;+\; \frac{y^2}{4} \;-\; 1 \;=\; 0 \quad \text{und} \\ h_2(x, y, w, z) \;&:=\; z \;-\; 3\cos w \;=\; 0\,. \end{aligned}$$

Die Menge $M := \{(x, y, w, z) \in \mathbb{R}^4 : h_j = 0,\ j = 1, 2\}$ ist nicht beschränkt, also nicht kompakt. Trotzdem existiert eine Lösung (nicht nur aus anschaulichen, sondern auch aus mathematischen Gründen). Man braucht sich für die Suche nach einem Minimum ja nur auf Teilmengen der Form $\hat{M} := \bigl\{(x, y, w, z) \in M : f(x, y, w, z) \leq f(\hat{x}, \hat{y}, \hat{w}, \hat{z})\bigr\}$ für irgendein $(\hat{x}, \hat{y}, \hat{w}, \hat{z}) \in M$ zu beschränken. Jedes Minimum von f auf $\hat{M}$ ist natürlich auch Minimum auf M, denn jedes $(x, y, w, z) \in M \setminus \hat{M}$ hat sicher einen größeren Funktionswert. Setzt man in unserem Fall etwa $(\hat{x}, \hat{y}, \hat{w}, \hat{z}) = (1, 0, \pi/2, 0) \in M$, so ist $\hat{M}$ beschränkt, da aus $(x-w)^2 + (y-z)^2 = f(x, y, w, z) \leq f(\hat{x}, \hat{y}, \hat{w}, \hat{z}) = f(1, 0, \pi/2, 0) = (1 - \pi/2)^2$ und der Beschränktheit von M bzgl. x und y sicher auch die Beschränktheit bzgl. w und z folgt. Also existiert eine Lösung, die wir mit (x, y, w, z) bezeichnen.
Wir rechnen aus:

$$\nabla f(x, y, w, z) \;=\; 2 \begin{pmatrix} x - w \\ y - z \\ w - x \\ z - y \end{pmatrix},$$

und

$$\nabla h_1(x, y, w, z) \;=\; 2 \begin{pmatrix} x \\ y/4 \\ 0 \\ 0 \end{pmatrix}, \quad \nabla h_2(x, y, w, z) \;=\; \begin{pmatrix} 0 \\ 0 \\ -3\sin w \\ 1 \end{pmatrix}.$$

Die constraint qualification ist im optimalen Punkt erfüllt, denn $\nabla h_1(x, y, w, z)$ und $\nabla h_2(x, y, w, z)$ sind für $(x, y, w, z) \neq 0$ linear unabhängig. Also existieren Multiplikatoren $\lambda, \mu \in \mathbb{R}$ mit

$$\nabla f(x, y, w, z) \;+\; \lambda\,\nabla h_1(x, y, w, z) \;+\; \mu\,\nabla h_2(x, y, w, z) \;=\; 0\,,$$

d.h.

$$\begin{aligned} x - w &+ \lambda x &= 0 \\ y - z &+ \lambda y/4 &= 0 \\ -2(x-w) &- 3\mu \sin w &= 0 \\ -2(y-z) &+ \mu &= 0 \end{aligned}$$

Zu diesen vier Gleichungen für die 6 Unbekannten x, y, w, z, λ, μ kommen noch die zwei Gleichungen $x^2 + y^2/4 = 1$ und $z = 3\cos w$. Also erhalten wir ein nichtlineares System von 6 Gleichungen und 6 Variablen. Man muss dieses System numerisch lösen, etwa mit dem Newtonverfahren. Startwerte für x, y, w, z sind leicht zu erhalten, indem man sich die Skizze ansieht. Wir starten mit $(x^0, y^0, w^0, z^0) = (1, 0, \pi/2, 0)$. Schwieriger ist es, Startwerte für die Multiplikatoren zu finden. Man kann aber die erste und die vierte Gleichung nach λ bzw. μ auflösen und dann $\lambda^0 = w^0/x^0 - 1 = \pi/2 - 1$ sowie $\mu^0 = 2(y^0 - z^0) = 0$ nehmen. Mit diesen Startwerten erhalten wir:

k	x^k	y^k	w^k	z^k	λ^k	μ^k	$\|\|F(x^k)\|\|$
1	1.000	0.000	1.571	0.000	0.571	0.000	$5.7079633e-01$
2	1.000	0.706	1.302	0.806	0.302	−0.201	$1.3396653e-01$
3	0.839	1.263	1.107	1.361	0.316	−0.196	$1.0550283e-01$
4	0.805	1.190	1.120	1.306	0.388	−0.233	$3.7674328e-03$
5	0.804	1.190	1.120	1.308	0.394	−0.234	$1.1061641e-05$
6	0.804	1.190	1.120	1.308	0.394	−0.234	$5.7518120e-10$

