Lineare Optimierung

Bernhard Schmitt

Winter-Semester 2013/14

Inhaltsverzeichnis

1 Optimierungs-Probleme

1.1 Strukturen

Eine präzise Vorstellung für die "Optimierung" einer Eigenschaft erfordert, dass man deren Qualität F quantitativ (als reelle Zahl) angeben kann und dass man sich über Einflußgrößen x dieser Qualität im Klaren ist. Wenn man dann die in Frage kommenden Werte der Parameter x zu einer Menge X zusammenfaßt ist das Qualitätsmaß $F : X \to \mathbb{R}$ eine reelle Funktion auf X. In der *Optimierungsaufgabe*

$$\min\{F(x) : x \in X\} \quad \text{bzw.} \quad \left\{ \begin{array}{c} \min F(x) \\ x \in X \end{array} \right. \tag{P}$$

wird eine Minimalstelle $\hat{x} \in X$ gesucht mit $F(\hat{x}) \leq F(x)\, \forall x \in X$.

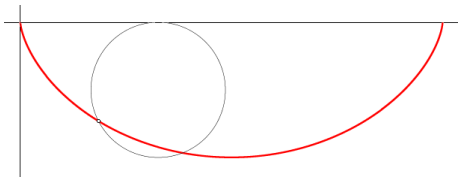
Bezeichnung: F heißt *Zielfunktion*, X *zulässiger Bereich*, jedes $x \in X$ *zulässiger Vektor bzw. Element*, $\hat{x}$ eine *(globale) Lösung* von (P) und $F(\hat{x})$ der *Wert* von (P).

Ein wesentlicher Teil der Problematik besteht meist darin, dass zwar die Zielfunktion F explizit vorliegt, der zulässige Bereich X aber nur implizit gegeben ist, etwa durch Systeme von Gleichungen oder Ungleichungen. Daher zerfällt schon die Grundaufgabe (P) in mehrere Teile:

1. Frage $X = \emptyset$?
2. für $X \neq \emptyset$:
   (a) $F(x)$ beschränkt auf X, d.h. $\inf\{F(x) : x \in X\} > -\infty$?
   Wird dann das Infimum auch angenommen ("Minimum")?
   (b) Wenn ja: berechne ein $\hat{x} \in X$ mit $F(\hat{x}) \leq F(x)\, \forall x \in X$.

Die einsetzbaren Methoden unterscheiden sich auch nach der Art und Anzahl der "Freiheitsgrade", die in der Menge X auftreten. Die Frage, ob ein Minimum oder Maximum gesucht wird, ist aber unerheblich, Eines kann durch Übergang zu $-F(x)$ in das Andere überführt werden.

Beispiel 1.1.1 a) Problem der *Brachistochrone* von Galilei:
Ein Körper soll nur durch den Einfluß der Schwerkraft zwischen zwei Punkten bewegt werden. Gesucht ist die Kurve, auf der der Körper in minimaler Zeit vom höheren zum niederen Punkt kommt.
Johann Bernoulli: Lösung ist Zykloide

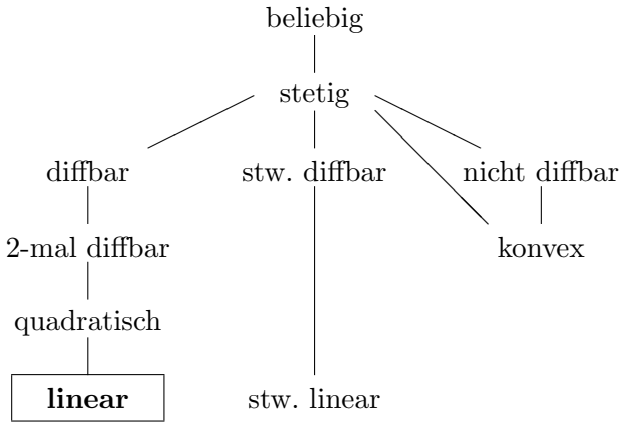


b) Transportproblem: *Ein Unternehmen mit mehreren Produktionsstandorten beliefert verschiedene Abnehmer mit seinen Produkten (Massen-/Stückgut). Gesucht ist ein Transportplan mit möglichst geringen Kosten*

Einordnung der Beispiele: Da die Weghöhe beim Brachistochronen-Problem an jedem reellen Punkt s der Strecke unbekannt ist, hat man eine *unendliche Anzahl an Freiheitsgraden* (überabzählbar). Zur korrekten Beschreibung wäre die Menge X als ein Raum geeigneter Funktionen $x(s)$, $s \in [a, b]$, zu wählen. Derartige Probleme werden in der Variationsrechnung und Steuerungstheorie (*optimal control*) behandelt. Beim Transportproblem sind dagegen die endlich vielen, vom Produktionsort P_i zum Kunden K_j zu liefernden Mengen unbekannt. Bei Massengütern können diese (nichtnegative) reelle Werte, bei Stückgütern ganzzahlige Werte annehmen. Die Grundmenge X ist also (ein Teil) eines geeigneten $\mathbb{R}^n$ oder $\mathbb{Z}^n \subseteq \mathbb{R}^n$. In dieser Vorlesung wird nur der Fall $X \subseteq \mathbb{R}^n$ behandelt.

Eine weitere Klassifikation des Problems ergibt sich aus den

Eigenschaften der Zielfunktion F:



Die Gestalt des *zulässigen Bereichs* X ist in der Regel nicht explizit bekannt, sondern durch Einschränkungen an die Parameter x. Die Art dieser *Nebenbedingungen* schränkt ebenfalls die Auswahl möglicher Verfahren ein. Daher ist es zweckmäßig, die Nebenbedingungen aufzuteilen in funktionale und mengenmäßige. Ab jetzt sei also

$$X := \{x \in \mathbb{R}^n : \ f(x) \leq 0, \ g(x) = 0, \ x \in C\}, \tag{1.1.1}$$

mit $f : \mathbb{R}^n \to \mathbb{R}^p$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $C \subseteq \mathbb{R}^n$. Generell werden Ungleichungen wie in dieser Beschreibung komponentenweise verstanden, $f_i(x) \leq 0$, $i = 1, \ldots, p$, für $f = (f_i)_{i=1}^p$. Auch die Eigenschaften der Funktionen f, g gehen in die Klassifikation von Optimierungsproblemen ein, da durch Umformulierungen mit Zusatzvariablen wie $x_{n+1} := F(x)$, die Zielfunktion auch in Nebenbedingungen verlagert werden kann. Als Grundmengen C treten oft folgende Fälle auf

- $\mathbb{R}^n$, $\mathbb{R}^n_+$, $\mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}$ — die Nichtnegativität ließe sich auch bei f unterbringen
- $B_r(y)$ — Kugel um y vom Radius r, allgemeiner: Ellipsoid
- $\mathbb{Z}^n$, $\mathbb{R}^{n_1} \times \mathbb{Z}^{n_2}$ — ganzzahlige, gemischt-ganzzahlige Probleme,
- $\mathbb{B}^n = \{0, 1\}^n$ — boolesche Optimierungsprobleme.

In dieser Vorlesung werden nur *Lineare Programme* (LP) behandelt, das sind *kontinuierliche Optimierungsprobleme* ($C = \mathbb{R}^n$) mit Funktionen

$$F(x) = c^\mathsf{T} x + d, \;\; f_i, \;\; g_j \text{ affin linear.}$$

Bei einer (in der Praxis üblichen) großen Anzahl von Unbekannten n ist eine Sonderbehandlung bei speziellen Strukturen sinnvoll, etwa bei linearen *Transport-* oder *Fluß-Problemen*. Lösungsmethoden für Optimierungsprobleme haben offensichtlich im Unternehmensbereich (Kostenminimierung) eine erhebliche ökonomische Bedeutung. Aber auch in theoretischer Hinsicht (Komplexitätstheorie) sind sie eine große Herausforderung. Naheliegende Fragestellungen sind:

Theorie:
- Allgemeine Aussagen, z.B. zur Struktur
- Existenz und Eindeutigkeit
- Kriterien für Optimalität
- Empfindlichkeit der Lösungen (Stabilität des *Problems*)
- Komplexität des *Problems*

Praxis:
- Algorithmenentwicklung
- Empfindlichkeit der berechneten Lösung (Stabilität des *Algorithmus*)
- Komplexität des *Algorithmus*

In die erste Kategorie fallen bei Linearen Programmen Erkenntnisse zur Geometrie des zulässigen Bereichs X. Diese hat zentrale Bedeutung, denn X ist ein konvexes Polyeder (Vielflächner), das Minimum wird auf dem Rand angenommen, da nicht-konstante lineare Funktionen keine inneren Extrema besitzen. Daher werden in §3 auch Grundlagen der Konvexen Geometrie behandelt.

1.2 Beispiele

Produktionsplanung

In einem Unternehmen können n verschiedene Produkte P_j erzeugt werden unter Nutzung von m unterschiedlichen Resourcen R_i (Arbeitszeit, Rohstoffe, Energie,...). Der Gewinn bei Produktion einer Einheit von Produkt P_j sei c_j.

Die zu erzeugende Menge des Produkts P_j wird als Unbekannte x_j eingeführt. Eine triviale Nebenbedingung ist offensichtlich $x_j \geq 0$, der erzielte Gesamtgewinn ist $\sum_{j=1}^n c_j x_j = F(x_1, \ldots, x_n)$ und stellt die *Zielfunktion* des Problems dar. Nimmt man weiter an, dass zur Poduktion von P_j jeweils a_{ij} Einheiten von durch Größen b_i beschränkte Resourcen R_i, $i = 1, \ldots, m$, verwendet werden, sind ausserdem die Restriktionen

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

einzuhalten. Insgesamt lautet das Problem somit

$$\begin{aligned} &\max \sum_{j=1}^{n} c_j x_j \\ &\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = 1, \dots, m \\ &x_i \ge 0, \quad i = 1, \dots, n \end{aligned}$$

Hier bietet sich die Vektor-/Matrix-Notation für eine kompaktere Schreibweise an. Mit $x = (x_1, \dots, x_n)^\mathsf{T}$, $c := (c_1, \dots, c_n)^\mathsf{T}$, $b = (b_1, \dots, b_m)^\mathsf{T}$, $A = (a_{ij})_{i,j=1}^{m,n}$ ist $F(x) = c^\mathsf{T} x$ und man hat die äquivalente Formulierung

$$\begin{aligned} \max \quad & c^\mathsf{T} x \\ & Ax \le b \\ & x \ge 0. \end{aligned}$$

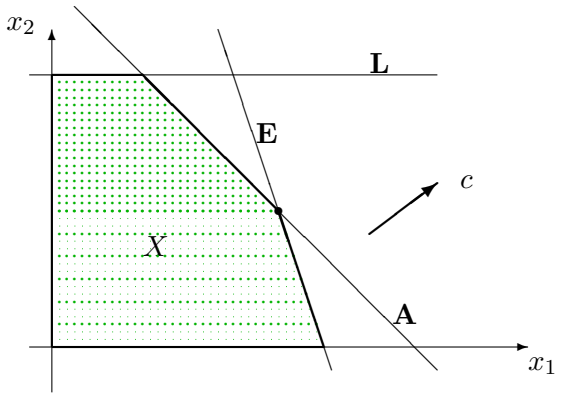
Die Ungleichungen bei Vektoren sind dabei wieder komponentenweise zu verstehen. Da alle Restriktionen Ungleichungen sind, ist der zulässige Bereich $X := \{x \in \mathbb{R}^n : \ Ax \le b,\, x \ge 0\}$.

Beispiel 1.2.1 Fall $n = 2$, $m = 3$, die Produkte P_1 (Gewinn $c_1 = 4$ EUR) und P_2 (Gewinn $c_2 = 3$ EUR) sollen mit Hilfe der Resourcen **A**rbeitszeit, **L**agerkapazität, **E**nergie produziert werden. Die Einschränkungen seien

A:	$x_1 + x_2 \le 16$	(gleicher Arbeitsaufwand)
L:	$x_2 \le 12$	(Rohstoffe nur für P_2 zu lagern)
E:	$3x_1 + x_2 \le 36$	(3-facher Energiebedarf P_1)

Gesamtformulierung und zulässiger Bereich:

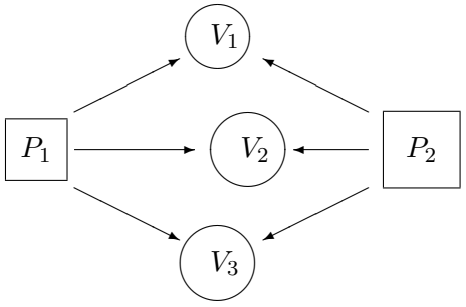
$$\begin{aligned} &\max \quad (4,3) \cdot x \\ &\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 3 & 1 \end{pmatrix} x \le \begin{pmatrix} 16 \\ 12 \\ 36 \end{pmatrix}, \\ &x \ge 0. \end{aligned}$$



Der Pfeil c ist der (konstante!) Gradient der Zielfunktion $F(x) = c^\mathsf{T} x = 4x_1 + 3x_2$, das Maximum wird im markierten **Rand**punkt $(\hat{x}_1, \hat{x}_2) = (10, 6)$ angenommen mit dem Wert $F(\hat{x}) = 58$.

Transportprobleme

Hier soll ein Massengut (beliebig teilbar) von m Produktions-/Lagerstätten P_i mit Kapazität s_i zu n Verbrauchern V_j mit Bedarf r_j transportiert werden. Die Gesamtmengen bei Produktion und Verbrauch sollen dabei gleich sein $\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} r_j$ (oBdA).



Als Unbekannte werden die von P_i nach V_j transportierten Mengen $x_{ij} \geq 0$ eingeführt, der Transport einer Einheit auf dieser Strecke habe den Preis c_{ij}. Für den optimalen *Transportplan*, der minimale Kosten verursacht, ergibt sich das Programm

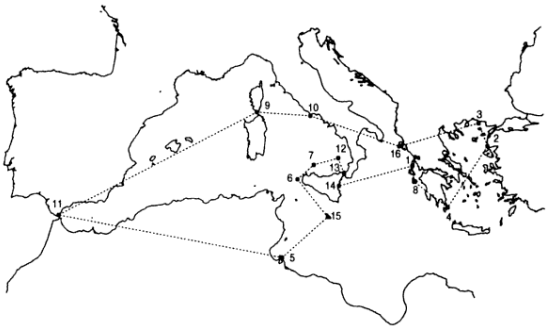
$$\begin{aligned}
&\min \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} && && \text{(Gesamt-Transportkosten)}\\
&\sum_{j=1}^{n} x_{ij} = s_i, && i = 1,\ldots,m && \text{(alle Produkte abtransportiert)}\\
&\sum_{i=1}^{m} x_{ij} = r_j, && j = 1,\ldots,n && \text{(jeder Bedarf abgedeckt)}\\
&x_{ij} \geq 0 && \forall i,j
\end{aligned}$$

Die Restriktionen sind hier ausschließlich lineare Gleichungen und reine Vorzeichen–Bedingungen an alle Variable. Zum LGS gehört ein affin-linearer Lösungsraum, der zulässige Bereich X ist daher der Durchschnitt dieses Lösungsraums mit dem *Positivkegel* $\mathbb{R}^{mn}_+$. Diese Struktur wird bei dem Standard-Lösungsverfahren zugrunde gelegt. Beim Transport von Stückgut sind aber nur ganzzahlige Werte $x_{ij} \in \mathbb{Z}_+$ zulässig. Dann liegt ein ganzzahliges Optimierungsproblem vor.

Modifikation: Transport in Netzwerk (Graph), wenn nur ein Teil der Transportstrecken vorhanden ist. Hierbei können reine Umschlagknoten (ohne Produktion und Verbrauch) auftreten.

Das Problem des Handlungsreisenden (TSP)

Dieses Problem ("traveling salesman problem") hat in der Komplexitätstheorie die Bedeutung eines extrem schwierigen Referenz-Problems. In der Grundform soll ein Reisender eine Anzahl von n Orten je einmal besuchen und zum Ausgangspunkt zurückkehren. Ziel ist eine Tour mit minimaler Gesamtstrecke. Dies ist also die moderne Form der klassischen *Odyssee* (rechts: eine optimale Lösung derselben).



Dazu sei $N = \{1,\ldots,n\}$ die Menge der Orte und $w_{ij} \geq 0$ die Entfernung von i nach j. Ist die Rundreise (*Tour*) gegeben durch die Liste $(p(1),\ldots,p(n))$ der besuchten Orte, so können in der Gesamtstrecke $\sum_{j=1}^{n-1} w_{p(j)p(j+1)} + w_{p(n)p(1)}$ die Summanden w offensichtlich nach dem ersten Index umsortiert werden. Im zweiten Index steht dann eine *zyklische Permutation* $\pi \in S_n$ mit $\pi(p(j)) = p(j+1)$. Die Menge der zyklischen n-Permutationen $S_{z,n} \subseteq S_n$ enthält alle diejenigen, welche aus einem einzigen Zyklus bestehen. Das Problem lautet daher

$$\min\{\sum_{i=1}^{n} w_{i,\pi(i)} : \ \pi \in S_{z,n}\} \qquad \text{(TSP)}$$

In der allgemeinen Form sind die Entfernungsangaben $w_{ij} \geq 0$ nicht weiter eingeschränkt. Sinnvolle Spezialfälle sind aber offensichtlich das

symmetrische TSP:	$w_{ij} = w_{ji}$	(z.B., keine Einbahnstraßen)
euklidsche TSP:	$w_{ij} \leq w_{ik} + w_{kj}\, \forall i,j,k$	(Gültigkeit der Dreieckungleichung)

In der Form (TSP) liegt ein kombinatorisches Optimierungsproblem vor. Wegen $|S_{z,n}| = (n-1)!$ ist eine reine Enumeration aller Möglichkeiten zur Lösung nur für kleine n möglich, denn, z.B., ist $5! = 120$, $10! = 368800$, $30! > 2 \cdot 10^{32}$. Der z.Z. schnellste Rechner (Tianhe-2 mit 33800 Tera-FLOPS $>$ 33 PetaFLOPS) schafft ca. $3 \cdot 10^{21}$ Operationen pro Tag.

Eine alternative Formulierung als (LP) ist möglich durch Betrachtung des charakteristischen Vektors $x = (x_{ij}) \in \mathbb{B}^k$, $k = n(n-1)$ beim allgemeinen und $k = \binom{n}{2} = n(n-1)/2$ beim symmetrischen Problem. Beim symmetrischen Problem haben die Variablen x_{ij}, $i < j$, folgende Bedeutung

$$x_{ij} = \begin{cases} 1 & \text{der Weg zwischen } i \text{ und } j \text{ wird benutzt,} \\ 0 & \text{sonst.} \end{cases}$$

Damit sich eine Tour ergibt, müssen zu jedem Ort genau zwei Wege benutzt werden, also

$$\sum_{j<i} x_{ji} + \sum_{j>i} x_{ij} = 2 \quad \forall 1 \leq i \leq n. \tag{1.2.1}$$

Allerdings sind dadurch Teiltouren noch nicht ausgeschlossen. Zusätzlich kann man dazu fordern, dass in keiner echten Teilmenge $U \subseteq N$ ein Kreis auftritt, $\sum_{i,j \in U} x_{ij} \leq |U| - 1$, bzw. die Menge wieder verlassen wird

$$\sum_{i \in U, j \notin U} x_{ij} \geq 2 \quad \forall U \subset N,\ 1 \leq |U| \leq n-1. \tag{1.2.2}$$

Diese Formulierung des (TSP) ist damit

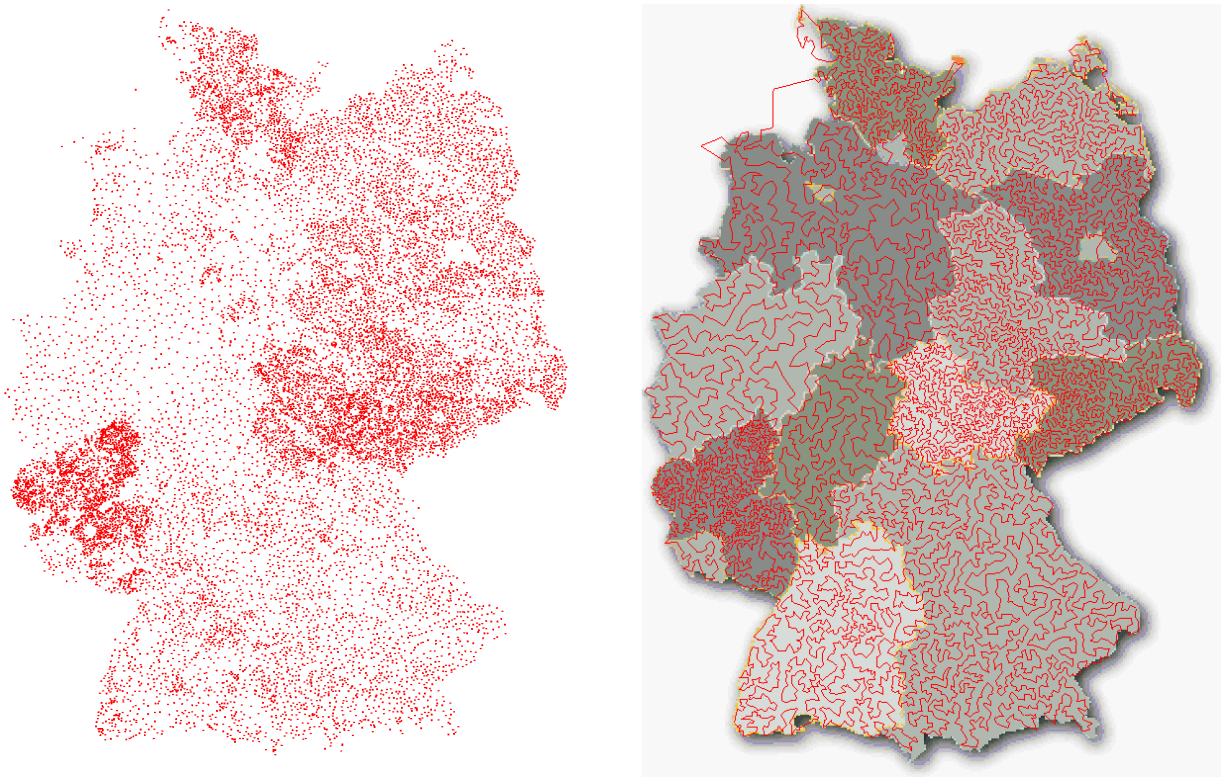
$$\begin{gathered} \min \textstyle\sum_{i,j=1}^{n} w_{ij} x_{ij} \\ x \in X := \{x \in \mathbb{B}^{\binom{n}{2}} : (1.2.1), (1.2.2) \text{ gelten}\}. \end{gathered} \tag{TSPB}$$

Dieses (TSPB) ist also ein boolesches lineares Programm mit n Gleichungen und $\sum_{k=1}^{n-1} \binom{n}{k} = 2^n - 2$ Ungleichungen. Wegen dieser vielen Bedingungen und der booleschen Variablen ist auch diese (und jede) Form des (TSP) schwierig zu lösen.

Daten zur Geschichte des Problems, Lösungsrekorde:

1930	Karl Menger	Formulierug des Problems, einzige Lösungs-
1934	Hasler Whitney	möglichkeit vollständige Enumeration
1954	G.B. Dantzig, D.R. Fulkerson,	Lösen 42-Städte-Problem mit Schnittebenen-
	S.M. Johnson	Verfahren und linearen Programmen,
1972	R.M Karp	TSP ist NP-vollständig,
1979	Crowder, Padberg	318 Orte, Branch-and-Cut-Verfahren,
1995	Applegate, Bixby, Chvátal, Cook	7397-Städte-Problem, Parallelrechner
2001	dito	15112 Städte Deutschland
2004	dito+Helsgaun	24978 Städte Schweden
2006	A+B+C+C+E+G+H	85900 Punkte VLSI (s.u.)

Der aktuelle Rekord (www.tsp.gatech.edu/) berechnet die optimale Rundreise durch 85900 Punkte einer VLSI-Schaltung, ein Vorgänger-Rekord 2001 betraf 15112 deutsche Städte (elib.zib.de):



Statt des Booleschen Problems (TSPB) kann man auch seine *stetige Relaxation* betrachten, mit dem zulässigen Bereich

$$X_1 := \{x \in \mathbb{R}^{\binom{n}{2}} : \ 0 \le x \le \mathbb{1}, \text{ und } (1.2.1), (1.2.2)\} \supset X. \tag{1.2.3}$$

Da dessen zulässige Menge X umfaßt, erhält man daraus zumindestens eine untere Schranke W_1 für den Wert W des (TSPB): $W \ge W_1$. Bei den erwähnten *Schnittebenen-Verfahren* legt man tatsächlich (1.2.3) zugrunde und eliminiert schrittweise unbrauchbare Lösungen durch Hinzunahme weiterer Nebenbedingungen, die nichtganzzahlige Lösungen *abschneiden*.

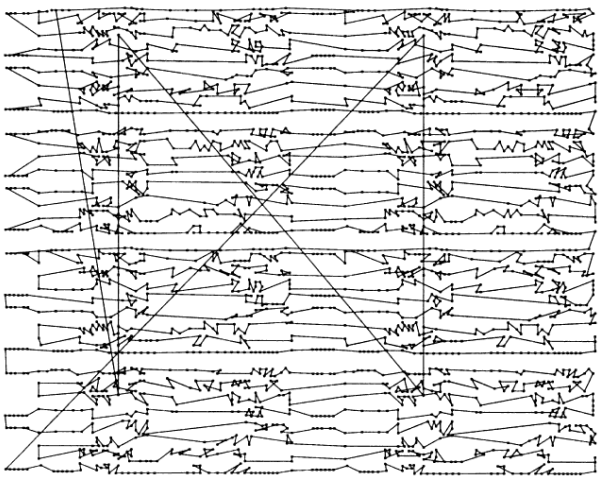
Anwendungen Viele praktische Fragen lassen sich als TSP formulieren:

— Leiterplatinen-Produktion, Computerverdrahtung

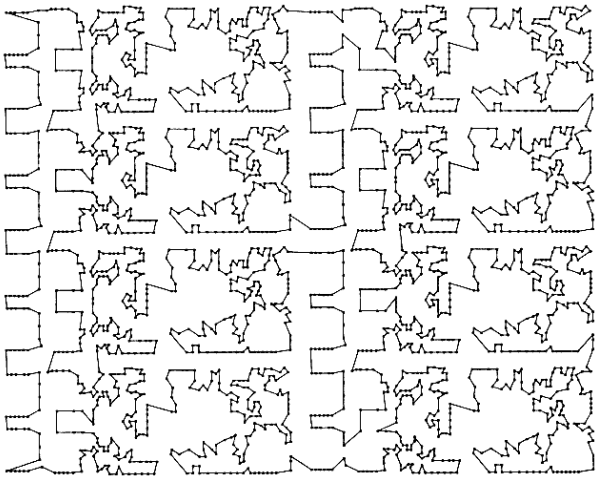
— Tourenplanung

— Ablaufplanung (job-shop scheduling)

Zur Bestückung von Platinen mit Bauteilen sind für deren Anschlußdrähte Bohrungen in den Leiterplatten anzubringen. Da die Zeit pro Bohrung konstant ist, wird die Gesamtzeit v.a. durch die Fahrzeit zwischen den Bohrpunkten bestimmt. Unter der Annahme, dass die Fahrzeit proportional zur Entfernng ist, entspricht c_{ij} dem euklidschen Abstand der Punkte. Die im folgenden Beispiel mit $n = 2392$ Punkten per Hand geplante Tour ist um 90% länger als die optimale.



"manuelle" Lösung mit Länge 718876



Optimale Lösung der Länge 378032

1.3 Lineare Programme

Für Lineare Optimierungsprobleme hat sich der Begriff *Lineare Programme* eingebürgert. In dem allgemeinen Rahmen der Form (P) mit dem zulässigen Bereich (1.1.1) sind alle auftretenden Funktionen (affin) linear, es gelten also Darstellungen der Form

$$F(x) = c^\mathsf{T} x, \quad f_i(x) = a_i^\mathsf{T} x + \alpha_i,\ g_j(x) = b_j^\mathsf{T} x + \beta_j,$$

mit Vektoren $a_i, b_j \in \mathbb{R}^n$, $i = 1, \ldots, p$, $j = 1, \ldots, m$. Dabei wurde F oBdA als linear angenommen, da eine Konstante zwar den Wert des Problems, aber nicht die Lösung $\hat{x}$ ändert. In den Beispielen traten Ungleichungsrestriktionen of in sehr einfacher Form auf, als reine Vorzeichenbeschränkungen. Wegen ihrer vielfältigen Sonderrolle werden diese im folgenden gesondert notiert, man teilt die Unbekannten auf in *freie* und *vorzeichenbeschränkte Variable*. Zusammen mit der Aufteilung in Ungleichungen und Gleichungen können die Restriktionen in einer Blockmatrix gesammelt werden. Die allgemeine Form eines linearen Programms lautet daher

$$\left.\begin{array}{rl} \min\ c_1^\mathsf{T} x_1 + c_2^\mathsf{T} x_2 & \\ A_{11}x_1 + A_{12}x_2 & \geq b_1 \\ A_{21}x_1 + A_{22}x_2 & = b_2 \\ x_1 & \geq 0 \end{array}\right\} \quad \begin{array}{c} x_1, c_1 \in \mathbb{R}^{n_1},\ x_2, c_2 \in \mathbb{R}^{n_2},\ n = n_1 + n_2, \\ b_1 \in \mathbb{R}^{m_1},\ b_2 \in \mathbb{R}^{m_2},\ m = m_1 + m_2, \\ A_{ij} \in \mathbb{R}^{m_i \times n_j},\ i, j = 1, 2. \end{array} \tag{LP}$$

Allerdings kann man durch elementare Umformungen daraus auch jedes der folgenden, einfacheren Standardprogramme erzeugen mit $A \in \mathbb{R}^{m \times n}$,

$$\min\{\, c^\mathsf{T} x :\ Ax \geq b\} \tag{LP1}$$

$$\min\{\, c^\mathsf{T} x :\ Ax \geq b,\ x \geq 0\} \tag{LP2}$$

$$\min\{\, c^\mathsf{T} x :\ Ax = b,\ x \geq 0\} \tag{LP3}$$

Bei diesen ist in der allgemeinen Form (LP) jeweils nur ein Matrixblock nichttrivial, nämlich $A_{12} \neq 0$ bei (LP1), $A_{11} \neq 0$ bei (LP2) und $A_{21} \neq 0$ bei (LP3). Folgende *elementare Umformungen* können eingesetzt werden, die auf äquivalente Probleme führen:

1. eine Gleichung $a^\mathsf{T}x = \alpha$ kann durch die beiden Ungleichungen $a^\mathsf{T}x \geq \alpha$, $-a^\mathsf{T}x \geq -\alpha$ ersetzt werden.

2. eine freie Variable ξ kann als Differenz $\xi = \xi^+ - \xi^-$ von zwei nichtnegativen Variablen $\xi^+, \xi^- \geq 0$ geschrieben werden.

3. Ungleichungen $a^\mathsf{T}x \geq \alpha$ können durch Einführung einer *Schlupfvariablen* $\eta \geq 0$ durch die Gleichung $a^\mathsf{T}x - \eta = \alpha$ ersetzt werden.

4. jede Vorzeichenbeschränkung $\xi \geq 0$ kann als Ungleichungsrestriktion $\xi \geq 0$ einer freien Variablen ξ nach A_{12} verlagert werden.

Durch diese Umformungen können sich die Dimensionen m, n vergrößern, die wesentlichen Eigenschaften aus §1.1 ($X \neq \emptyset$? $\inf\{F(x) : x \in X\} > -\infty$?) bleiben aber unverändert. Allerdings unterscheiden sich die geometrischen Eigenschaften der zulässigen Bereiche bei den 3 Standardformen. Dies eröffnet die Möglichkeit, je nach Fragestellung die passende zu wählen, es gilt:

(LP1) $X = \{x : Ax \geq b\} = \bigcap_{i=1}^m \{(e_i^\mathsf{T}A)x \geq b_i\}$ mit den Einheitsvektoren $e_i \in \mathbb{R}^n$. Da jede Ungleichung der Form $a^\mathsf{T}x \geq \beta$ einen abgeschlossenen *Halbraum* definiert, ist X als Durchschnitt von Halbräumen ein *Polyeder*. Hier erwartet man Dimensionen $m > n$.

(LP2) $X = \{x : Ax \geq b, x \geq 0\}$ ist Durchschnitt des gerade erwähnten Polyeders mit dem *positiven Kegel* $\{x \in \mathbb{R}^n : x \geq 0\} = \mathbb{R}^n_+$, also wieder ein *Polyeder*.

(LP3) $X = \{x : Ax = b, x \geq 0\}$ ist als Durchschnitt $U \cap \mathbb{R}^n_+$ ein "dünnes" Polyeder. Dabei wird der Positivkegel geschnitten mit dem affinen Unterraum $U := \{x : Ax = b\} = \{\hat{x}\} + kern(A)$ aller Lösungen des Gleichungssystems. Für einen Kern ist i.d.R. $m < n$ erforderlich.

2 Simplex – Verfahren

2.1 Bezeichnungen

Es wird der n-dimensionale Vektorraum $\mathbb{R}^n$ zugrundegelegt. Die Vektoren der Einheitsbasis heißen $e_i = (\delta_{ij})_{j=1}^n$ und es sei $\mathbb{1} := \sum_{i=1}^n e_i$ der Vektor aus Einsen. Allgemein werden Elemente $x \in \mathbb{R}^n$ als *Spaltenvektoren* geschrieben,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \Big(x_i\Big)_{i=1}^n.$$

Meist wird die Euklidnorm $\|x\| = \|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}$ verwendet, eine andere interessante Norm ist die Maximumnorm $\|x\|_\infty := \max_{i=1}^n |x_i|$. Ungleichungen zwischen Vektoren sind komponentenweise zu verstehen. Eine solche wird in der Definition $\mathbb{R}^n_+ := \{x : \ x \geq 0\}$ des *positiven Kegels* verwendet (s.o.). Die Menge der reellen $m \times n$-Matrizen heißt $\mathbb{R}^{m\times n}$. Im Folgenden werden oft Untermatrizen aus ausgewählten Spalten oder Zeilen einer Matrix betrachtet. Zu $A \in \mathbb{R}^{m\times n}$ seien daher $a_j = Ae_j \in \mathbb{R}^m$ die Spalten und $a^{(i)} = A^\mathsf{T} e_i \in \mathbb{R}^n$ die Zeilen von A. Dann gelten folgende Schreibweisen

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (a_{ij}) = (a_1, \ldots, a_n) = \begin{pmatrix} {a^{(1)}}^\mathsf{T} \\ \vdots \\ {a^{(m)}}^\mathsf{T} \end{pmatrix}.$$

Elemente einer Vektorfolge werden ebenfalls durch einen oberen Index unterschieden, $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})^\mathsf{T}$.

2.2 Matrix – Umformungen

Das später behandelte *Simplex-Verfahren* benutzt die Problemform (LP3) und durchläuft spezielle Lösungen des Linearen Gleichungssystems $Ax = b$, $m < n$, welche durch reguläre quadratische Untermatrizen von A gegeben sind. Die Lösung von regulären Gleichungssystemen spielt daher eine zentrale Rolle bei der Optimierung. Zwischen aufeinanderfolgenden Schritten des Simplexverfahrens ändern sich die Systeme aber nur wenig. Um Rechenaufwand zu sparen nutzt man daher oft Aktualisierungs-Formeln ("matrix update"). Denn bei Änderung einer Matrix durch eine Rang-1-Matrix ist die Inverse explizit bekannt und läßt sich effizient berechnen.

Satz 2.2.1 *Die Matrix $B \in \mathbb{R}^{m\times m}$ sei regulär, mit Vektoren $u, v \in \mathbb{R}^m$ sei $\beta := 1 + v^\mathsf{T} B^{-1} u \neq 0$. Dann ist auch die Matrix $B + uv^\mathsf{T}$ regulär und ihre Inverse ist*

$$(B + uv^\mathsf{T})^{-1} = B^{-1} - \frac{1}{1 + v^\mathsf{T} B^{-1} u}\, B^{-1} uv^\mathsf{T} B^{-1}. \tag{2.2.1}$$

Wenn dabei in B nur die Spalte Nummer $s \in \{1,\dots,m\}$ durch einen anderen Vektor a ersetzt wird, d.h., $v = e_s$ und $u = a - b_s$ gilt, ist $\beta = e_s^\mathsf{T} B^{-1} a$ und die Zeilen der Inversen ändern sich nach den Regeln Bew

$$e_i^\mathsf{T}(B + ue_s^\mathsf{T})^{-1} = \begin{cases} \frac{1}{\beta}\, e_s^\mathsf{T} B^{-1}, & i = s, \\ e_i^\mathsf{T} B^{-1} - e_i^\mathsf{T} B^{-1} a \Big(\frac{1}{\beta} e_s^\mathsf{T} B^{-1}\Big), & i \neq s. \end{cases} \tag{2.2.2}$$

In den Zeilen mit $i \neq s$ treten insbesondere die durch die Klammer hervorgehobenen Werte der neuen Zeile s auf. Einfacher ist die Formel (2.2.1) für den Fall $B = I$ mit $(I+uw^\mathsf{T})^{-1} = I - \frac{1}{\beta} uw^\mathsf{T}$, $\beta = 1 + w^\mathsf{T} u$. Aber auch hieraus folgt schon die allgemeine Version, denn mit $w^\mathsf{T} := v^\mathsf{T} B^{-1}$ ist

$$(B + uv^\mathsf{T})^{-1} = \Big((I + uw^\mathsf{T})B\Big)^{-1} = B^{-1}(I - \frac{1}{\beta} uw^\mathsf{T}) = B^{-1} - \frac{1}{\beta} B^{-1} uv^\mathsf{T} B^{-1}.$$

Die Formel (2.2.2) wird in der klassischen Tabellenform des Simplexverfahrens (Handrechnung) benutzt, da der Rechenaufwand bei $O(m^2)$ arithmetischen Operationen (FLOP: *FLoating point OPeration*) liegt. Dies hat aber den Nachteil, dass sich bei größeren Problemen und insbesondere für kleine Werte β *Rundungsfehler* ansammeln.