(d) (Pferdewette)
Wir benutzen die folgenden Bezeichnungen bei einer Pferdewette mit n Pferden:

- x_0: Gesamteinsatz (fest gewählt)
- x_i: Einsatz auf Pferd Nummer i
- p_i: Wahrscheinlichkeit dafür, dass Pferd i gewinnt
- s_i: Wettbetrag aller anderen Wetter auf Pferd i
- $c \in (0,1)$: ausgezahlter Teil des Gesamteinsatzes

Der Gesamtwettbetrag ist dann $x_0 + \sum_{i=1}^n s_i$. Damit ist der ausgeschüttete Gewinn: $c\big(x_0 + \sum_{i=1}^n s_i\big)$. Der mir ausgezahlte Betrag, wenn Pferd i gewinnt, ist

$$c\left(x_0 + \sum_{j=1}^n s_j\right) \frac{x_i}{s_i + x_i}\,.$$

Der zu erwartende Reingewinn ist

$$R(x) \;:=\; c\left(x_0 + \sum_{j=1}^n s_j\right) \sum_{i=1}^n \frac{p_i\, x_i}{s_i + x_i} \;-\; x_0\,.$$

Das Optimierungsproblem besteht also darin, $R(x)$ zu maximieren unter den Nebenbedingungen $x \geq 0$ und $h(x) := \sum_{i=1}^n x_i - x_0 = 0$. Statt $R(x)$ zu maximieren kann man genauso gut

$$f(x) \;:=\; -\sum_{i=1}^n \frac{p_i\, x_i}{s_i + x_i}$$

minimieren. Es ist $g(x) := -x$. Die Existenz einer Lösung x^* ist klar, da die Menge M der zulässigen Punkte kompakt ist. Auch die constraint qualification ist erfüllt. Wir müssen zeigen, dass es $\hat{x} \in \mathbb{R}^n$ gibt mit $x^* + \hat{x} > 0$ und $e^\top \hat{x} = 0$. (Hier ist $e = (1, \ldots, 1)^\top \in \mathbb{R}^n$.) Dies ist gleichbedeutend mit der Forderung, dass es $\tilde{x} > 0$ gibt mit $e^\top \tilde{x} = x_0$. Setze $\tilde{x} = \frac{x_0}{n}\, e$, also $\hat{x} = x_0/n\, e - x^*$. Schließlich ist

$$\frac{\partial f(x)}{\partial x_i} \;=\; -\frac{p_i}{s_i + x_i} \;+\; \frac{p_i\, x_i}{(s_i + x_i)^2} \;=\; -\frac{s_i\, p_i}{(s_i + x_i)^2}\,.$$

Also ist die Multiplikatorenregel anwendbar und liefert die Existenz von $u \in \mathbb{R}^n$, $u \geq 0$, und $\lambda \in \mathbb{R}$ mit

$$(*) \qquad -\frac{s_i\, p_i}{(s_i + x_i^*)^2} \;-\; u_i \;+\; \lambda \;=\; 0 \quad \text{und} \quad u_i x_i^* = 0\,.$$

Wie können wir diese Gleichungen zur Berechnung von x^* ausnutzen? Das folgende Vorgehen ist typisch für viele Anwendungen. Wir betrachten zunächst $\lambda \in \mathbb{R}$ als Parameter. Für jedes $i = 1, \ldots, n$ machen wir eine Fallunterscheidung:
1. Fall: $p_i/s_i > \lambda$. Wäre $x_i^* = 0$, so würde aus $(*)$ folgen, dass

$$u_i \;=\; \lambda \;-\; \frac{p_i}{s_i} \;<\; 0\,,$$

ein Widerspruch. Also ist $x_i^* > 0$, also $u_i = 0$, also

$$(s_i + x_i^*)^2 \;=\; \frac{s_i\, p_i}{\lambda}\,, \quad \text{also } \lambda > 0 \quad \text{also}$$