Für große (Computer-) Anwendungen greift man zur Lösung auf den *Gauß-Algorithmus* oder verwandte Methoden zurück. Auch dieser läßt sich so anpassen, dass geringfügige Änderungen der Matrix mit geringem Aufwand berücksichtigt werden können. Dazu ist es nützlich, die Zeilenumformungen im Gauß-Algorithmus als Matrixmultiplikation zu interpretieren. Mit $z \in \mathbb{R}^m$ und $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ betrachtet man

$$L_j(z) := \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -z_{j+1} & 1 & \\ & & \vdots & & \ddots \\ & & -z_m & & & 1 \end{pmatrix}, \quad L_j(z)A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ a_{j+1,1} - z_{j+1}a_{j1} & \dots & a_{j+1,n} - z_{j+1}a_{jn} \\ \vdots & & \vdots \\ a_{m1} - z_m a_{j1} & \dots & a_{mn} - z_m a_{jn} \end{pmatrix}.$$

Die Matrix L_j beschreibt also den Effekt einer vollständigen Elimination in Spalte j und läßt sich auch kompakt in der Form $L_j = I - ze_j^\mathsf{T}$ schreiben. Wegen $e_j^\mathsf{T} z = 0$ ist ihre Inverse nach (2.2.1) einfach $L_j^{-1} = I + ze_j^\mathsf{T}$. Beim Gauß-Algorithmus werden der Reihe nach Umformungen $A \to L_1 A \to L_2 L_1 A$ etc. angewendet, um die Matrix auf *obere Dreieckgestalt* (Stufenform) zu bringen. Da Produkte von unteren Dreieckmatrizen wieder solche Dreieckmatrizen sind, kann das Ergebnis des Gauß-Algorithmus folgendermaßen zusammengefaßt werden.

Satz 2.2.2 *Wenn der einfache Gauß-Algorithmus, der die Matrix $A = A_1 \in \mathbb{R}^{m\times n}$, $m \le n$, mit Zeilenumformungen $A_{j+1} = (a_{ik}^{(j+1)}) := L_j(z^{(j)})A_j$, $j = 1,\dots,m-1$, und*

$$z^{(j)} = \frac{1}{a_{jj}^{(j)}} \Big(0,\dots,0,a_{j+1,j}^{(j)},\dots,a_{mj}^{(j)}\Big)^\mathsf{T}, \tag{2.2.3}$$

in obere Dreieckgestalt $R := A_m$ überführt, durchführbar ist ($a_{jj}^{(j)} \neq 0 \forall j$), erzeugt er eine LR-Zerlegung *der Matrix als Produkt einer unteren Dreieckmatrix $L = L_1^{-1} \cdots L_{m-1}^{-1}$ und einer oberen $R = A_m$:*

$$A = LR, \quad L = \begin{pmatrix} 1 & & & \\ z_2^{(1)} & 1 & & \\ \vdots & & \ddots & \\ z_m^{(1)} & \dots & z_m^{(m-1)} & 1 \end{pmatrix}, \; R = \begin{pmatrix} r_{11} & r_{12} & \dots & \dots & . & r_{1n} \\ & r_{22} & \dots & \dots & . & r_{2n} \\ & & \ddots & & & \vdots \\ & & & r_{mm} & . & r_{mn} \end{pmatrix}.$$

Die Berechnung der LR-Zerlegung hat einen Aufwand von i.w. $(n - \frac{1}{3}m)m^2$ arithmetischen Operationen, also $\frac{2}{3}m^3$ FLOP für $m = n$.

Im Satz wurde implizit vorausgesetzt, dass die *Pivot-Elemente* $a_{jj}^{(j)} = r_{jj}$, durch welche dividiert wird, von Null verschieden sind. Bei einer Rechnung mit Maschinenzahlen endlicher Genauigkeit muß aber nicht nur der Fall $a_{jj}^{(j)} = 0$ durch Zeilenvertauschungen vermieden werden, sondern auch die Verwendung kleiner Pivot-Werte $a_{jj}^{(j)} \cong 0$. Sonst zeigen sich die gleichen Probleme wie bei Verwendung der Rang-1-Formel (2.2.2). Daher bringt man durch Vertauschungen möglichst große Elemente in die Hauptdiagonale (s.u.).

Durch Berechnung einer LR-Zerlegung wird die Berechnung der Inversen absolut überflüssig. Denn mit der Zerlegung kostet die Auflösung eines quadratischen linearen Gleichungssystem $Bx = c$ nur noch den Aufwand der Lösung von zwei gestaffelten (Dreieck-) Systemen:

$$x = B^{-1}c = R^{-1}L^{-1}c \iff Ly = c, \; Rx = y.$$

Außerdem kann diese Auflösung ohne Zusatzvariable (am Platz) durchgeführt werden. Die folgenden Anweisungen überschreiben die rechte Seite $c = (c_i)$ zunächst mit der Zwischenlösung y, dann mit der Gesamtlösung x:

löst $Ly = c$, $c := y$

für $i = 2$ bis m {
  für $j = 1$ bis $i - 1$ { $c_i := c_i - l_{ij}c_j$;}
}

löst $Rx = c$, $c := x$

für $i = m$ abwärts bis 1 {
  für $j = i + 1$ bis m {$c_i := c_i - r_{ij}c_j$;}
  $c_i := c_i / r_{ii}$;
}

Der Rechenaufwand beträgt pro Teilsystem i.w. m^2 Operationen. Damit ist der Gesamtaufwand zur Lösung von $Bx = LRx = c$ mit $2m^2$ Operationen *nicht höher* als die reine Multiplikation $B^{-1}c$, jeweils für jede neue rechte Seite c.

Zeilenvertauschungen bei einer $m \times n$-Matrix A können formal mit Hilfe einer Permutationsmatrix $P \in \mathbb{B}^{m \times m}$ dargestellt werden. So wird etwa mit einer Permutation π die entsprechende Umordnung der Zeilen in $A = (a_{ij})$ folgendermaßen bewirkt (δ: Kronecker-Symbol):

$$A' = (a'_{kj}) = (a_{\pi(i),j}) \iff A' = PA, \; P = \left(\delta_{\pi(i),j}\right)_{i,j=1}^{m}.$$

Permutationsmatrizen entstehen durch Vertauschungen bei der Einheitsmatrix und sind unitär, die Transponierte $P^{\mathsf{T}} = P^{-1}$ bewirkt die inverse Permutation. In der praktischen Realisierung bestimmt man im Gaußalgorithmus vor Elimination der j-ten Spalte das betragsmaximale

Element unterhalb von a_{jj} und tauscht dessen Zeile mit der j-ten. Dann ist $a_{jj}^{(j)}$ in (2.2.3) betragsmaximal und alle Elemente von L daher im Betrag kleiner gleich eins. Die Permutationen protokolliert man am Besten in einem Indexfeld P[1..m], in dem man alle Zeilenvertauschungen der Matrix A synchron durchführt. Der obige Satz 2.2.2 kann damit in folgender Weise verallgemeinert werden:

Für jede reguläre Matrix $A \in \mathbb{R}^{m\times m}$ gibt es eine Permutationsmatrix P so, dass die LR-Zerlegung $PA = LR$ existiert.

Beispiel 2.2.3 Die folgende Matrix A besitzt offensichtlich keine LR-Zerlegung, da schon das erste Pivotelement verschwindet,

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix}. \quad \text{Mit } P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

gilt aber

$$PA = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & \frac{3}{2} \end{pmatrix} = LR.$$

Bei der Elimination ist hier die Diagonale jeweils größer als die Elemente darunter, daher sind tatsächlich alle Beträge im L-Faktor nicht größer als eins.

Anpassung der LR-Zerlegung Der Aufwand bei einem Gauß-Eliminationsschritt, also der "Multiplikation" mit einer Matrix $L_j(z^{(j)})$ ist proportional zur Zahl der nichttrivialen Elemente von $z^{(j)}$, also der Anzahl solcher Elemente in der j-ten Spalte von B_j. Tauscht man in der (quadratischen) Matrix B mit $B = LR$ wieder die Spalte s aus, $C := B + ue_s^{\mathsf{T}}$, $u = a - b_s$, tritt in $L^{-1}C$ dort eine volle Spalte auf, deren Elimination (etwa bei $s = 1$) fast den vollen Aufwand einer Neuzerlegung verursacht. Denn bei Elimination in Spalte s füllt sich der vorher freie Bereich hinter dieser Spalte i.a. vollständig auf! Dies läßt sich dadurch vermeiden, dass man die neue Spalte a am Ende einfügt, und die Spalten $s+1$ bis m nach vorne schiebt:

$$B = \left(\begin{array}{c|c|c} & b_{1s} & \\ & \vdots & \\ & b_{ms} & \end{array}\right) \quad \mapsto \quad C' = \left(\begin{array}{c|c|c} & & a_1 \\ & & \vdots \\ & & a_m \end{array}\right)$$

Der R-Faktor ändert sich dann folgendermaßen mit dem Vektor $c := L^{-1}a$ am Ende:

$$R = L^{-1}B = \left(\begin{array}{c} \overset{s}{} \\ \\ \\ \\ \end{array}\right) \quad \mapsto \quad L^{-1}C' = \left(\begin{array}{c} \overset{s}{} \\ \\ \\ \\ \end{array}\right) \begin{matrix} c_1 \\ \vdots \\ \\ c_m \end{matrix} =: R'.$$

Jetzt tritt ab Spalte s nur je ein Element unter der Diagonale auf, welches man mit Zeilenoperationen, die nur je eine Zeile betreffen (Aufwand $O(m)$ pro Elimination!) eliminieren kann,

evtl. nach Zeilenvertauschung. Dabei wendet man die Umformungen gleichzeitig auf L und $R' = L^{-1}C'$ an, um danach wieder eine gültige LR-Zerlegung von C' zu bekommen. Bei der Elimination von $r'_{s+1,s}$ mit $L_s(z)$ etwa hat $z = \zeta e_{s+1}$ nur ein nichttriviales Element und durch

$$C' = LR' = (LL_s^{-1})\,(L_sR') = \Big(L(I + \zeta e_{s+1}e_s^{\mathsf{T}})\Big)\,\Big((I - \zeta e_{s+1}e_s^{\mathsf{T}})R'\Big),$$

wird beim R-Faktor nur die Zeile $s+1$ geändert, beim L-Faktor nur die Spalte s. Daher ist der Gesamtaufwand für diese Anpassung der LR-Zerlegung in der Größenordnung $O(m^2)$.

2.3 Basen

Bei der numerischen Durchführung der Optimierung geht man vom Programm (LP3) aus

$$\min\{c^{\mathsf{T}}x :\ x \in X\}, \quad X := \{x \in \mathbb{R}^n :\ Ax = b,\, x \geq 0\},$$

und betrachtet ohne Einschränkung den Fall $A \in \mathbb{R}^{m\times n}$, $\mathrm{Rang}(A) = m < n$. Denn für $\mathrm{Rang}(A) < m$ wäre der affine Unterraum $U = \{x :\ Ax = b\}$ entweder leer, oder es könnten Gleichungen entfernt werden.

Der zulässige Bereich $X = \{x :\ Ax = b,\ x \geq 0\} = U \cap \mathbb{R}^n_+$ ist der Schnitt des affinen Unterraums U mit dem positiven Oktanten $\mathbb{R}^n_+$. Da die Zielfunktion $x \mapsto c^{\mathsf{T}}x$ linear ist, ist ihr Gradient c^{T} konstant und daher gibt es keine inneren Extrema. Daher liegt das Optimum auf dem Rand von $X = U \cap \mathbb{R}^n_+$ und somit auf dem Rand des Positivkegels $\mathbb{R}^n_+$. Trivialerweise hat $x \in X$ daher Komponenten, die entweder positiv oder null sind, letzteres insbesondere auf dem Rand von X. Daher sind zur Beschreibung folgende Bezeichnungen nützlich. Zu einem Punkt $x \in \mathbb{R}^n$ sei

$$J^+(x) := \{i : x_i > 0\},\ J^-(x) := \{i : x_i < 0\},\ J(x) := J^-(x) \cup J^+(x)$$

die Menge der (positiven, negativen bzw. aller) *Stützindizes* von x. Für $x \geq 0$ ist $J(x) = J^+(x)$.

In (LP3) kann man zu unterbestimmten Gleichungssystem für eine spezielle Lösung $\bar{x} \in X$ einige Spalten von A "auslassen", denn mit $J^+(\bar{x}) = \{j_1, \ldots, j_\ell\} \subseteq N := \{1, \ldots, n\}$ ist

$$b = A\bar{x} = a_{j_1}\bar{x}_{j_1} + a_{j_2}\bar{x}_{j_2} + \ldots + a_{j_\ell}\bar{x}_{j_\ell}, \quad \ell \leq n. \tag{2.3.1}$$

Dies entspricht einem Gleichungssystem der Dimension $m \times \ell$. Als Bezeichnung wird zur Indexmenge $J = \{j_1, \ldots, j_\ell\} \subseteq \{1, \ldots, n\}$, $|J| = \ell$, daher folgende Untermatrix von A eingeführt

$$A_J = (a_{j_1}, \ldots, a_{j_\ell}) \in \mathbb{R}^{m\times\ell}.$$

Die analoge Bezeichnung (vgl. §2.1) für ausgewählte Zeilen $L = \{l_1, \ldots, l_k\} \subseteq \{1, \ldots, m\}$ der Matrix ist

$$A^{(L)} = \begin{pmatrix} {a^{(l_1)}}^{\mathsf{T}} \\ \vdots \\ {a^{(l_k)}}^{\mathsf{T}} \end{pmatrix} \in \mathbb{R}^{k\times n}. \tag{2.3.2}$$

Wie in (2.3.1) werden damit die verschwindenden Komponenten von $\bar{x}$ aus dem Gleichungssystem $A\bar{x} = b$ entfernt. Denn mit $J := J(\bar{x})$ und dem Komplement $K = N \setminus J$ gilt (etwa nach geeigneter Umordnung) $A = (A_J, A_K)$, $\bar{x}^\mathsf{T} = (\bar{x}_J^\mathsf{T}, \bar{x}_K^\mathsf{T})$ und

$$b = A\bar{x} = \sum_{j=1}^{n} a_j \bar{x}_j = A_J \bar{x}_J + A_K \bar{x}_K = (A_J, A_K) \begin{pmatrix} \bar{x}_J \\ \bar{x}_K \end{pmatrix} \tag{2.3.3}$$
$$J = J(\bar{x}) \;\Rightarrow\; A_J \bar{x}_J = b,\; \bar{x}_K = 0.$$

Dieser Umgang mit Indexmengen hat für die Optimierung eine fundamentale Bedeutung. Man stellt sich dabei vor, dass an jede Matrixspalte und x-Variable ihr Index angeheftet ist und sich in dem Produkt nur zusammenpassende Paare bilden. Umgekehrt berechnet man bei gegebener Indexmenge J aus der letzten Beziehung in (2.3.3) direkt eine *spezielle* Lösung von $Ax = b$, wenn die Untermatrix A_J regulär ist, also insbesondere $|J| = m$ gilt. Eine solche Lösung ist aber nicht unbedingt schon zulässig.

Definition 2.3.1 *a) Ein $\bar{x} \in X$ heißt zulässige* Basislösung*, wenn $Rang(A_{J(\bar{x})}) = |J(\bar{x})|$ ist.*

b) Zu $J = \{j_1, \ldots, j_m\} \subseteq \{1, \ldots, n\}$ heißt A_J Basis*, wenn $B := A_J \in \mathbb{R}^{m\times m}$ regulär ist,* $\det(B) \neq 0$*. Die Basis A_J heißt* zulässig*, wenn $A_J^{-1} b \geq 0$ gilt.*

Zu jeder Basis A_J bekommt man über (2.3.3) die *Basislösung* $\bar{x}^\mathsf{T} = (\bar{x}_J^\mathsf{T}, \bar{x}_K^\mathsf{T})$ mit $\bar{x}_J := A_J^{-1} b$, $\bar{x}_K := 0$, $K = N \setminus J$. Mit einer geeigneten Ergänzung des Systems (2.3.3) durch

$$\begin{pmatrix} A_J & A_K \\ 0 & I_{n-\ell} \end{pmatrix} \begin{pmatrix} \bar{x}_J \\ \bar{x}_K \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \tag{2.3.4}$$

ist auch die ganze Basislösung $\bar{x}$ Lösung eines regulären Systems. Mit $\ell = |J(\bar{x})|$ hat dessen Gesamtmatrix Dimension $(m + n - \ell) \times n$, ihr Rang ist $\mathrm{Rang}(A_J) + n - \ell$ und das System also eindeutig lösbar für $\mathrm{Rang}(A_J) = \ell$. Für $\ell < m$ ist das System (2.3.4) allerdings nicht quadratisch (überbestimmt). Man nennt dann $\bar{x}$ eine *ausgeartete Basislösung*. Allgemein gilt

Satz 2.3.2 *Es sei $\bar{x} \in X$ zulässige Basislösung. Dann besitzt $\bar{x}$ höchstens m positive Komponenten, es gilt also $|J(\bar{x})| \leq m$, und die Untermatrix $A_{J(\bar{x})}$ kann zu einer Basis $A_J \in \mathbb{R}^{m\times m}$, $J \supseteq J(\bar{x})$, erweitert werden.*

Beweis Da $Rang(A) = m$ ist und daher $Rang(A_{J(\bar{x})}) = |J(\bar{x})| =: \ell \leq m$ sein muss, hat $\bar{x}$ höchstens $\ell \leq m$ positive Komponenten. Die Gesamtmatrix A besitzt maximalen Rang m, es existiert also eine Basis des $\mathbb{R}^m$ aus Spalten von A. Nach dem Basis-Austauschsatz können daher die ℓ linear unabhängigen Spalten von $A_{J(\bar{x})}$ zu einer vollen Basismatrix A_J, mit $|J| = m$ und $J \supseteq J(\bar{x})$ ergänzt werden. ∎

Da die im Satz genannte Ergänzung nicht eindeutig ist, gehören zu einer ausgearteten Basislösung mehrere *verschiedene Basen*. Dies kann im demnächst behandelten Simplexverfahren zu Problemen führen ($\rightarrow$ läuft im Kreis), da es nicht durch die Orte $\bar{x}$, sondern die zugehörigen Basen B gesteuert wird.

Der geometrische Hintergrund für die folgenden Überlegungen ist die Tatsache, dass die zulässige Menge X ein *konvexes Polyeder* ist und Basislösungen gerade den *Ecken* dieser Menge entsprechen. Diese Begriffe und Eigenschaften werden aber erst im Geometrie-Kapitel §3 genauer definiert. Eines der zentralen Ergebnisse dort besagt, dass man beim Linearen Programm nur Basislösungen untersuchen muß.

Basisdarstellung von X: Zu jeder Basislösung $\bar{x}$ von X gibt es eine Basis $B = A_J$ mit $A_J\bar{x}_J = b$, $\bar{x}_K = 0$, $J \cup K = \{1, \ldots, n\}$. Aber nicht nur dieser spezielle Punkt $\bar{x}$, sondern jeder Punkt $x \in X$ kann mit Hilfe dieser Basis dargestellt werden. Dazu wird analog zu (2.3.3) die Gesamtmatrix $A = (A_J, A_K)$ aufgeteilt und das Gleichungssystem $Ax = b$ umgeformt. Da A_J^{-1} existiert, gilt nämlich für $x \in U$

$$Ax = A_J x_J + A_K x_K = b \iff x_J = A_J^{-1}b - A_J^{-1}A_K x_K = \bar{x}_J - A_J^{-1}A_K x_K. \tag{2.3.5}$$

Dies ist die aus der Linearen Algebra bekannte Parameterdarstellung des Lösungsraums U mit den Variablen $x_K \geq 0$ als "freien" und den x_J als "abhängigen" Variablen und der speziellen Lösung $\bar{x}$. Nach Einführung von $n - m = |K|$ echten Parametern $\lambda_K \geq 0$ heißt das also

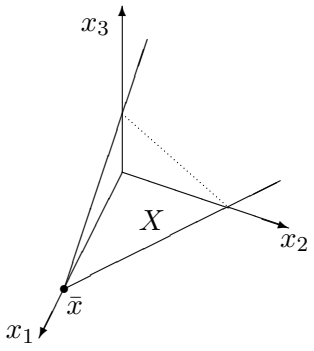
$$Ax = b,\, x \geq 0 \quad \iff \quad x = \begin{pmatrix} x_J \\ x_K \end{pmatrix} = \begin{pmatrix} \bar{x}_J \\ 0_K \end{pmatrix} + \begin{pmatrix} -A_J^{-1}A_K \\ I_{n-m} \end{pmatrix} \lambda_K = \bar{x} - W_K\lambda_K \geq 0. \tag{2.3.6}$$

Im letzten Schritt wurde die Abkürzung

$$\begin{pmatrix} A_J^{-1}A_K \\ -I_{n-m} \end{pmatrix} =: \begin{pmatrix} W_K^{(J)} \\ W_K^{(K)} \end{pmatrix} = W_K = \left(w_{i,k_j}\right) \in \mathbb{R}^{n\times(n-m)}, \quad K = \{k_1, \ldots, k_{n-m}\},$$

benutzt. Im Simplexverfahren spielt nur der Teil $W_K^{(J)}$, vgl. (2.3.2), eine Rolle und das hier gewählte Vorzeichen führt dort zu einfacheren Regeln. Die Spalten von W_K sind wegen I_{n-m} linear unabhängig und bilden eine Basis von $\ker(A)$. In einer Umgebung der Basislösung $\bar{x}$ sieht die zulässige Menge X also aus wie ein Kegel aus positiven Linearkombinationen der Vektoren $-w_j$. Denn nach (2.3.6) gilt

$$x - \bar{x} \in \{-\sum_{j\in K} w_j\lambda_j : \lambda_j \geq 0\} \quad \text{und } x \in \mathbb{R}^n_+.$$



Im Bild befindet sich, von $\bar{x}$ aus gesehen, der Bereich $X \subseteq \mathbb{R}^3$ in dem angegebenen Kegel, der allerdings an der gepunkteten Linie den Positivkegel $\mathbb{R}^3_+$ verläßt.

Mit den Spalten von W können nun spezielle, von $\bar{x}$ ausgehende Strahlen (Halbgeraden) in X beschrieben werden, bei denen genau eine K-Komponente positiv ist. Dazu wird zu festem $\ell \in K$ und $t \in \mathbb{R}_+$ der *elementare Strahl*

$$x(t) := \bar{x} - tw_\ell \quad \iff \begin{cases} x_J(t) = & \bar{x}_J - tA_J^{-1}a_\ell \\ x_k(t) = & t\delta_{k\ell}, \quad k \in K, \end{cases} \tag{2.3.7}$$

betrachtet. Da der Vektor $x(t)$ die Gestalt (2.3.6) hat, ist das System $Ax(t) \equiv b$ erfüllt $\forall t \in \mathbb{R}$. Für die Zugehörigkeit $x(t) \overset{!}{\in} X$ muß nur noch das Vorzeichen $x(t) \geq 0$ geprüft werden für Werte $t = x_\ell(t) \geq 0$. Außerdem interessiert natürlich, wie sich die Zielfunktion auf $x(t)$ ändert.

Durch Einsetzen der Basisdarstellung (2.3.5) in die Zielfunktion und Betrachtung der Vorzeichenbedingungen kann man wichtige Aussagen zur Bedeutung einer Basislösung machen. Denn mit einer Basislösung $\bar{x}$ und zugehöriger Basis A_J gilt für beliebige zulässige Punkte $x \in X$

$$\begin{aligned} c^\mathsf{T} x &= c_J^\mathsf{T} x_J + c_K^\mathsf{T} x_K \overset{(2.3.5)}{=} c_J^\mathsf{T}(\bar{x}_J - A_J^{-1} A_K x_K) + c_K^\mathsf{T} x_K \\ &= c_J^\mathsf{T} \bar{x}_J + \Big(c_K^\mathsf{T} - c_J^\mathsf{T} A_J^{-1} A_K\Big) x_K = \underbrace{c^\mathsf{T} \bar{x}}_{\text{aktuelle ZF}} + \underbrace{\gamma_K^\mathsf{T} x_K}_{\text{Änderung}} . \end{aligned} \tag{2.3.8}$$

Damit wird das Verhalten der Zielfunktion in der Nähe von $\bar{x}$ alleine durch die Nichtbasis-Variablen x_K beschrieben. Da $c^\mathsf{T}\bar{x}$ der Zielwert in der aktuellen Ecke ist, beschreibt der n-Vektor

$$\gamma^\mathsf{T} := c^\mathsf{T} - c_J^\mathsf{T} A_J^{-1} A \quad \text{mit} \quad \gamma_K^\mathsf{T} = c_K^\mathsf{T} - c_J^\mathsf{T} A_J^{-1} A_K = -c^\mathsf{T} W_K \tag{2.3.9}$$

der sogenannten *reduzierten Kosten* die Änderung der Zielfunktion bei Vergrößerung der Nichtbasis-Variablen $x_K \geq 0$. Für die Basisindizes gilt offensichtlich $\gamma_J^\mathsf{T} = c_J^\mathsf{T} - c_J^\mathsf{T} A_J^{-1} A_J = 0^\mathsf{T}$.

Satz 2.3.3 (Optimalität) *Gegeben sei eine Basis A_J mit Basislösung $\bar{x} \in X$. Wenn alle reduzierten Kosten nicht-negativ sind, $\gamma \geq 0$, dann ist $\bar{x}$ (Minimal-) Lösung von (LP3).*

Beweis Mit der Basisdarstellung (2.3.6) können **alle** Punkte $x \in X$ erreicht werden. Damit gilt aber für die Zielfunktion nach (2.3.8) bei jedem beliebigen $x \in X$, dass

$$c^\mathsf{T} x = c_J^\mathsf{T} \bar{x}_J + \gamma_K^\mathsf{T} x_K = c^\mathsf{T} \bar{x} + \underbrace{\gamma_K^\mathsf{T}}_{\geq 0} \underbrace{x_K}_{\geq 0} \geq c^\mathsf{T} \bar{x},$$

also ist $\bar{x}$ (eine) Lösung. ∎

Die Aussage bezieht sich auf eine gewählte Basis, für eine bestimmte Basislösung $\bar{x}$ ist das Kriterium aber nur hinreichend, da zu einer ausgearteten Basislösung $\bar{x}$ verschiedene Basen existieren können, die möglicherweise nicht alle das Optimalitätskriterium des Satzes erfüllen.

Wenn also negative Kosten $\gamma_\ell < 0$ existieren, kann die Zielfunktion evtl. noch verkleinert werden, indem man auf einem Strahl (2.3.7) entlangläuft. Wenn dieser allerdings als Ganzes in X liegt, kann die Zielfunktion beliebig klein werden und dann existiert keine Lösung für (LP3).

Satz 2.3.4 (Unbeschränktheit) *Gegeben sei eine Basis A_J mit Basislösung $\bar{x} \in X$. Wenn für ein $\ell \in K$ gilt $\gamma_\ell < 0$ und $w_\ell^{(J)} = A_J^{-1} a_\ell \leq 0$, dann ist (LP3) unbeschränkt.*

Beweis Zu dem genannten $\ell \in K$ mit $\gamma_\ell < 0$ wird der Strahl (2.3.7) $x(t) = \bar{x} - t w_\ell \in U := \{x : Ax = b\}$ im Lösungsraum des LGS betrachtet. Für $w_\ell^{(J)} \leq 0$ gilt sogar für den Gesamtvektor $-w_\ell \geq 0$. Also ist

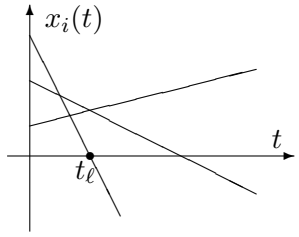
$$x(t) = \underbrace{\bar{x}}_{\geq 0} + t \underbrace{(-w_\ell)}_{\geq 0} \geq 0 \ \forall t \geq 0,$$

wegen $x(t) \in X\, \forall t \geq 0$ gibt es keine Einschränkung an t. Auf diesem Strahl fällt aber die Zielfunktion

$$c^\mathsf{T} x(t) = c_J^\mathsf{T}\bar{x} - tc^\mathsf{T} w_\ell = c_J^\mathsf{T}\bar{x} - t(c_J^\mathsf{T} w_\ell^{(J)} - c_\ell) = c_J^\mathsf{T}\bar{x} + t(c_\ell - c_J^\mathsf{T} A_J^{-1} a_\ell) = c_J^\mathsf{T}\bar{x} + t\underbrace{\gamma_\ell}_{<0} \to -\infty \ (t \to \infty)$$

beliebig weit, das Problem ist unbeschränkt. ∎

Wenn dagegen $w_\ell^{(J)} \not\leq 0$ gilt, kann man dem Strahl (2.3.7) nur ein endliches Stück weit folgen, bis man den zulässigen Bereich X verläßt. Die Zulässigkeit von (2.3.7) erfordert mit $t = x_\ell \geq 0$:



$$x_i(t) = \bar{x}_i - tw_{i\ell} \geq 0 \quad \forall i \in J.$$

In dieser Bedingung spielen nur die positiven Komponenten von $w_\ell^{(J)}$ eine Rolle, der maximal zulässige Wert für t ist daher

$$t_\ell := \min\{\frac{\bar{x}_i}{w_{i\ell}} : i \in J, w_{i\ell} > 0\} = \frac{\bar{x}_p}{w_{p\ell}} \geq 0. \tag{2.3.10}$$

Dieser Wert wurde gerade so bestimmt, dass eine Komponente $x_p(t_\ell)$ null wird, und deren Index $p \in J$ ist einer, in dem das Minimum in (2.3.10) angenommen wird. Für eine nicht ausgeartete Basislösung ist $\bar{x}_J > 0$ und daher $t_\ell > 0$. Im neuen Punkt ist nun die Komponente $x_\ell(t_\ell) = t_\ell > 0$ und mit $x_p(t_\ell) = 0$ ändert sich die Stützmenge zu $J(x(t_\ell)) = J \setminus \{p\} \cup \{\ell\}$. In diesem Punkt liegt also ein Kandidat für eine *neue Basis* vor, deren Regularität aber zu prüfen ist.

Satz 2.3.5 (Basiswechsel) *Gegeben sei die Basis $B = A_J$ mit Basislösung $\bar{x} \in X$. Sei für*

$$\ell \notin J: \quad \gamma_\ell = -c^\mathsf{T} w_\ell = c_\ell - c_J^\mathsf{T} A_J^{-1} a_\ell < 0 \ \text{ und } \ J^+(w_\ell) \neq \emptyset.$$

Mit einem Index $p \in J$, in dem das Minimum in (2.3.10) angenommen wird, bildet man die neue Menge $J' := J \setminus \{p\} \cup \{\ell\}$. Dann ist $B' = A_{J'}$ neue Basis mit Basislösung $x' = x(t_\ell)$, wobei $x'_{J'} = (B')^{-1}b \geq 0$, und neuem Zielfunktionswert $c^\mathsf{T} x' \leq c^\mathsf{T}\bar{x}$. Die Ungleichung gilt streng $c^\mathsf{T} x' < c^\mathsf{T}\bar{x}$, wenn $t_\ell > 0$ ist in (2.3.10).

Beweis Das Hauptproblem ist die Regularität der Matrix B'. Es sei s die Position von a_p in B und $a_p = Be_s$. Die neue Spalte a_ℓ werde bei B' an der Stelle s eingefügt, damit gilt also $B' = B + (a_\ell - a_p)e_s^\mathsf{T}$ und $B'e_s = a_\ell$. Die Bedingung zur Anwendung der Rang-1-Formel (2.2.2) ist erfüllt, da

$$\beta = e_s^\mathsf{T} B^{-1} a_\ell = e_p^\mathsf{T} A_J^{-1} a_\ell = w_{p\ell} > 0.$$

Denn die Zeile p von A_J^{-1} steht bei B^{-1} in Zeile s, und $w_{p\ell} > 0$ ist das Element, das den Wert t_ℓ in (2.3.10) bestimmt. Also ist B' regulär. Die zu B' gehörige Basislösung $x' = A_{J'}^{-1}b$ kann ebenfalls mit (2.2.2) bestimmt werden, es gilt mit $A_J^{-1}a_\ell = w_\ell^{(L)}$ und der Definiton von t_ℓ:

$$\begin{aligned} x'_\ell &= \ e_s^\mathsf{T}(B')^{-1}b = \tfrac{1}{\beta}\, e_s^\mathsf{T} B^{-1} b = \tfrac{1}{\beta}\bar{x}_p = t_\ell, \\ x'_i &= \ e_i^\mathsf{T} A_J^{-1} b - \tfrac{1}{\beta} w_{i\ell}\bar{x}_p = \bar{x}_i - t_\ell w_{i\ell}, \qquad i \in J. \end{aligned}$$

Insbesondere gilt $x'_p = 0$. Für die Zielfunktion im neuen Punkt x' erhält man demnach

$$c^\mathsf{T} x' = c^\mathsf{T}\bar{x} + t_\ell(c_\ell - \sum_{i\in J} c_i w_{i\ell}) = c^\mathsf{T}\bar{x} + t_\ell \underbrace{(-c^\mathsf{T} w_\ell)}_{\gamma_\ell < 0} \leq c^\mathsf{T}\bar{x}.$$

Für $t_\ell > 0$ (d.h. $\bar{x}_J > 0$) tritt hier auch eine echte Änderung $t_\ell\gamma_\ell < 0$ auf. ∎

2.4 Das revidierte Simplex-Verfahren

In der Basislösung $\bar{x}$ mit Basis A_J sind am Vektor γ der reduzierten Kosten alle diejenigen Richtungen ablesbar, in denen die Zielfunktion fällt, nämlich alle $x_K \geq 0$ mit $\gamma_K^\mathsf{T} x_K < 0$. Aus Effizienzgründen beschränkt man sich aber darauf, dass pro Schritt nur eine einzige Komponente x_ℓ, $\ell \in K$, des aktuellen Vektors $\bar{x}_K = 0$ vergrößert wird und die Zielfunktion dabei nicht wächst. Man bewegt sich also nur auf einem *elementaren Strahl* (2.3.7). Daher besteht der Ablauf, ausgehend von einer zulässigen Startbasis A_J, grob aus folgenden Schritten:

1. Berechne $\bar{x}_J$ und γ_K zu $K = \{1, \ldots, n\} \setminus J$,
2. suche $\gamma_\ell < 0$, $\ell \in K$,
3. wenn aber $\gamma_K \geq 0$, nach S. 2.3.3 _____: OPTIMUM!,
4. wenn $w_\ell^{(J)} \leq 0$, nach S. 2.3.4 _____: UNBESCHRÄNKT!
5. bestimme Minimalindex p, $w_{p\ell} > 0$, in (2.3.10),
6. Basiswechsel zu $J := J \setminus \{p\} \cup \{\ell\}$.

Die erforderlichen Berechnungen sollten möglichst effizient erfolgen. Benötigt werden dazu in jeder besuchten Basis die Größen

$$\gamma_K^\mathsf{T} = c_K^\mathsf{T} - (c_J^\mathsf{T} A_J^{-1}) A_K, \quad w_\ell^{(J)} = A_J^{-1} a_\ell, \quad \bar{x}_J = A_J^{-1} b.$$

Wenn die Berechnung von γ_K in der angegebenen Weise geklammert wird, mit $y^\mathsf{T} := c_J^\mathsf{T} A_J^{-1}$, kostet die Bestimmung der drei Lösungen

$$y^\mathsf{T} A_J = c_J^\mathsf{T}, \; A_J w_\ell^{(J)} = a_\ell, \; A_J \bar{x}_J = b,$$

bei vorhandener LR-Zerlegung $A_J = LR$ nur einen Aufwand von höchstens $6m^2$ Operationen. Außerdem kann diese LR-Zerlegung mit der Technik aus §2.2 mit einem $O(m^2)$-Aufwand zu einer Zerlegung von $A_{J'}$, $J' = J \setminus \{p\} \cup \{\ell\}$, umgebaut werden. Die Dimension $n > m$ geht nur bei $\gamma_K = c_K^\mathsf{T} - y^\mathsf{T} A_K$ in Schritt 2 ein, der Aufwand wäre hier $2m(n-m)$ Operationen, wenn alle Komponenten bestimmt würden. Man muss aber nur einen Teil der γ_j berechnen, wenn man eines der *ersten* $\gamma_\ell < 0$ akzeptiert. Das Vorgehen ergibt den

Simplex-Algorithmus

Eingabe:	Zulässige Basis A_J, $J \subseteq \{1, \ldots, n\}$	
Schritt 1	$x_J := A_J^{-1} b$, $y^\mathsf{T} := c_J^\mathsf{T} A_J^{-1}$, $K := \{1, \ldots, n\} \setminus J$,	
2	suche $\gamma_\ell < 0$ unter $\gamma_j := c_j - y^\mathsf{T} a_j$, $j \in K$.	
3	wenn $\gamma_j \geq 0 \, \forall j \in K$: _____	**STOP**, Optimum!
4	$w_\ell^{(J)} := A_J^{-1} a_\ell$, wenn $w_{i\ell} \leq 0 \, \forall i \in J$: _____	**STOP**, unbeschränkt!
5	Bestimme $p \in J$: $x_p / w_{p\ell} = \min\{x_i / w_{i\ell} : \; w_{i\ell} > 0, i \in J\} = t_\ell$	
6	$J := J \setminus \{p\} \cup \{\ell\}$, weiter mit 1	

Beispiel 2.4.1 Simplexverfahren mit $m = 3$, $n = 6$ bei (LP3) mit $c^\mathsf{T} = (-9, -6, -7, 0, 0, 0)$,

$$A = \begin{pmatrix} 3 & 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 4 & 3 & 4 & 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 20 \\ 11 \\ 40 \end{pmatrix}.$$

Das Problem ist aus einem (LP2) durch Einführung von Schlupfvariablen entstanden. Hier gehört zu $J = \{4, 5, 6\}$ eine Startbasis $A_J = I_3$ mit Basislösung $\bar{x}_J = b \geq 0$. Auftretende Simplex-Basen:

B-1 $\boxed{1.}$ $J = \{4, 5, 6\}$, $A_J = I$, $\bar{x}_J = \begin{pmatrix} 20 \\ 11 \\ 40 \end{pmatrix}$, $y^\mathsf{T} = 0^\mathsf{T}$, $\gamma_K^\mathsf{T} = (\gamma_1, \gamma_2, \gamma_3) = (-9, \underline{-6}, -7)$.