$$x_i^* \;=\; \sqrt{\frac{s_i\, p_i}{\lambda}} \;-\; s_i\,.$$

2. Fall: $p_i/s_i \leq \lambda$. Wäre $x_i^* > 0$, so $u_i = 0$, also $(s_i + x_i^*)^2 = s_i\, p_i/\lambda$, also $\lambda > 0$ und

$$x_i^* \;=\; \sqrt{\frac{s_i\, p_i}{\lambda}} \;-\; s_i \;=\; \frac{s_i}{\sqrt{\lambda}} \left[\sqrt{\frac{p_i}{s_i}} - \sqrt{\lambda}\right] \;\leq\; 0\,,$$

ein Widerspruch. Also ist $x_i^* = 0$.

Damit haben wir $x_i^* = \rho_i(\lambda)$, wobei

$$\rho_i(\lambda) \;:=\; \begin{cases} \sqrt{s_i\, p_i/\lambda} \;-\; s_i\,, & \text{falls } p_i/s_i > \lambda\,, \\ 0\,, & \text{falls } p_i/s_i \leq \lambda\,. \end{cases}$$

Der Parameter λ ist schließlich aus der Gleichung $\sum_{i=1}^n \rho_i(\lambda) = x_0$ zu bestimmen. Dies kann etwa mit der Regula Falsi geschehen.

8.3 Notwendige und hinreichende Bedingungen zweiter Ordnung

Wir betrachten das gleiche Problem wie im letzten Abschnitt. Wir benötigen für das folgende eine stärkere constraint qualification. Sei dazu wieder $x^* \in M$ ein lokales Minimum von f auf M. Definiere wieder die Menge der aktiven Indizes durch $I(x^*) = \big\{i \in \{1,\dots,p\} : g_i(x^*) = 0\big\}$.

(CQ2) Die Vektoren

$$\big\{\nabla g_i(x^*) : i \in I(x^*)\big\} \;\cup\; \big\{\nabla h_j(x^*) : j = 1,\dots,m\big\}$$

seien linear unabhängig.

Ist also $I = I(x^*) = \{i_1,\dots,i_q\}$ mit $1 \le i_1 < \dots < i_q \le p$ und $g_I(x) := \big(g_{i_j}(x)\big)_{j=1,\dots,q} \in \mathbb{R}^q$, so bedeutet (CQ2) genau, dass

$$\operatorname{Rang} \begin{bmatrix} g_I{}'(x^*) \\ h'(x^*) \end{bmatrix} \;=\; q+m\,.$$

Insbesondere kann dies nur für $q+m \le n$ der Fall sein. (CQ2) ist stärker als (CQ1):

Lemma 8.5 *Sei $x^* \in M$ optimal. Ist (CQ2) erfüllt, so auch (CQ1).*

Beweis: Bedingung (i) ist natürlich erfüllt. Für (ii) müssen wir zeigen, dass es $\hat z \in \mathbb{R}^n$ gibt mit

$$g_I{}'(x^*)\hat z < 0 \quad \text{und} \quad h'(x^*)\hat z = 0 \tag{8.1}$$

sowie $$g_i(x^*) + \nabla g_i(x^*)^\top \hat z < 0 \quad \text{für alle } i \notin I\,.$$

Da die Matrix $\begin{bmatrix} g_I{}'(x^*) \\ h'(x^*) \end{bmatrix} \in \mathbb{R}^{(q+m)\times n}$ den Rang $q+m$ hat, so existiert $\tilde z \in \mathbb{R}^n$ mit $\nabla g_i(x^*)^\top \tilde z = -1$ für alle $i \in I$ und $h'(x^*)\tilde z = 0$. Setze $\hat z = t\tilde z$ für noch zu bestimmendes $t > 0$. Dann ist $\nabla g_i(x^*)^\top \hat z = -t < 0$ für alle $i \in I$ und $h'(x^*)\hat z = 0$. Für $i \notin I$ ist $g_i(x^*) + \nabla g_i(x^*)^\top \hat z = g_i(x^*) + t\,\nabla g_i(x^*)^\top \tilde z$. Wegen $g_i(x^*) < 0$ für alle $i \notin I$ gilt für hinreichend kleines $t > 0$ auch $g_i(x^*) + \nabla g_i(x^*)^\top \hat z < 0$. □