$\boxed{2{+}4.}$ wähle $\ell = 2$: $w_2^{(J)} = \begin{pmatrix} w_{42} \\ w_{52} \\ w_{62} \end{pmatrix} = Ia_2 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$,

$\boxed{5.}$ (2.3.10): $\left.\begin{array}{lll} x_4(t) = 20 - t & \geq 0 \\ x_5(t) = \underline{11 - t} & \geq 0 \\ x_6(t) = 40 - 3t & \geq 0 \end{array}\right\} \Rightarrow t_2 = 11,\ p = 5.$

B-2 $\boxed{1.}$ $J = \{2, 4, 6\}$, $K = \{1, 3, 5\}$, $A_J = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 3 & 0 & 1 \end{pmatrix}$, $A_J^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -3 & 1 \end{pmatrix}$, $\bar{x}_J = \begin{pmatrix} 11 \\ 9 \\ 7 \end{pmatrix}$,

$$y^\mathsf{T} = (c_2, c_4, c_6)A_J^{-1} = (0, -6, 0),\ \gamma_K^\mathsf{T} = (c_1, c_3, c_5) - (0, -6, 0)\begin{pmatrix} 3 & 2 & 0 \\ 1 & 1 & 1 \\ 4 & 4 & 0 \end{pmatrix} = (\underline{-3}, -1, 6).$$

$\boxed{2{+}4.}$ wähle $\ell = 1$: $w_1^{(J)} = \begin{pmatrix} w_{21} \\ w_{41} \\ w_{61} \end{pmatrix} = A_J^{-1}a_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$,

$\boxed{5.}$ (2.3.10): $\left.\begin{array}{lll} x_2(t) = 11 - t & \geq 0 \\ x_4(t) = \underline{9 - 2t} & \geq 0 \\ x_6(t) = 7 - t & \geq 0 \end{array}\right\} \Rightarrow t_1 = \frac{9}{2},\ p = 4.$

Kontrolle: insbesondere ist $x_1 = t_1$ vom Schritt vorher!

B-3 $\boxed{1.}$ $J = \{1, 2, 6\}$,$A_J = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \end{pmatrix}$, $A_J^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 \\ -1 & 3 & 0 \\ -1 & -5 & 2 \end{pmatrix}$, $\bar{x}_J = \frac{1}{2}\begin{pmatrix} 9 \\ 13 \\ 5 \end{pmatrix}$,

$$y^\mathsf{T} = (c_1, c_2, c_6)A_J^{-1} = \tfrac{1}{2}(-3, -9, 0),\ \gamma_K^\mathsf{T} = (c_3, c_4, c_5) - \tfrac{1}{2}(-3, -9, 0)\begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 4 & 0 & 0 \end{pmatrix} = (\tfrac{1}{2}, \tfrac{3}{2}, \tfrac{9}{2}).$$

$\boxed{3.}$ $\gamma_K > 0$, $\bar{x}_J > 0$: eindeutiges Minimum!

Zwei offene Fragen zum Simplex-Algorithmus müssen noch genauer behandelt werden:

— Bestimmung einer Start-Basis (Anlaufrechnung, vgl. §2.6)
— Der Algorithmus ist endlich, wenn Basen *nicht* wiederholt auftreten.

Das zentrale Ergebnis von Kapitel 3 wird der Dekompositionssatz sein, der eine *endliche Darstellung* des Polyeders X durch Ecken und Kanten garantiert. Dies sind auch die im Simplexverfahren verwendeten Größen und daher terminiert dieses in endlicher Zeit, wenn jede Basis nur einmal auftritt. Allerdings ist dies beim "Kreisen" des Simplex-Verfahrens nicht gegeben, dort werden Basen zyklisch wiederholt ohne dass sich $\bar{x}$ ändert. Dieses Problem tritt aber nur in ausgearteten Basislösungen auf, in normalen $\bar{x} \in X$ mit $|J(\bar{x})| = m$ gibt es beim Basiswechsel nach Satz 2.3.5 dagegen eine echte Abnahme der Zielfunktion, was eine Rückkehr zu $\bar{x}$ ausschließt. Ausgeartete Basen treten eher selten auf (nicht-generischer Fall), wenn $\bar{x}$ "zufälligerweise" auf mehr als $n-m$ Hyperebenen $\{x : {a^{(i)}}^{\mathsf{T}} x = b_i\}$ bzw. $\{x : x_j = 0\}$ liegt. Vor allem bei Problemen mit (kleinen) ganzzahligen Koeffizienten ist dieser Fall aber nicht auszuschließen. Das Kreisen kann durch Zusatzmaßnahmen verhindert werden (§2.7).

Gesamtaufwand des Simplex-Verfahrens Der einzelne Simplex-Schritt, der im Algorithmus formuliert wurde, ist zwar effizient durchführbar mit einem Aufwand von $O\big(m(m+n)\big)$ Operationen. Der Gesamtaufwand hängt aber von der Anzahl der untersuchten Basen ab und kann durch Änderungen bei den Auswahlentscheidungen in Schritt 2 und 5 im Einzelfall verbessert werden. Unglücklicherweise fallen aber generelle Aussagen zur Anzahl der zu untersuchenden Basen eher negativ aus.

Beispiel 2.4.2 (Klee-Minty) Zu $n \in \mathbb{N}$, $\epsilon \in (0, \frac{1}{2})$ betrachte man

$$\min\{-e_n^{\mathsf{T}} x : x \in X\},$$
$$X := \{x : 0 \le x_1 \le 1,\ \epsilon x_i \le x_{i+1} \le 1 - \epsilon x_i,\ i = 1, \ldots, n-1\}.$$

Es läßt sich zeigen, dass das Polyeder X genau 2^n Ecken besitzt, und einen Simplexpfad, der alle besucht. Dieses Problem kann auch nicht durch verbesserte Auswahlstrategien umgangen werden, auch dafür gibt es meist Gegenbeispiele mit exponentiellem Aufwand. In der Praxis arbeitet das Simplexverfahren aber sehr effizient, bei genügend allgemeiner Verteilung der Restriktionen ist beim Problem (LP1) im Mittel mit $O(\sqrt[n-1]{m} \cdot n^3)$ Schritten zu rechnen.

2.5 Tabellenform des Simplex-Verfahrens

Beim revidierten Simplexverfahren werden nur die für die Durchführung der einzelnen Schritte erforderlichen Größen berechnet. Der zugehörige Verwaltungsaufwand (Indexmenge J) ist nur gering, für Handrechnung aber irritierend. In der älteren Tabellenform des Simplexverfahrens wird immer das gesamte System umgeformt und notiert in der ursprünglichen Reihenfolge der Spalten, $H \doteq A_J^{-1} A$. Der Punkt deutet dabei die unterschiedliche Indizierung der Zeilen bei H und $A_J^{-1} A$ an. Dieses System $Hx \doteq A_J^{-1} Ax = A_J^{-1} b = \bar{x}_J$ wird außerdem ergänzt durch die aktuelle Zielfunktion $\omega = c^{\mathsf{T}} \bar{x}$, den gesamten Kostenvektor $\gamma^{\mathsf{T}} = c^{\mathsf{T}} - c_J^{\mathsf{T}} A_J^{-1} A$ und als Tableau

geschrieben in der Form

$$\left(\begin{array}{c|c} -c^{\mathsf{T}}\bar{x} & c^{\mathsf{T}} - c_J^{\mathsf{T}} A_J^{-1} A \\ \hline A_J^{-1} b & A_J^{-1} A \end{array}\right) = \begin{pmatrix} -\omega & \gamma^{\mathsf{T}} \\ \bar{x}_J & H \end{pmatrix} =: \bar{H} = \left(h_{ij}\right)_{i,j=0}^{m,n}. \tag{2.5.1}$$

Die zusätzlichen Daten werden also als nullte Zeile und Spalte des Tableaus geführt. Wegen $H_J \doteq A_J^{-1} A_J = I$ stehen in den Spalten zu Basisindizes $j \in J$ die Einheitsvektoren, dort gilt $\gamma_j = 0$ und $He_j \in \{e_1, \ldots, e_m\} \subseteq \mathbb{R}^m$. Zur Vereinfachung der folgenden Regeln wird zur Indizierung der Zeilen von H die Position i und nicht der Basisindex j_i aus $J = \{j_1, \ldots, j_m\}$ verwendet, da die entsprechende Zuordnung der Zeilen wechselt. Die Zuordnung der Komponenten aus der nullten Spalte $(h_{i0}) = \bar{x}_J$ (*Steuerspalte*) wird durch die Position der Einheitsvektoren hergestellt, es gilt $h_{i0} = x_{j_i}$ und e_i steht in Spalte j_i von H. In der nullten *Steuerzeile* stehen die reduzierten Kosten $h_{0j} = \gamma_j, j \geq 1$. Der aktuelle Zielfunktionswert wird negativ in $h_{00} = -c_J^{\mathsf{T}} \bar{x}_J$ notiert, dann gilt mit $c_0 := 0$ in der nullten Zeile die einheitliche Vorschrift $h_{0j} = c_j - \sum_i c_{j_i} h_{ij}$, $j = 0, \ldots, n$.

Die Anordnung hat den Vorteil, dass jetzt ein Basiswechsel zu dem Tableau, welches zur neuen Basis $A_{J'}$ mit $J' = J \setminus \{p\} \cup \{\ell\}$ gehört, durch Anwendung der Rang-1-Formel (2.2.2) auf das *Gesamttableau* $\bar{H}$ durchgeführt werden kann. Für $p = j_s$ entspricht das "Pivot-Element" $h_{s\ell} = w_{p\ell}$ aus (2.3.10). Die Formeln für den Basiswechsel lauten einheitlich für alle Daten:

$$\left.\begin{array}{ll} h'_{sj} & = \dfrac{h_{sj}}{h_{s\ell}}, \\ h'_{ij} & = h_{ij} - h_{i\ell} h'_{sj},\ i \in \{0, \ldots, m\} \setminus \{s\} \end{array}\right\}\ j = 0, \ldots, n. \tag{2.5.2}$$

In der zweiten Zeile, für $i \neq s$, wurde insbesondere zur Vereinfachung berücksichtigt, dass bei der Korrektur die auftretenden Quotienten $h_{sj}/h_{s\ell} = h'_{sj}$ schon in Zeile s berechnet wurden.

Satz 2.5.1 *Es sei $\bar{H}$ das Simplex-Tableau (2.5.1) zur zulässigen Basis A_J. Dann wird der Übergang zum Tableau $\bar{H}'$, das zur Basis A'_J mit $J' = J \setminus \{j_s\} \cup \{\ell\}$, $h_{s\ell} \neq 0$, gehört, durch (2.5.2) hergestellt.*

Beweis Das Tableau zur Basis $B \doteq A_J$ ist die Matrix $H \doteq B^{-1}A$. Durch Austausch in Spalte s wechselt die Basis zu $A_{J'} \doteq B' = B + ue_s^{\mathsf{T}}$ mit $p = j_s$ und $u = a_\ell - a_p$ und $B^{-1}u = h_\ell - e_s$. Zur Berechnung des neuen Tableaus dient Satz 2.2.1, dabei ist $\beta = 1 + e_s^{\mathsf{T}} B^{-1} u = 1 + e_s^{\mathsf{T}}(h_\ell - e_s) = h_{s\ell}$ und es gilt

$$H' = (B')^{-1}A = B^{-1}A - \frac{1}{\beta} B^{-1} u \underbrace{e_s^{\mathsf{T}} B^{-1} A}_{} = H - \frac{1}{\beta}(h_\ell - e_s) h^{(s)\mathsf{T}}$$

Zeilenweise bedeutet dies wie in (2.2.2)

$$e_i^{\mathsf{T}} H' = \begin{cases} \frac{1}{h_{s\ell}} e_s^{\mathsf{T}} H, & i = s, \\ e_i^{\mathsf{T}} H - h_{i\ell}\left(\frac{1}{h_{s\ell}} e_s^{\mathsf{T}} H\right), & i \neq s. \end{cases}$$

Insbesondere gilt die Formel sinngemäß auch für

$$\bar{x}_{J'} \doteq h'_0 = (B')^{-1} b = B^{-1} b - \frac{1}{\beta} B^{-1} u \underbrace{e_s^{\mathsf{T}} B^{-1} b}_{} = h_0 - \frac{1}{\beta}(h_\ell - e_s) h_{s0} = h_0 - (h_\ell - e_s) h'_{s0}.$$

Beim Kostenvektor (Steuerzeile) ist zu beachten, dass die Zeilennummern $\in \{1, \ldots, m\}$ von H die Indexposition in der Liste $J = \{j_1, \ldots, j_m\}$ angeben. Daher sei $\tilde{c}^{\mathsf{T}} = (\tilde{c}_1, \ldots, \tilde{c}_m) \doteq (c_{j_1}, \ldots, c_{j_m}) = c_J^{\mathsf{T}}$. Damit

ist dann ${h^{(0)}}^\mathsf{T} = \gamma^\mathsf{T} = c^\mathsf{T} - c_J^\mathsf{T} A_J^{-1} A = c^\mathsf{T} - \tilde{c}^\mathsf{T} H$. Beim Basiswechsel ändert sich $\tilde{c'}^\mathsf{T} = \tilde{c}^\mathsf{T} + (c_\ell - c_p) e_s^\mathsf{T}$ und führt zum neuen Kostenvektor

$$\begin{aligned}
\gamma'^\mathsf{T} &= c^\mathsf{T} - \tilde{c'}^\mathsf{T} H' = c^\mathsf{T} - (\tilde{c}^\mathsf{T} + (c_\ell - c_p) e_s^\mathsf{T}) H' \\
&= c^\mathsf{T} - \tilde{c}^\mathsf{T}\big(H - \frac{1}{\beta}(h_\ell - e_s){h^{(s)}}^\mathsf{T}\big) - (c_\ell - c_p) e_s^\mathsf{T} H' \\
&= \underbrace{c^\mathsf{T} - \tilde{c}^\mathsf{T} H}_{\gamma^\mathsf{T}} + \tilde{c}^\mathsf{T}(h_\ell - e_s)\frac{1}{\beta}{h^{(s)}}^\mathsf{T} - (c_\ell - c_p)\frac{1}{\beta}{h^{(s)}}^\mathsf{T} \\
&= \gamma^\mathsf{T} - \underbrace{\big(c_\ell - c_p + c_p - \tilde{c}^\mathsf{T} h_\ell\big)}_{\gamma_\ell}\frac{1}{\beta}{h^{(s)}}^\mathsf{T} = \Big(h_{0j} - h_{0\ell} h'_{sj}\Big)_{j=1}^{s}.
\end{aligned}$$

Wegen $\tilde{c}_s = c_{j_s} = c_p$ ist der geklammerte Ausdruck gerade $\gamma_\ell = h_{0\ell}$. Der Wert $-c_{J'}^\mathsf{T} x_{J'} = -\tilde{c'}^\mathsf{T} h'_0$ ist ein Spezialfall davon. ■

Damit läßt sich das Tableau-Verfahren angeben (Schritte wie in §2.4). Die Formulierung nimmt dabei keinerlei Bezug auf die Bedeutung der Zeilenindizes.

Simplex-Tableau-Verfahren

Eingabe:	Zulässiges Tableau $\bar{H}$
2	suche $h_{0\ell} < 0$, $1 \le \ell \le n$,
3	wenn $h_{0j} \ge 0 \,\forall 1 \le j \le n$: ——— **STOP**, Optimum!
4	wenn $h_{i\ell} \le 0 \,\forall 1 \le i \le m$: ——— **STOP**, unbeschränkt!
5	Bestimme s: $h_{s0}/h_{s\ell} = \min\{h_{i0}/h_{i\ell} :\ h_{i\ell} > 0, 1 \le i \le m\}$
6	Basiswechsel nach (2.5.2), weiter mit 2

Beispiel 2.5.2 Mit dem Ablauf aus Beispiel 2.4.1.1 bekommt man beim Tableauverfahren folgende Tabellen. In den Steuer-Zeilen und -Spalten ist jeweils das ausgewählte Element $h_{0\ell} = \gamma_\ell$ bzw. $h_{0s} = \bar{x}_p$, $p = j_s$, unterstrichen, außerdem wurde das Pivotelement für den Basiswechsel eingerahmt. Unter den Tabellen ist die Position der Basisindizes angegeben. Das erste Tableau ist zulässig, das dritte Tableau optimal, da keine negativen Kosten mehr auftreten.

0	−9	−6	−7	0	0	0
20	3	1	2	1	0	0
11	1	1	1	0	1	0
40	4	3	4	0	0	1
J:				j_1	j_2	j_3

→

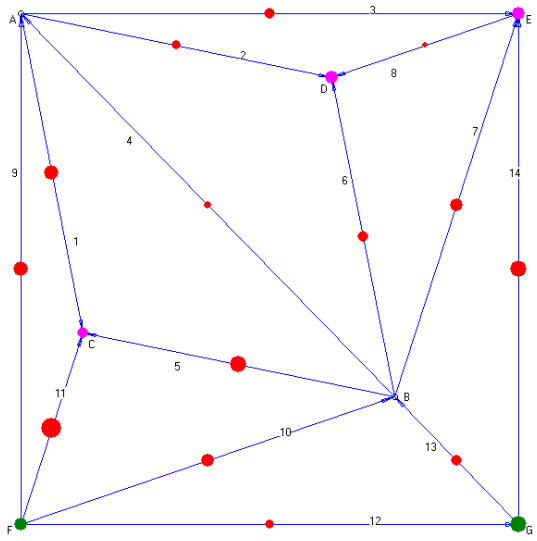
66	−3	0	−1	0	6	0
9	2	0	1	1	−1	0
11	1	1	1	0	1	0
7	1	0	1	0	−3	1
		j_2		j_1		j_3

→

$\frac{159}{2}$	0	0	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{9}{2}$	0
$\frac{9}{2}$	1	0	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0
$\frac{13}{2}$	0	1	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$	0
$\frac{5}{2}$	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{5}{2}$	1
	j_1	j_2				j_3

Das Tabellenverfahren hat (für Handrechnung) den vordergründigen Vorteil, dass der Basiswechsel mit einer einheitlichen Vorschrift für alle Daten des Linearen Programms durchgeführt werden kann. Für große Probleme ist aber ein wesentlicher Nachteil, dass immer wieder die ganze Matrix umgeformt (und damit zerstört) wird und sich die Pivotwahl nicht nach der Größe von $h_{s\ell}$ richtet. Insbesondere können kleine Pivotwerte $h_{s\ell} \cong 0$ zu großen Rundungsfehlern führen und die Fehler der Schritte summieren sich in H. Außerdem ist der Aufwand für einen Schritt immer $(2m+1)(n+1)$ Operationen.

Beispiel 2.5.3 (Rechner-Demo) In dem gezeigten Transportnetz soll ein Produkt von den Produzenten F und G zu den Abnehmern C,D,E geliefert werden, die Knoten A und B sind nur Umschlagplätze mit Bedarf 0. Transporte verlaufen längs der numerierten Kanten j in der angezeigten Richtung (Menge $x_j \geq 0$). Das zugehörige (LP3) ist in der folgenden Tabelle beschrieben, die Transportkosten der Kanten in der nullten Zeile, der Bedarf in den Knoten in der nullten Spalte. Die Kosten sollen minimiert werden. Die Restriktionen sind Bilanzgleichungen in den einzelnen Knoten, die Differenz aller eingehenden und ausgehenden Mengen entspricht dem Bedarf des Knotens. Die Zeile zu Knoten G fehlt, da sie redundant ist (Bedarf=-15), die Summe aller Zeilen der Gesamtmatrix ist null.



		53	18	29	8	60	28	37	5	44	38	98	14	23	59
$A:$	0	-1	-1	-1	1					1					
$B:$	0				-1	-1	-1	-1			1			1	
$C:$	6	1				1						1			
$D:$	10		1				1		1						
$E:$	8			1				1	-1						1
$F:$	-9									-1	-1	-1	-1		

2.6 Anlaufrechnung

Das Simplexverfahren setzt die Kenntnis einer zulässigen Startbasis voraus. Eine Startbasis konstruiert man durch Betrachtung von erweiterten Hilfsproblemen, welche die gleichen Restriktionen, aber eine andere Zielfunktion verwenden.

Zwei-Phasen-Methode

Diese basiert auf der Beobachtung, dass man beim Übergang von einem Problem (LP2) mit $b \leq 0$ zur Form (LP3) durch Einführung von Schlupfvariablen $Ax - y = b$ direkt eine Startbasis mit zulässiger Basislösung $\bar{x} = 0$, $\bar{y} = -b \geq 0$ angeben kann (vgl. Beispiel 2.4.1). Diese Kenntnis nutzt man beim Problem (LP3)

$$\min\{c^\mathsf{T} x : Ax = b,\, x \geq 0\}, \quad b \geq 0 \text{ (oBdA)},$$

und führt dort künstliche Schlupfvariable ein. Da b die rechte Seite eines Gleichungssystems ist, ist die Vorzeichenbedingung an die b_i keine Einschränkung. Zu (LP3) wird demnach mit

$\mathbb{1} = (1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^m$ das Hilfsproblem (Phase I)

$$\min \mathbb{1}^\mathsf{T} y : \; Ax + y = b, \; x \geq 0, \; y \geq 0, \tag{2.6.1}$$

mit der Matrix $D := (A, I_m) \in \mathbb{R}^{m \times (n+m)}$ betrachtet. Die Variablen können zu einem Vektor $z^\mathsf{T} = (x^\mathsf{T}, y^\mathsf{T})$ zusammengefaßt werden. Mit $J = \{n+1, \ldots, n+m\}$ ist $D_J = I_m$ eine Basis und die Basislösung $\bar{z}_J = \bar{y} = b \geq 0$ zulässig. Die neue Zielfunktion $\mathbb{1}^\mathsf{T} y = \sum_{i=1}^m y_i \geq 0$ ist eine *Straffunktion*, sie bestraft die künstlichen Schlupfvariablen und ist nach unten durch null beschränkt, das Hilfsproblem also lösbar. Mit der Lösung $\hat{z}^\mathsf{T} = (\hat{x}^\mathsf{T}, \hat{y}^\mathsf{T})$, die das Verfahren mit der Indexmenge $J \subseteq \{1, \ldots, n+m\}$ bestimmt, gilt die

Fallunterscheidung:

a) $\hat{y} \neq 0$: Das Ausgangsproblem (LP3) ist inkonsistent.

b) $\hat{y} = 0$: $\hat{x}$ ist zulässig bei (LP3), dabei

b1) $J \subseteq \{1, \ldots, n\}$: A_J bildet eine zulässige Basis für (LP3).

b2) $J \not\subseteq \{1, \ldots, n\}$: $P := J \cap \{n+1, \ldots, n+m\} \neq \emptyset$, die Lösung $\hat{z}$ ist ausgeartet. Für $p = j_s \in P$ ist $\hat{z}_p = \hat{y}_{p-n} = h_{s0} = 0$ und ein Austauschschritt mit einem beliebigen Pivot $h_{s\ell} = w_{p\ell} \neq 0$, $\ell \in \{1, \ldots, n\} \setminus J$ ändert wegen $t_\ell = 0$ nicht die Basislösung $\hat{z}$, verkleinert aber P. Wenn bei $P \neq \emptyset$ kein Austausch mehr möglich ist, gilt also $h_{sj} = w_{pj} = 0, j = 1, \ldots, n$ und die Matrix $D_J^{-1} A$ hat eine Nullzeile, A also einen Rangdefekt. Dann kann Zeile $p - n$ (zur Schlupfvariable z_p) aus A entfernt werden.

Im Fall b) kann die Rechnung mit dem Simplex-Verfahren aus §2.4 fortgesetzt werden, für das Tabellenverfahren aus §2.5 ist dazu die Steuerzeile aus c neu zu berechnen. Die Neuberechnung im Tabellenverfahren läßt sich umgehen, indem man zusätzlich zur der Steuerzeile $h^{(0)^\mathsf{T}} = (-\mathbb{1}^\mathsf{T} A, 0^\mathsf{T})$ für das Hilfsproblem (2.6.1) die zusätzliche Zeile $h^{(-1)^\mathsf{T}} = (c^\mathsf{T}, 0^\mathsf{T})$ mitführt und umformt. Nach Beendigung von Phase I ersetzt man dann $h^{(0)^\mathsf{T}}$ durch $h^{(-1)^\mathsf{T}}$.

Wenn das Ausgangsproblem (LP3) selbst schon Schlupfvariable enthält in einigen Gleichungen, muß an dieser Stelle evtl. nicht noch eine weitere eingeführt werden. →

Groß-M-Methode

Das Umschalten von Phase I auf Phase II (Originalproblem) erspart man sich, wenn man in (2.6.1) die gemischte Zielfunktion

$$c^\mathsf{T} x + M \mathbb{1}^\mathsf{T} y = (c^\mathsf{T}, M \mathbb{1}^\mathsf{T}) z$$

mit einer "genügend großen" Konstanten M betrachtet. Diese muß die künstlichen Variablen y so stark bestrafen, dass sie im Optimum nicht mehr auftreten. Allerdings ist eine geeignete Wahl von M nicht einfach zu treffen, insbesondere, wenn (LP3) inkonsistent ist.

Wenn allerdings ursprünglich das Problem (LP2) mit $b \not\leq 0$ vorliegt, hat die Methode den Vorteil, dass nur eine Zusatzvariable benötigt wird. Dazu sei $b_q = \max\{b_i : 1 \leq i \leq m\} > 0$. Im erweiterten System $Ax - y = b$ subtrahiert man nun jede Zeile von der Zeile q, ihre rechte

Seite $b_q - b_i$ wird dadurch nichtnegativ. Die Zeile q selbst bleibt unverändert, bekommt aber eine zusätzliche Variable $y_{m+1} \geq 0$. Damit ergibt sich das Problem

$$\begin{array}{llll} \min \quad c^\mathsf{T} x & & +My_{m+1} & \\ \sum_{j=1}^{n}(a_{qj}-a_{ij})x_j \quad -y_q \quad +y_i & & & = b_q - b_i \geq 0,\ i \neq q, \\ \sum_{j=1}^{n} a_{qj}x_j \quad -y_q & & +y_{m+1} & = b_q > 0, \\ x_j \geq 0, & & y_i, y_q, y_{m+1} & \geq 0 \end{array} \tag{2.6.2}$$

Die Matrix mit den Spalten zu den Indizes $J = \{n+1, \ldots, n+m+1\} \backslash \{n+q\}$ bildet eine zulässige Basis aus Einheitsvektoren mit Basislösung $\bar{x} = 0$, $\bar{y}_q = 0$, $\bar{y}_i = b_q - b_i \geq 0\,(i \neq q)$, $\bar{y}_{m+1} = b_q > 0$. Wenn dann im Optimum $(\hat{x}^\mathsf{T}, \hat{y}^\mathsf{T})$ die Zusatzvariable verschwindet, $\hat{y}_{m+1} = 0$, hat man natürlich auch eine Lösung des Ausgangsproblems gefunden. Im umgekehrten Fall ist allerdings nicht klar, ob nur M zu klein gewählt wurde, oder ob das Ausgangsproblem inkonsistent ist. Die Zwei-Phasen-Methode bietet hier eine verläßlichere Entscheidung.

Beispiel 2.6.1 Beim folgenden Problem (LP2), einschließlich Schlupfvariablen,

$$\begin{array}{rrrrrrr} \min & 2x_1 & -3x_2 & & & & \\ & -2x_1 & +3x_2 & -y_1 & & & = 5 \\ & -x_1 & +2x_2 & & -y_2 & & = 2 \\ & -x_1 & -2x_2 & & & -y_3 & = -6 \end{array}$$

tritt das größte Element von b in der ersten Zeile auf. Subtraktion der übrigen Zeilen von der ersten und Einführung der Zusatzvariablen y_4 führt auf das folgende zulässige Tableau $\bar{H}$. Die Kosten für die Steuerzeile sind $\gamma^\mathsf{T} = (c^\mathsf{T}, 0^\mathsf{T}, M) - Me_q^\mathsf{T} H$, es wird also das M-fache der q-ten Gesamtzeile vom Zielvektor subtrahiert. Das M in der letzten Spalte hebt sich dabei auf.

$-5M$	$2+2M$	$\underline{-3-3M}$	M	0	0	0
$\underline{5}$	-2	$\boxed{3}$	-1	0	0	1
3	-1	1	-1	1	0	0
11	-1	5	-1	0	1	0

$\rightarrow$

5	0	0	-1	0	0	$M+1$
5/3	$-2/3$	1	$-1/3$	0	0	1/3
4/3	$-1/3$	0	$-2/3$	1	0	$-1/3$
8/3	7/3	0	2/3	0	1	$-5/3$

Der Wert von $M \geq 0$ wurde nicht festgelegt, er war hier unwichtig. Nach einem Schritt ist die Zusatzvariable eliminiert und das Verfahren läßt sich mit der verkleinerten Tabelle fortsetzen.

2.7 Ausgeartete Ecken und praktische Aspekte

Die Steuerung beim Simplexverfahren erfolgt allein über die (Indexmenge der) Basen. Da zu einer ausgearteten Basislösung verschiedene Basen gehören, kann es vorkommen, dass das Verfahren zwar die Basis wechselt, aber in der gleichen Basislösung verharrt. Dann besteht auch die Gefahr, dass das Verfahren (bei unveränderter Pivotwahl) zu einer früheren Basis zurückkehrt und dann in dieser Schleife gefangen bleibt ("Kreisen" beim Simplexverfahren). Dieses Problem kann insbesondere bei Restriktionen mit kleinen ganzzahligen Koeffizienten wie im Beispiel 2.5.2 auftreten. Im Verfahren sind ausgeartete Ecken daran zu erkennen, dass das Minimum in Schritt 5

bzw. (2.3.10), das die maximal mögliche Schrittweite

$$t_\ell = \min\{\frac{\bar{x}_i}{w_{i\ell}} : i \in J, w_{i\ell} > 0\}$$

bestimmt, gleichzeitig in mehreren Indizes $p_1, p_2, \ldots$ angenommen wird. Dann gilt also $x_{p_1}(t_\ell) = x_{p_2}(t_\ell) = \ldots = 0$ und $x(t_\ell)$ ist wegen $|J(x(t_\ell))| < m$ also ausgeartet. Eine einfache Abhilfe gegen das Kreisen besteht darin, dass man die Auswahl unter diesen Indizes durch Zusatzregeln wieder eindeutig macht. In der Literatur gibt es dazu unterschiedliche Strategien.

Die folgenden *kleinste Index*-Regeln wählen jeweils den in Frage kommenden kleinsten Original-Index (Komponentenindex im $\mathbb{R}^n$) und verhindert dadurch ein Kreisen. Die Schritte 2 und 5 des Simplexverfahrens aus §2.4 sind dazu so zu präzisieren:

2	bestimme $\ell \in K$: $\quad \ell = \min\{j \in K : \gamma_j < 0\}$
5	bestimme $p \in J$: $\quad p = \min\{i \in J : \bar{x}_i/w_{i\ell} = t_\ell\}$

(2.7.1)

Die Durchführung dieser Regel erfordert beim Tabellenverfahren und auch beim revidierten Verfahren (abhängig von der Indexverwaltung dort) einen geringen Organisationsaufwand (Index-Sortierung), da die zugehörigen Daten im Verfahren oft den Platz wechseln.

Das Simplexverfahren basiert darauf, dass an mehreren Stellen eine Auswahl anhand des Vorzeichens berechneter Daten, etwa der Kosten γ_K getroffen wird. Leider treten aber bei der Durchführung im Rechner Rundungsfehler auf und daher kann statt exakter Kosten $\gamma_j = 0$ ein berechneter Wert $\tilde{\gamma}_j < 0$, $\tilde{\gamma}_j \cong 0$ auftreten. In der Praxis müssen daher die Entscheidungen in (2.7.1) durch eine sorgfältig gewählte Toleranz ϵ ($\cong$ Rechengenauigkeit 10^{-15}) modifiziert werden: $\min\{j \in K : \gamma_j < -\epsilon\}$. Analog ist bei der Bestimmung von p vorzugehen, es ist der minimale Index mit $\bar{x}_i/w_{i\ell} \leq t_\ell + \epsilon$ zu verwenden.

Bei sehr kritischen Anwendungen kann man versuchen, Rundungsfehler ganz zu vermeiden. Ein Gleichungssystem mit rationalen Koeffizienten kann durch Erweiterung ganzzahlig gemacht werden und die Gauß-Elimination kann dann divisionsfrei ganzzahlig durchgeführt werden. Die dann auftretenden Koeffizienten können allerdings eine erhebliche Größenordnung annehmen.

Damit ist die Standardmethode zur Lösung von Linearen Programmen behandelt. Im Folgenden muß aber die Arbeitsgrundlage des Verfahrens, der Dekompositionssatz für Polyeder, noch erarbeitet werden. Außerdem werden weitere Eigenschaften von Ungleichungssystemen behandelt, etwa Lösbarkeits-Kriterien, die auf eine schlagkräftige Theorie über duale Programme führt. Damit werden strategische Diskussionen zu gestellten Optimierungsaufgaben möglich wie die, durch gezielte Änderungen bei einem gegebenen Problem eine zusätzliche Verkleinerung des Optimalwerts zu bewirken. Mit einem dualen Simplexverfahren lassen sich solche Änderungen auch effizient umsetzen.

3 Konvexe Geometrie

Mit dem Simplex-Verfahren kann für jedes einzelne Programm (LP) eine Lösung berechnet werden oder es wird die Unlösbarkeit festgestellt. Die theoretische Grundlage für diese Behauptung ist aber noch offen, die geometrische Struktur der zulässigen Menge X muss geklärt werden, denn auf ihrem Rand liegen die Maxima der linearen Zielfunktion. Die zentrale Aussage für Polyeder wie X lautet, dass tatsächlich nur endlich viele Punkte bzw. Richtungen von X zu prüfen sind.

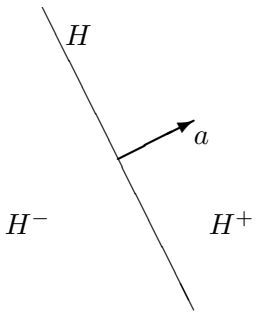
3.1 Spezielle Teilmengen

Die zulässigen Bereiche von (LP$*$) lassen sich als Durchschnitte einfacher Gebilde darstellen. Jeder $(n-1)$-dimensionale affine Unterraum $H \subseteq \mathbb{R}^n$ ist eine *Hyperebene*. Sie kann durch eine einzelne lineare Gleichung charakterisiert werden

$$H = \{x : a^\mathsf{T}(x-y) = 0\} = \{x : a^\mathsf{T}x = \alpha\}, \quad a \neq 0,\ y \in H,\ \alpha = a^\mathsf{T}y, \tag{3.1.1}$$

wobei a der (bis auf Skalierung eindeutige) *Normalenvektor* von H ist und $y \in H$ beliebig. Kompaktschreibweise $H = H(a,y) = H(a,\alpha)$. Modifikationen der Darstellung $H(a,\alpha)$ führen auf die *offenen Halbräume*

$$\begin{aligned} H^+(a,\alpha) &:= \{x : a^\mathsf{T}x > \alpha\}, \\ H^-(a,\alpha) &:= \{x : a^\mathsf{T}x < \alpha\}. \end{aligned} \tag{3.1.2}$$



Die Zerlegung $\mathbb{R}^n = H^- \cup H \cup H^+$ ist damit disjunkt. Die entsprechenden *abgeschlossenen Halbräume* sind $H^\oplus := H^+ \cup H$, $H^\ominus := H^- \cup H$. Jeder r-dimensionale affine Unterraum, $r < n$, ist Durchschnitt von $n-r$ Hyperebenen.

Zu einer beliebigen Menge $M \subseteq \mathbb{R}^n$, $M \neq \emptyset$, wird die *affine Hülle* $\mathrm{aff}(M)$ definiert als kleinster affiner Unterraum $U \subseteq \mathbb{R}^n$ mit $M \subseteq U$, also

$$\mathrm{aff}(M) = \bigcap_{U \supseteq M} U \qquad (U \subseteq \mathbb{R}^n \text{ affiner Unterraum}) \tag{3.1.3}$$

$$= \{\sum_{i=1}^{k} \lambda_i x^{(i)} : x^{(i)} \in M,\ \lambda_i \in \mathbb{R},\ \sum_{i=1}^{k} \lambda_i = 1,\ k \in \mathbb{N}\}. \tag{3.1.4}$$

Außerdem wird die (affine) *Dimension* $\dim M = \dim \mathrm{aff}(M)$ gesetzt. Umgekehrt ist der größte, bei jeder Verschiebung, in M "passende" (lineare) Unterraum der *Linealraum* $L(M)$ von M:

$$x + L(M) \subseteq M \ \forall x \in M. \tag{3.1.5}$$

Für $0 \in M$ ist offensichtlich $L(M) \subseteq M$, für beschränktes M ist $L(M) = \{0\}$ trivial.