Satz 8.6 *(notwendige Optimierungsbedingung 2. Ordnung)*
Sei x^* lokales Minimum von f auf M, und sei (CQ2) erfüllt. Es seien zusätzlich f, g und h zweimal stetig differenzierbar in x^*. Dann existiert $u^* \in \mathbb{R}^p$, $v^* \in \mathbb{R}^m$, $u^* \ge 0$, mit

(a) $\nabla_x L(x^*, u^*, v^*) = 0$, *wobei wieder*

$$L(x,u,v) \;=\; f(x) \;+\; u^\top g(x) \;+\; v^\top h(x)\,, \quad x \in \mathbb{R}^n\,,\; u \in \mathbb{R}^p,\; v \in \mathbb{R}^m\,,$$

die Langrangefunktion ist.

(b) $u_i^*\, g_i(x') = 0$ *für alle* $i = 1,\dots,p$.

(c) Definiere den Unterraum

$$V \;=\; \big\{z \in \mathbb{R}^n : h'(x^*)z = 0\,,\ \nabla g_i(x^*)^\top z = 0 \text{ für alle } i \in I(x^*)\big\}\,.$$

Dann gilt

$$z^\top \nabla_x^2 L(x^*, u^*, v^*)\, z \;\geq\; 0 \quad \text{für alle } z \in V\,,$$

d.h. $\nabla_x^2 L(x^*, u^*, v^*)$ ist auf V positiv semidefinit.

Beweis: Es ist nur (c) zu zeigen.
Sei $z \in V$ festgehalten. Wende jetzt den Satz von Lyusternik auf die Funktion $\tilde{h} : \mathbb{R}^n \to \mathbb{R}^{q+m}$, definiert durch

$$\tilde{h}(x) \;=\; \begin{bmatrix} g_I(x) \\ h(x) \end{bmatrix}\,, \quad x \in \mathbb{R}^n\,,$$

an, wobei wieder $I = I(x^*) = \{i_1, \ldots, i_q\}$. Dies ist möglich wegen $\tilde{h}'(x^*)z = 0$ und $(CQ2)$ und liefert $\delta > 0$ und $r : (-\delta, \delta) \to \mathbb{R}^n$ mit $r(0) = r'(0)$ und

$$\begin{aligned} g_i\big(x^* + tz + r(t)\big) &= 0 \quad \text{für } i \in I(x') \text{ und} \\ h\big(x^* + tz + r(t)\big) &= 0 \end{aligned}$$

für alle $t \in (-\delta, \delta)$. Für $i \notin I$ ist $g_i(x^*) < 0$. Daher können wir $\delta > 0$ so klein wählen, dass außerdem $g_i\big(x^* + tx + r(t)\big) \leq 0$ gilt für $i \notin I(x^*)$ und $t \in (-\delta, \delta)$. Setze zur Abkürzung

$$\gamma(t) \;=\; x^* + tz + r(t)\,, \quad |t| < \delta.$$

Dann ist $\gamma(t) \in M$ und $(f \circ \gamma) : (-\delta, \delta) \to \mathbb{R}$ hat in $t = 0$ ein lokales Minimum. Die zweimalige Differenzierbarkeit von g und h liefert, dass auch γ zweimal stetig differenzierbar ist. Für die Formeln $\frac{d}{dt}\big(f \circ \gamma\big)(0) = 0$ und $\frac{d^2}{dt^2}\big(f \circ \gamma\big)(0) \geq 0$ benötigen wir die erste und zweite Ableitung (Kettenregel!):

$$\begin{aligned} \frac{d}{dt}\big(f \circ \gamma\big)(t) &= \nabla f\big(\gamma(t)\big)^\top \big(z + r'(t)\big)\,, \\ \frac{d^2}{dt^2}\big(f \circ \gamma\big)(t) &= \big(z + r'(t)\big)^\top \nabla^2 f\big(\gamma(t)\big)\big(z + r'(t)\big) \;+\; \nabla f\big(\gamma(t)\big)^\top r''(t)\,. \end{aligned}$$