Beispiel 3.1.1 Für eine Hyperebene $H = H(a,\alpha) \subseteq \mathbb{R}^n$, $a \neq 0$, ist $\dim H = n-1$ und für $\alpha \neq 0$, ist $\mathrm{aff}(H \cup \{0\}) = \mathbb{R}^n$ und $L(H) = H(a,0)$.

Die beiden Darstellungen (3.1.3,3.1.4) können als Charakterisierungen der affinen Hülle von "außen" bzw. "innen" gesehen werden, wobei die zweite affine Kombinationen von Vektoren verwendet. Da unterschiedliche Arten von Linearkombinationen auch im folgenden auftreten, werden sie gemeinsam eingeführt.

Definition 3.1.2 *Zu Vektoren* $x^{(1)}, \ldots, x^{(k)}$ *heißt die Linearkombination* $z := \sum_{i=1}^{k} \lambda_i x^{(i)}$ *mit* $\lambda_i \in \mathbb{R}$ *eine*

- positive Kombination *für* $\lambda_i > 0,\ i = 1, \ldots, k,$
- konische Kombination *für* $\lambda_i \geq 0,\ i = 1, \ldots, k,$
- affine Kombination *für* $\sum_{i=1}^{k} \lambda_i = 1,$
- konvexe Kombination *für* $\sum_{i=1}^{k} \lambda_i = 1, \quad \lambda_i \geq 0,\ i = 1, \ldots, k.$

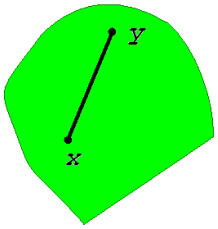
Die $k+1$ Punkte $x^{(0)}, \ldots, x^{(k)} \in \mathbb{R}^n$ heißen *affin linear unabhängig* bzw. in *allgemeiner Lage*, wenn die k Differenzen $x^{(1)} - x^{(0)}, \ldots, x^{(k)} - x^{(0)}$ linear unabhängig sind. Andernfalls sind $x^{(0)}, \ldots, x^{(k)}$ affin linear abhängig, was äquivalent zur Existenz eines nichttrivialen Tupels $(\lambda_0, \ldots, \lambda_k) \neq 0$ ist mit

$$\sum_{i=0}^{k} \lambda_i = 0, \quad \sum_{i=0}^{k} \lambda_i x^{(i)} = 0. \tag{3.1.6}$$

3.2 Konvexe Mengen

Definition 3.2.1 *Eine Menge* $M \subseteq \mathbb{R}^n$ *heißt* konvex, *wenn*

$$[x, y] := \{\lambda x + (1-\lambda) y : \ 0 \leq \lambda \leq 1\} \subseteq M \quad \forall x, y \in M.$$



Zu jedem Paar von Punkten $x, y \in M$ liegt hier die *ganze* Verbindungsstrecke $[x, y]$ in M. Die "offene" Strecke wird mit $(x, y) = \{\lambda x + (1-\lambda) y : \ 0 < \lambda < 1\}$ bezeichnet (enthält Endpunkte nicht für $x \neq y$). Das folgende Beispiel c) zeigt, dass Konvexität für uns eine zentrale Bedeutung besitzt.

Beispiel 3.2.2

a) Affine Unterräume $U \subseteq \mathbb{R}^n$ sind konvex, da mit $x, y \in U$ sogar $[x, y] \subseteq \mathrm{aff}(x, y) \subseteq U$ gilt.

b) Der Durchschnitt $\bigcap_{i \in I} M_i$ konvexer Mengen $M_i \subseteq \mathbb{R}^n$, $i \in I$, ist konvex.

c) Halbräume $H^{\pm}$, $H^{\ominus}, H^{\oplus}$ sind konvex. Die Menge

$$X := \{x \in \mathbb{R}^n : \ \sum_{j=1}^{n} a_{ij} x_j \geq b_i,\ i \in I\} = \bigcap_{i \in I} H^{\oplus}(a^{(i)}, b_i)$$

der Lösungen eines linearen Ungleichungssystems $Ax \geq b$ ist als Durchschnitt der Halbräume $H^{\oplus}(a^{(i)}, b_i)$ konvex.

d) Der *Einheitssimplex* $\Delta_n := \{x \in \mathbb{R}^n : \mathbb{1}^\mathsf{T} x = 1,\, x \geq 0\}$ ist ebenso konvex wie $\Delta_n' := \{x \in \mathbb{R}^n : \mathbb{1}^\mathsf{T} x \leq 1,\, x \geq 0\}$.

e) Streckung und Addition erhalten die Konvexität. Mit $\lambda \in \mathbb{R}$ und konvexen Mengen $M, N \subseteq \mathbb{R}^n$ sind auch folgende Mengen konvex

$$\begin{aligned} \lambda M &:= \{\lambda x :\ x \in M\}, \\ M + N &:= \{x + y :\ x \in M,\, y \in N\}. \end{aligned}$$

Definition 3.2.3 *Zu $M \subseteq \mathbb{R}^n$ ist die* konvexe Hülle *konv(M) die kleinste konvexe Menge, die M enthält.*

Offensichtlich gilt für Mengen $M \subseteq \mathbb{R}^n$: M konvex $\iff$ $M = \text{konv}(M)$. Den Zusammenhang zwischen Konvexität und Konvex-Kombinationen präzisieren die folgenden Sätze.

Satz 3.2.4 *$M \subseteq \mathbb{R}^n$ ist genau dann konvex, wenn jede konvexe Kombination von endlich vielen Punkten aus M wieder in M liegt.*

Beweis "$\Leftarrow$" Die Konvexität folgt aus dem Spezialfall $k = 2$.

"$\Rightarrow$" induktiv, die Behauptung für $k = 2$ entspricht der Definition. Nun sei M konvex und $x^{(1)}, \ldots, x^{(k+1)} \in M$, $k \geq 2$. Mit $\lambda_i \geq 0$, $\sum_{i=1}^{k+1} \lambda_i = 1$ sei $z := \sum_{i=1}^{k+1} \lambda_i x^{(i)}$. Für $\lambda_{k+1} = 1$ ist $z = x^{(k+1)} \in M$. Andererseits gilt für $\lambda_{k+1} < 1$

$$\begin{aligned} z &= \sum_{i=1}^{k} \lambda_i x^{(i)} + \lambda_{k+1} x^{(k+1)} = (1 - \lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1 - \lambda_{k+1}} x^{(i)} + \lambda_{k+1} x^{(k+1)} \\ &= (1 - \lambda_{k+1}) \underbrace{\sum_{i=1}^{k} \mu_i x^{(i)}}_{=:\tilde{z}} + \lambda_{k+1} x^{(k+1)}, \end{aligned}$$

mit $\mu_i := \lambda_i / (1 - \lambda_{k+1}) \geq 0$, $i = 1, \ldots, k$, und $\sum_{i=1}^{k} \mu_i = 1$. Damit ist $\tilde{z} \in M$ nach I.V. und auch $z \in M$ als einfache Konvexkombination von $\tilde{z}$ und $x^{(k+1)}$. ∎

Spezielle Charakterisierungen der konvexen Hülle von M sind auch:

- von außen: Durchschnitt aller konvexen Obermengen:

$$\text{konv}(M) = \bigcap_{M \subseteq N \subseteq \mathbb{R}^n} N \quad (N \text{ konvex})$$

- von innen: Menge aller konvexen Kombinationen von Punkten aus M:

$$\text{konv}(M) = \bigcup_{k \in \mathbb{N}} \{\sum_{i=1}^{k} \lambda_i x^{(i)} :\ x^{(i)} \in M,\, \lambda \in \Delta_k\}. \tag{3.2.1}$$

Der Einheitssimplex ist die konvexe Hülle aller Einheitsvektoren $\Delta_n = \text{konv}(\{e_1, \ldots, e_n\})$ und $\Delta_n' = \text{konv}(\Delta_n \cup \{0\})$. Dieses Beispiel läßt erwarten, dass in der Darstellung (3.2.1) nur eine Höchstanzahl von Summanden zu betrachten ist. Das bestätigt folgender Satz.

Satz 3.2.5 (Caratheodory) *Die Menge $M \subseteq \mathbb{R}^n$, $M \neq \emptyset$, besitze Dimension m. Dann kann jeder Punkt $z \in konv(M)$ durch höchstens $m+1$ Punkte konvex kombiniert werden, d.h., es existieren $x^{(1)}, \ldots, x^{(k)} \in M$, $k \leq m+1$, $\lambda \in \Delta_k$ so, dass $z = \sum_{i=1}^{k} \lambda_i x^{(i)}$ gilt.*

Beweis Für beliebiges $z \in \mathrm{konv}(M)$ gibt es ein $s \in \mathbb{N}$ so, dass

$$z = \sum_{i=1}^{s} \lambda_i x^{(i)}, \quad (\lambda_i) \in \Delta_s, \ x^{(i)} \in M.$$

ZZ Für $s > m+1$ können Punkte $x^{(i)}$ aus der Darstellung entfernt werden, nur der Fall $\lambda_i > 0 \forall i$ ist dabei nichttrivial. Tatsächlich sind für $s > m+1$ die Vektoren $x^{(2)} - x^{(1)}, \ldots, x^{(s)} - x^{(1)}$ linear abhängig, da ihre Anzahl größer ist als $\dim(M)$. Nach (3.1.6) existiert daher $(\alpha_1, \ldots, \alpha_s) \neq 0$ mit

$$\sum_{i=1}^{s} \alpha_i x^{(i)} = 0, \quad \sum_{i=1}^{s} \alpha_i = 0.$$

Man wählt den Index j so, dass $|\alpha_j|/\lambda_j = \max\{|\alpha_i|/\lambda_i : 1 \leq i \leq s\}$ und $\alpha_j > 0$ (oBdA). Dann ist $x^{(j)} = -\sum_{i \neq j} \alpha_i x^{(i)}/\alpha_j$ und somit auch

$$z = \sum_{\substack{i=1 \\ i \neq j}}^{s} \big(\lambda_i - \alpha_i \frac{\lambda_j}{\alpha_j}\big) x^{(i)} \quad \text{mit } \lambda_i - \alpha_i \frac{\lambda_j}{\alpha_j} = \lambda_i \Big(1 - \frac{\alpha_i \lambda_j}{\lambda_i \alpha_j}\Big) \geq 0$$

eine konische Darstellung von z mit $s-1$ Punkten. Die Darstellung ist auch konvex, da $\sum_{i \neq j} \alpha_i = -\alpha_j$ und $\sum_{i \neq j} (\lambda_i - \alpha_i \lambda_j/\alpha_j) = \sum_{i \neq j} \lambda_i + \lambda_j = 1$ ist. Die Elimination kann solange wiederholt werden, bis höchsten $m+1$ Punkte auftreten. ∎

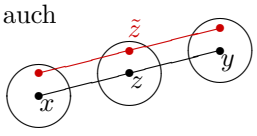
Zum Zusammenspiel von Konvexität und Topologie:

Satz 3.2.6 *Bei einer nichtleeren konvexen Menge $M \subseteq \mathbb{R}^n$ sind auch das Innere $\overset{\circ}{M}$ und der Abschluß $\bar{M}$ konvex.*

Beweis Das Innere sei nicht leer und $x, y \in \overset{\circ}{M}$. Dann liegen auch ε-Kugeln um diese in $\overset{\circ}{M}$, also $x + B_\varepsilon(0) \subseteq \overset{\circ}{M}$, $y + B_\varepsilon(0) \subseteq \overset{\circ}{M}$, mit $\varepsilon > 0$. Zu $\lambda \in [0,1]$ ist n.V. $z := \lambda x + (1-\lambda) y \in M$, es gilt auch $\lambda(x-z) + (1-\lambda)(y-z) = 0$.

ZZ $z \in \overset{\circ}{M}$. Dazu sei $\tilde{z} \in z + B_\varepsilon(0)$ beliebig, also $\|\tilde{z} - z\| \leq \varepsilon$. Dann gilt tatsächlich auch

$$\tilde{z} = \lambda \underbrace{(x + \tilde{z} - z)}_{\in B_\varepsilon(x)} + (1-\lambda) \underbrace{(y + \tilde{z} - z)}_{\in B_\varepsilon(y)} \quad \in M.$$



$\bar{M}$: zu $\bar{x}, \bar{y} \in \bar{M}$ existieren Folgen mit $x^{(i)}, y^{(i)} \in M$ und $\bar{x} = \lim_{i \to \infty} x^{(i)}$, $\bar{y} = \lim_{i \to \infty} y^{(i)}$. Zu $\lambda \in [0,1]$ ist dann $z^{(i)} = \lambda x^{(i)} + (1-\lambda) y^{(i)} \in M \, \forall i$ n.V.. Damit folgt

$$\lambda \bar{x} + (1-\lambda)\bar{y} = \lambda \lim_{i \to \infty} x^{(i)} + (1-\lambda) \lim_{i \to \infty} y^{(i)} = \lim_{i \to \infty} z^{(i)} = z \in \bar{M}$$

und somit die Konvexität von $\bar{M}$. ∎

Bei der Übertragung topologischer Eigenschaften auf die konvexe Hülle ist Vorsicht angebracht. Die Abgeschlossenheit von M überträgt sich nur bei beschränkten Mengen auf $\mathrm{konv}(M)$.

Satz 3.2.7 *Die Menge $M \subseteq \mathbb{R}^n$ sei*

$$\left.\begin{array}{l} \textit{offen} \\ \textit{beschränkt} \\ \textit{kompakt} \end{array}\right\} \quad \Rightarrow \quad \textit{konv}(M) \textit{ ist } \left\{\begin{array}{l} \textit{offen} \\ \textit{beschränkt} \\ \textit{kompakt} \end{array}\right.$$

Beweis Sei M offen. Zu $z \in \mathrm{konv}(M)$ existiert $k \in \mathbb{N}$ und $(\lambda_i) \in \Delta_k$, $x^{(i)} \in M$ mit $z := \sum_{i=1}^k \lambda_i x^{(i)}$. N.V. ist für ein $\varepsilon > 0$ auch $B_\varepsilon(x^{(i)}) \subseteq M \ \forall i = 1, \ldots, k$ und für ein $\tilde{z} \in B_\varepsilon(z)$ ist

$$\tilde{z} = \tilde{z} + \sum_{i=1}^k \lambda_i (x^{(i)} - z) = \sum_{i=1}^k \lambda_i (\underbrace{x^{(i)} + \tilde{z} - z}_{\in B_\varepsilon(x^{(i)})}),$$

eine Konvexkombination aus M heraus.

Sei M kompakt, die Beschränktheit ist dann trivial. Ist nun $z \in \mathrm{konv}(M)$ ein Häufungspunkt von $\mathrm{konv}(M)$, so existiert eine Folge $z^{(j)} \in \mathrm{konv}(M)$ mit $\lim_j z^{(j)} = z$. Nach Satz 3.2.5 hat jedes Folgenelement eine konvexe Darstellung mit *fester Anzahl* $n+1$:

$$z^{(j)} = \sum_{i=1}^{n+1} \lambda_{j,i} x^{(j,i)}, \quad (\lambda_{j,i})_{i=1}^{n+1} \in \Delta_{n+1}, \ x^{(j,i)} \in M.$$

Da auch das $n+1$-fache cartesische Produkt $M \times \ldots \times M$ und Δ_{n+1} kompakt sind, existieren konvergente Teilfolgen mit Indizes (j_k) für die Vektorfolgen

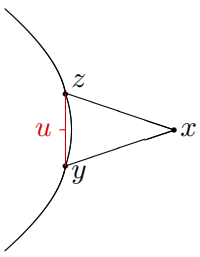
$$\left((\lambda_{j,i})_{i=1}^{n+1}\right)_{j \geq 0}, \quad \left((x^{(j,i)})_{i=1}^{n+1}\right)_{j \geq 0}.$$

Deren Limites seien $\lambda_i := \lim_{k\to\infty} \lambda_{j_k,i}$, $x^{(i)} := \lim_{k\to\infty} x^{(j_k,i)}$. Damit folgt

$$z = \lim_{j\to\infty} z^{(j)} = \lim_{k\to\infty} z^{(j_k)} = \sum_{i=1}^{n+1} \lambda_i x^{(i)} \in \mathrm{konv}(M),$$

denn es ist $(\lambda_i)_{i=1}^{n+1} \in \Delta_{n+1}$ und $x^{(i)} \in M$, da M abgeschlossen ist. ∎

Zu einem beliebigen Punkt $x \in \mathbb{R}^n$ gibt es in einer nichtleeren konvexen, abgeschlossenen Menge M einen *eindeutigen*, nächstgelegenen Punkt. Denn bei festem x ist $y \mapsto f_x(y) := \|y - x\|^2$ eine stetige Funktion und muss mit einem beliebigen $y_0 \in M$ nur auf der Kugel $B_r(x)$, $r^2 = f_x(y_0)$, bzw. der kompakten Menge $M \cap B_r(x)$ betrachtet werden. Dieses Minimum ist eindeutig aufgrund der Parallelogrammgleichung



$$\|\frac{y+z}{2}\|^2 = \frac{1}{2}\|y\|^2 + \frac{1}{2}\|z\|^2 - \frac{1}{4}\|y - z\|^2. \tag{3.2.2}$$

Bei zwei Minimalstellen mit $\|y - z\| > 0$ wäre f_x in $u := (y+z)/2 \in M$ echt kleiner: $f_x(u) < f_x(y) = f_x(z)$. Dies zeigt den

Satz 3.2.8 *Die Menge $M \subseteq \mathbb{R}^n$, $M \neq \emptyset$, sei konvex und abgeschlossen. Dann gibt es zu jedem $x \in \mathbb{R}^n$ einen eindeutigen, nächstgelegenen Punkt*

$$\hat{y} \in M: \quad \hat{y} = \textit{arg min}\{f_x(y) : y \in M\}.$$

Die Zuordnung $p_M : \mathbb{R}^n \to M$, $x \mapsto \hat{y}$ wird die Projektion *auf M genannt.*

Fixpunkte dieser Projektion $p_M(x) = x$ sind genau die Punkte $x \in M$, daher ist die Abbildung p_M auch *idempotent*, $p_M \circ p_M = p_M$. Bei einem affinen Unterraum $U \subseteq \mathbb{R}^n$ ist p_U die orthogonale Projektion auf U, mit $\hat{y} = p_M(x)$ ist

$$x = \hat{y} + (x - \hat{y}), \quad \text{wobei } (x - \hat{y})^\mathsf{T}(\hat{y} - y) = 0 \, \forall y \in U.$$

Bei einem linearen Unterraum ist auch p_U linear. Eine zur letzten Gleichung ähnliche Charakterisierung von $p_M(x)$ gilt im allgemeinen Fall.