Für $t = 0$ erhalten wir

$$\nabla f(x^*)^\top z \;=\; 0 \quad \text{und} \tag{8.2a}$$

$$z^\top \nabla^2 f(x^*) z \;+\; \nabla f(x^*)^\top r''(0) \;\geq\; 0\,. \tag{8.2b}$$

Jetzt definieren wir

$$\psi(t) \;=\; \sum_{i \in I} u_i^*\, g_i\big(\gamma(t)\big) \;+\; \sum_{i=1}^{m} v_i^*\, h_i\big(\gamma(t)\big)\,, \quad |t| < \delta\,.$$

Dann ist nach Konstruktion von $r(t)$ gerade $\psi(t) = 0$ für alle t. Daher können wir $\psi(t) = 0$ zweimal differenzieren und erhalten

$$\begin{aligned}
\psi'(t) &= \left[\sum_{i\in I} u_i^* \nabla g_i\big(\gamma(t)\big) + \sum_{i=1}^m v_i^* \nabla h_i\big(\gamma(t)\big)\right]^\top \big(z + r'(t)\big)\,, \\
\psi''(t) &= \big(z + r'(t)\big)^\top \left[\sum_{i\in I} u_i^* \nabla^2 g_i\big(\gamma(t)\big) + \sum_{i=1}^m v_i^* \nabla^2 h_i\big(\gamma(t)\big)\right] \big(z + r'(t)\big) + \\
&\quad + \left[\sum_{i\in I} u_i^* \nabla g_i\big(\gamma(t)\big) + \sum_{i=1}^m v_i^* \nabla h_i\big(\gamma(t)\big)\right]^\top r''(t)\,.
\end{aligned}$$

Für $t = 0$ erhalten wir

$$\begin{aligned}
&z^\top \left[\sum_{i\in I} u_i^* \nabla^2 g_i(x^*) + \sum_{i=1}^m v_i^* \nabla^2 h_i(x^*)\right] z + \\
&\quad + \left[\sum_{i\in I} u_i^* \nabla g_i(x^*) + \sum_{i=1}^m v_i^* \nabla h_i(x^*)\right]^\top r''(0) = 0\,. \qquad (8.3)
\end{aligned}$$

In der letzten Gleichung dürfen wir I durch $\{1,\ldots,p\}$ ersetzen, da $u_i^* = 0$ für $i \notin I$. Summation von (8.2b) und (8.3) liefert

$$\begin{aligned}
&z^\top \left[\nabla^2 f(x^*) + \sum_{i=1}^p u_i^* \nabla^2 g_i(x^*) + \sum_{i=1}^m v_i^* \nabla^2 h_i(x^*)\right] z + \\
&\quad + \left[\nabla f(x^*) + \sum_{i=1}^p u_i^* \nabla g_i(x^*) + \sum_{i=1}^m v_i^* \nabla h_i(x^*)\right]^\top r''(0) \geq 0\,.
\end{aligned}$$

Die zweite Klammer verschwindet, und damit ist die Behauptung gezeigt. □

Natürlich ist dieser Satz einer Verallgemeinerung der wohlbekannten Aussage, dass $f'(x^*) = 0$ und $f''(x^*) \geq 0$ sein muss, falls $x^* \in \mathbb{R}$ ein (lokales) Minimum von f auf $\mathbb{R}$ ist. Die Bedingungen $f'(x^*) = 0$ und $f''(x^*) > 0$ sind hinreichend dafür, dass x^* ein lokales Minimum ist. Das wollen wir nun auch übertragen. Dazu definieren wir die Menge der „stark aktiven" Indizes durch

$$I^+ = I^+(x^*, u^*) = \big\{i \in I(x^*) : u_i^* > 0\big\}\,.$$

Dann haben wir:

Satz 8.7 *(hinreichende Optimierungsbedingung 2. Ordnung)*
Sei $(x^*, u^*, v^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ ein Kuhn-Tucker Punkt, d.h. $x^* \in M$ und $u^* \geq 0$ und $u_i^*\, g_i(x^*) = 0$ für alle $i = 1,\ldots,p$ sowie $\nabla_x L(x^*, u^*, v^*) = 0$. Wieder seien f, g und h zweimal stetig differenzierbar in x^*. Außerdem sei $\nabla_x^2 L(x^*, u^*, v^*)$ auf dem Kegel

$$K = \big\{z \in \mathbb{R}^n : h'(x^*)z = 0\,,\ \nabla g_i(x^*)^\top z = 0 \text{ für } i \in I^+\,,\ \nabla g_i(x^*)^\top z \leq 0 \text{ für } i \in I(x^*)\backslash I^+\big\}$$

strikt positiv definit, d.h.

$$z^\top \nabla^2 L(x^*, u^*, v^*) z > 0 \quad \text{für alle } z \in K\,,\ z \neq 0\,.$$

Dann ist x^* striktes lokales Minimum von f auf M.

Beweis: Angenommen, dies sei nicht der Fall. Dann existiert eine Folge $x_k \in M$ mit $f(x_k) \leq f(x^*)$ und $x_k \to x^*$, $x_k \neq x^*$. Wir schreiben x_k in der Form $x_k = x^* + t_k z_k$ mit $\|z_k\| = 1$ und $t_k \to 0$. Es gibt eine konvergente Teilfolge von z_k, wir schreiben einfach $z_k \to z$, $k \to \infty$, für ein $\|z\| = 1$. Nun benutzen wir die zweimalige Differenzierbarkeit, z.B. von g_i, $i \in I(x^*)$:

$$0 \;\geq\; g_i(x_k) - \underbrace{g_i(x^*)}_{=0} \;=\; t_k \nabla g_i(x^*)^\top z_k \;+\; \frac{1}{2}\, t_k^2\, z_k^\top \nabla^2 g_i(x^*) z_k \;+\; \tilde{g}_i(t_k z_k) \tag{8.4}$$

mit $\tilde{g}(t_k z_k)/t_k^2 \to 0$, $k \to \infty$. Division der Ungleichung durch t_k und $k \to \infty$ liefert $\nabla g_i(x^*)^\top z \leq 0$ für alle $i \in I$. Genauso ist $h'(x^*)z = 0$ und $\nabla f(x^*)^\top z \leq 0$. Wir wollen noch $\nabla g_i(x^*)^\top z = 0$ für alle $i \in I^+$ zeigen. Dies folgt aber genau aus der Gleichunng

$$0 \;=\; \nabla_x L(x^*, u^*, v^*)^\top z \;=\; \nabla f(x^*)^\top z \;+\; \sum_{i \in I^+} u_i^*\, \nabla g_i(x^*)^\top z \;+\; v^{*\top} h'(x^*) z \;\leq\; 0\,,$$

da alle Summanden ≤ 0 sind.
Also ist $z \in K$ und daher $z^\top \nabla^2 L(x^*, u^*, v^*) z > 0$. Nun nutzen wir die zweite Ableitung aus, also (8.4) und analog

$$\begin{aligned}
0 \;&=\; h_i(x_k) - h_i(x^*) \;=\; t_k \nabla h_i(x^*)^\top z_k \;+\; \frac{1}{2}\, t_k^2\, z_k^\top \nabla^2 h_i(x^*) z_k \;+\; \tilde{h}_i(t_k z_k)\,,\; i = 1, \ldots, m\,,\\
0 \;&\geq\; f(x_k) - f(x^*) \;=\; t_k \nabla f(x^*)^\top z_k \;+\; \frac{1}{2}\, t_k^2\, z_k^\top \nabla^2 f(x^*) z_k \;+\; \tilde{f}(t_k z_k)\,.
\end{aligned}$$

Nun dividieren wir diese Ungleichungen durch t_k und multiplizieren sie mit den Lagrangemultiplikatoren und summieren auf:

$$\begin{aligned}
0 \;\geq\; &\left[\nabla f(x^*) \;+\; \sum_{i \in I} u_i^*\, \nabla g_i(x^*) \;+\; \sum_{i=1}^m v_i^*\, \nabla h_i(x^*)\right]^\top z_k \;+\\
&+\; \frac{1}{2}\, t_k\, z_k^\top \left[\nabla^2 f(x^*) \;+\; \sum_{i \in I} u_i^*\, \nabla^2 g_i(x^*) \;+\; \sum_{i=1}^m v_i^*\, \nabla^2 h_i(x^*)\right] z_k \;+\; o(t_k)\,.
\end{aligned}$$

Die erste Klammer $[\cdots]$ verschwindet. Weitere Division durch t_k und $k \to 0$ liefert $z^\top \nabla^2 L(x^*, u^*, v^*) z \leq 0$, ein Widerspruch. □