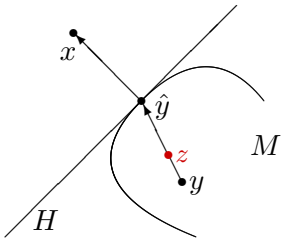
Satz 3.2.9 *Die nichtleere Menge $M \subseteq \mathbb{R}^n$ sei konvex und abgeschlossen und $\hat{y} \in M$. Dann gilt mit $x \in \mathbb{R}^n$*

$$\hat{y} = p_M(x) \quad \iff \quad (x - \hat{y})^\mathsf{T}(\hat{y} - y) \geq 0 \ \forall y \in M. \tag{3.2.3}$$

Für $x \notin M$ ist der nächstgelegene Punkt $\hat{y} = p_M(x)$ also dadurch charakterisiert, dass gilt $M \subseteq H^\ominus$, mit der Hyperebene $H = H(x - \hat{y}, \hat{y})$, die eingezeichneten Vektoren $x - \hat{y}$ und $\hat{y} - y$ aus (3.2.3) zeigen ungefähr in die gleiche Richtung.

Beweis″ $\Rightarrow$″ Mit $\hat{y} = p_M(x)$ und bel. $y \in M$ sowie $\lambda \in [0, 1]$ ist $z = \lambda y + (1 - \lambda)\hat{y} = \hat{y} + \lambda(y - \hat{y}) \in M$. Nach Voraussetzung gilt

$$\begin{aligned} f_x(\hat{y}) &\leq f_x(\hat{y} + \lambda(y - \hat{y})) = \|\hat{y} - x + \lambda(y - \hat{y})\|^2 \\ &= f_x(\hat{y}) + \underbrace{2\lambda(\hat{y} - x)^\mathsf{T}(y - \hat{y}) + \lambda^2\|y - \hat{y}\|^2}_{\geq 0}. \end{aligned}$$



Für $\lambda \to 0$ führt dies auf $(x - \hat{y})^\mathsf{T}(\hat{y} - y) \geq 0$.

″ $\Leftarrow$″ Für ein $\hat{y} \in M$ gelte $(x - \hat{y})^\mathsf{T}(\hat{y} - y) \geq 0$. Ist auch $x \in M$, führt die Wahl $y = x$ auf $-\|x - \hat{y}\|^2 \geq 0$ und zeigt $\hat{y} = x = p_M(x)$. Für $x \notin M$ ist $\|x - \hat{y}\| > 0$ und mit $y \in M$ folgt nach Cauchy-Schwarz

$$\begin{aligned} 0 &\leq (x - \hat{y})^\mathsf{T}(\hat{y} - y) = (x - \hat{y})^\mathsf{T}(\hat{y} - x + x - y) \\ &= -\|x - \hat{y}\|^2 + (x - \hat{y})^\mathsf{T}(x - y) \leq \underbrace{\|x - \hat{y}\|}_{0<} \big(\underbrace{-\|x - \hat{y}\| + \|x - y\|}_{0\leq} \big). \end{aligned}$$

Also gilt $f_x(\hat{y}) = \|x - \hat{y}\| \leq \|x - y\| = f_x(y) \ \forall y \in M$, daher ist $\hat{y} = p_M(x)$. ∎

Wie im linearen Fall sind alle Elemente von M Fixpunkte der Abbildung p_M. Diese ist auch nicht-expandierend, aber keine echte Kontraktion:

Satz 3.2.10 *Die Menge $M \subseteq \mathbb{R}^n$, $M \neq \emptyset$, sei konvex und abgeschlossen. Dann gilt für $x, y \in \mathbb{R}^n$*

$$\|p_M(x) - p_M(y)\| \leq \|x - y\|.$$

Hyperebenen der in Satz 3.2.9 auftretenden Art sind im Folgenden ein wichtiges Hilfsmittel.

Definition 3.2.11 *Sei $M \subseteq \mathbb{R}^n$ konvex, $M \neq \emptyset$. Eine Hyperebene $H = H(a, \alpha)$ mit $M \subseteq H^\ominus$, $H \cap \bar{M} \neq \emptyset$ heißt* Stützebene *für M und $a^\mathsf{T}x \leq \alpha$* zulässige Ungleichung *für M. Wenn $B := H \cap M \neq \emptyset$ ist, heißt B* Stützmenge.

In Satz 3.2.9 liegt also $p_M(x)$ für $x \notin M$ in der Stützmenge der dort zur abgeschlossenen(!) Menge M konstruierten Stützebene H. Diese *trennt* den Punkt x von der Menge M. Eine entsprechende Aussage gilt für beliebige disjunkte, konvexe Mengen.

Definition 3.2.12 *Zur Lage einer Hyperebene $H = H(a,\alpha)$ relativ zu nichtleeren Mengen $M, N \subseteq \mathbb{R}^n$ verwendet man folgende Begriffe.*

H trennt M und N,	*wenn $M \subseteq H^{\ominus},\ N \subseteq H^{\oplus}$*	*(bzw.umgekehrt)*
H trennt M und N echt,	*wenn $M \subseteq H^{\ominus},\ N \subseteq H^{+}$*	*(bzw.umgekehrt)*
H trennt M und N strikt,	*wenn $M \subseteq H^{-},\ N \subseteq H^{+}$*	*(bzw.umgekehrt)*
H trennt M und N stark,	*wenn für ein $\epsilon > 0$ gilt*	

$$a^{\mathsf{T}}x \le \alpha - \epsilon < \alpha + \epsilon \le a^{\mathsf{T}}y \quad \forall x \in M,\ y \in N.$$

Mit Satz 3.2.9 kann direkt eine Hyperbene konstruiert werden, die einen Punkt $x \notin \bar{M}$ außerhalb einer konvexen Menge von dieser strikt trennt. Etwas schwieriger wird der Nachweis, wenn x auf dem Rand von M liegt, die trennende Ebene ist dann eine Stützebene.

Satz 3.2.13 *Die nichtleere Menge $M \subseteq \mathbb{R}^n$ sei konvex.*

a) Ist M abgeschlossen und $x \notin M$, dann existiert eine Hyperebene mit $M \subseteq H^{-}(a,\alpha)$, $x \in H^{+}(a,\alpha)$, d.h.,

$$\forall y \in M: \quad a^{\mathsf{T}}y < \alpha < a^{\mathsf{T}}x.$$

b) Wenn x Randpunkt von M, $x \in \bar{M} \backslash \overset{\circ}{M}$, ist, existiert eine Hyperebene H mit $x \in H$, $M \subseteq H^{\ominus}$.

Beweis a) In Satz 3.2.9 ist $a := x - p_M(x) \neq 0$. Mit $\hat{y} = p_M(x)$ gilt für alle $y \in M$ nach (3.2.3)

$$0 \ge a^{\mathsf{T}}(y - \hat{y}) = a^{\mathsf{T}}(y - x + a) = a^{\mathsf{T}}y - a^{\mathsf{T}}x + \|a\|^2 \iff a^{\mathsf{T}}x \ge a^{\mathsf{T}}y + \|a\|^2.$$

Durch die Wahl $\alpha := a^{\mathsf{T}}x - \frac{1}{2}\|a\|^2$ geht die Hyperebene $H(a,\alpha)$ genau durch den Mittelpunkt $(x+\hat{y})/2$ und trennt x strikt von M:

$$a^{\mathsf{T}}x > a^{\mathsf{T}}x - \frac{1}{2}\|a\|^2 = \alpha \ge a^{\mathsf{T}}y + \frac{1}{2}\|a\|^2 > a^{\mathsf{T}}y \quad \forall y \in M.$$

b) Da x Randpunkt von M ist, existiert eine Folge $(x^{(j)})$ mit $x^{(j)} \notin \bar{M}$ und $x = \lim_{j\to\infty} x^{(j)}$. Zu jedem dieser $x^{(j)}$ existiert nach Teil a) eine strikt trennende Hyperebene $H(a^{(j)},\alpha_j)$. Normiert man abweichend von Teil a) durch $\|a^{(j)}\| = 1$, ist diese Folge $(a^{(j)})$ beschränkt und besitzt daher eine konvergente Teilfolge

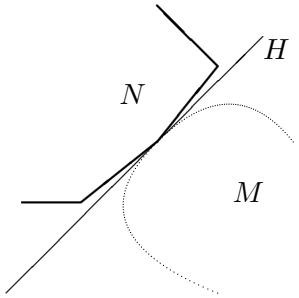
$$\lim_{k\to\infty} a^{(j_k)} = a, \quad \|a\| = 1.$$

Mit diesem a gilt $\forall y \in M$, dass

$$\begin{array}{ccc} {a^{(j_k)}}^{\mathsf{T}}y & < & {a^{(j_k)}}^{\mathsf{T}}x^{(j_k)} \\ \downarrow & & \downarrow \\ a^{\mathsf{T}}y & \le & a^{\mathsf{T}}x \end{array} \quad \text{für } k \to \infty.$$

Dies bedeutet aber gerade $y \in H^{\ominus}(a,x)$ $\forall y \in M$. ∎

Auch im Grenzfall sich berührender konvexer Mengen ist noch eine Trennung möglich.



Theorem 3.2.14 *Es seien $M, N \subseteq \mathbb{R}^n$ nichtleere, disjunkte, konvexe Mengen, $M \cap N = \emptyset$, und M offen. Dann existiert eine Hyperebene H, die M und N echt trennt, $M \subseteq H^-$, $N \subseteq H^\oplus$.*

Beweis Die Menge aller Differenzen $M - N := \{u - v : u \in M, v \in N\}$ ist konvex. Denn Punkte $x, y \in M - N$ sind Differenzen $x = u - v$, $y = w - z$ mit $u, w \in M$, $v, z \in N$. Für $\lambda \in [0, 1]$ gilt tatsächlich

$$\lambda x + (1-\lambda)y = \big(\underbrace{\lambda u + (1-\lambda)w}_{\in M}\big) - \big(\underbrace{\lambda v + (1-\lambda)z}_{\in N}\big) \in M - N.$$

Da wegen $M \cap N = \emptyset$ aber $0 \notin M - N$ ist, existiert nach Satz 3.2.13 eine Hyperebene $H(a, 0) \ni 0$, die die Null von $M - N$ trennt, also $M - N \subseteq H^\ominus(a, 0)$. Daher gilt

$$\forall w \in M,\, z \in N : y = w - z \in M - N \;\Rightarrow\; 0 \geq a^\mathsf{T} y = a^\mathsf{T} w - a^\mathsf{T} z, \text{ d.h. } a^\mathsf{T} w \leq a^\mathsf{T} z.$$

Offensichtlich ist daher $w \mapsto a^\mathsf{T} w$ beschränkt auf der offenen Menge M, das Supremum $\alpha := \sup\{a^\mathsf{T} w : w \in M\}$ existiert, wird aber nicht angenommem. Somit gilt $a^\mathsf{T} w < \alpha \leq a^\mathsf{T} z \,\forall w \in M,\, z \in N$. ∎

Bei ihrer Einführung wurde die konvexe Hülle als Durchschnitt allgemeiner konvexer Obermengen definiert. Mit den letzten Ergebnissen ist auch eine Charakterisierung nur mit Halbräumen (d.h. linearen Ungleichungen) möglich.

Satz 3.2.15 *$M \subseteq \mathbb{R}^n$ sei eine konvexe, abgeschlossene, echte Teilmenge des $\mathbb{R}^n$, $M \neq \emptyset$, $M \neq \mathbb{R}^n$. Bezeichnet $\mathcal{H}_M$ die Menge der Stützebene an M, dann gilt*

$$M = \bigcap_{H \in \mathcal{H}_M} H^\ominus.$$

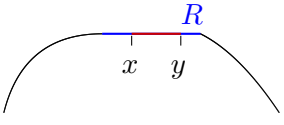
Beispiel 3.2.16 Bei der Einheitskugel $M := B_1(0)$ ist diese Aussage sofort nachvollziehbar. Für jedes $a \in \mathbb{R}^n$, $a \neq 0$, ist $H(a, \|a\|)$ eine Stützebene an M. Man sieht hier auch sofort, dass in der Darstellung $\bigcap_{H \in \mathcal{H}_M} H^\ominus$ unendlich viele Halbräume auftreten.

3.3 Randflächen und Ecken

Bekanntlich sind bei der Suche nach Extrema von Funktionen die Ränder des zulässigen Bereichs gesondert zu prüfen, insbesondere bei linearen Zielfunktionen. Auch eine Stützebene berührt eine konvexe Menge in (mindestens einem) Randpunkt. Die Definition des Randes ist bei abgeschlossenen konvexen Mengen aber auch mit rein geometrischen Begriffen möglich.

Definition 3.3.1 *Sei $R \neq \emptyset$ und beide Mengen $R \subseteq M \subseteq \mathbb{R}^n$ konvex.*
Dann heißt R Randfläche *von M, wenn*

$$\forall x, y \in M : \quad (x, y) \cap R \neq \emptyset \;\Rightarrow\; x, y \in R.$$



In der Definition tritt die offene Strecke (x, y) auf, Punkte einer Randfläche R können also nur aus Punkten von R selbst kombiniert werden. Abhängig von der Dimension einer Randfläche R verwendet man folgende Bezeichnungen:

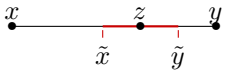
$\dim R = 0$: $R = \{y\}$ ist *Ecke* von M
$\dim R = 1$: R ist *Kante* von M
$\dim R = n-1$: R ist *Facette* von $M \subseteq \mathbb{R}^n$.

Satz 3.3.2 *Sei $M \subseteq \mathbb{R}^n$ nichtleer und konvex. Dann sind folgende Bedingungen äquivalent:*

$$\begin{aligned} &a) && z \in M \text{ ist Ecke von } M, \\ &b) && z \in (x,y),\ x,y \in M \ \Rightarrow x = y = z, \\ &c) && z = \frac{1}{2}(x+y),\ x,y \in M \ \Rightarrow x = y = z, \\ &d) && M \setminus \{z\} \text{ ist konvex.} \end{aligned}$$

Beweis Teil b) entspricht gerade der Definition des Begriffs a), gezeigt wird nur c) $\Rightarrow$ b). Dazu sei $x, y \in M$, $z = \lambda x + (1-\lambda)y$. Für $\lambda \in (0,1)$ existiert ein $\varepsilon > 0$ so, dass $0 < \lambda - \varepsilon < \lambda + \varepsilon < 1$. Damit sei

$$\left.\begin{aligned} \tilde{x} &= (\lambda+\varepsilon)x + (1-\lambda-\varepsilon)y \in M \\ \tilde{y} &= (\lambda-\varepsilon)x + (1-\lambda+\varepsilon)y \in M \end{aligned}\right\} \Rightarrow z = \frac{1}{2}(\tilde{x}+\tilde{y}).$$



Aus dieser Darstellung folgt aber n.V. $\tilde{x} = \tilde{y} = z$ und daher $0 = \tilde{x} - \tilde{y} = 2\varepsilon(x-y)$. Wegen $2\varepsilon > 0$ hat das auch $x = y = z$, also die Eckeneigenschaft, zur Folge. ■

Ecken sind die wichtigsten Teile des Randes, die Menge aller Ecken von M heißt $E(M)$.

Beispiel 3.3.3

a) Die Eckenmenge der Einheitskugel $M = B_1(0) = \{x : \|x\| \leq 1\}$ ist die Sphäre $E(M) = \{x : \|x\| = 1\}$. Dies folgt direkt aus der Parallelogrammgleichung (3.2.2) und Satz 3.3.2b. Die offene Kugel hat keine Ecken $E(\overset{\circ}{M}) = \emptyset$.

b) Auch Unterräume $U \subseteq \mathbb{R}^n$ haben keine Ecken, sind aber abgeschlossen.

c) Im folgenden treten aber in der Regel Mengen mit endlich vielen Ecken auf. Dazu gilt etwa: für $M = \text{konv}\{x^{(1)}, \ldots, x^{(m)}\}$ ist $E(M) \subseteq \{x^{(1)}, \ldots, x^{(m)}\}$.

Jede nichtleere kompakte Menge M enthält mindestens eine Ecke (Satz, denn $argmax\{\|x\| : x \in M\}$ ist Ecke). $E(M)$ enthält dann sogar so viele Punkte, dass die ganze Menge M daraus rekonstruiert werden kann (Theorem 3.3.7). Zum Beweis wird benötigt:

Satz 3.3.4 *Sei $M \neq \emptyset$, $M \subseteq \mathbb{R}^n$ konvex und kompakt und H eine Stützebene an M. Dann ist $R := H \cap M$ eine Randfläche von M und enthält eine Ecke von M.*

Beweis Als wichtigster Teil wird die Existenz der Ecke gezeigt. Da $R = H \cap M$ als nichtleerer Schnitt ebenfalls konvex und kompakt ist, besitzt R eine Ecke $z \in M \cap H$. Es sei $H = H(a, \alpha)$.

ZZ z ist Ecke der Menge M. Dazu sei $z = \frac{1}{2}(x+y)$ mit $x, y \in M$, also

$$a^\mathsf{T} x \le \alpha, \quad a^\mathsf{T} y \le \alpha, \quad a^\mathsf{T} z = \alpha \text{ (da } z \in H!).$$

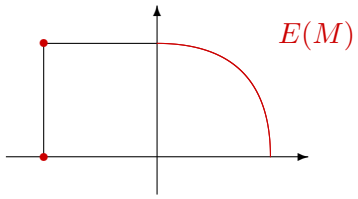
Daher gilt

$$0 = a^\mathsf{T} z - \alpha = \frac{1}{2} a^\mathsf{T}(x+y) - \alpha = \frac{1}{2}(\underbrace{a^\mathsf{T} x - \alpha}_{\le 0}) + \frac{1}{2}(\underbrace{a^\mathsf{T} y - \alpha}_{\le 0}) \le 0.$$

Also sind beide Klammern null: $x, y \in H \cap M = R \;\Rightarrow\; x = y = z$, da z Ecke von R war. ■

Aufgrund des Satzes ist jede Stützmenge auch Randfläche, aber i.a. nicht umgekehrt:

Beispiel 3.3.5 Bei der Vereinigung $M = ([-1,0] \times [0,1]) \cup (B_1(0) \cap \mathbb{R}^2_+)$ von Quadrat und Viertelkreis ist $e_2 = (0,1)^\mathsf{T}$ zwar eine Ecke, aber selbst nur Ecke einer Stützmenge.



Konvexität und Randflächen-Eigenschaft sind "monotone" bzw. transitive Eigenschaften.

Satz 3.3.6 *a) $M \subseteq \mathbb{R}^n$, $M \neq \emptyset$ sei konvex und kompakt. Dann ist jede Randfläche von M konvex und kompakt.*

b) Bei den konvexen Mengen $S \subseteq R \subseteq M \subseteq \mathbb{R}^n$, $S \neq \emptyset$, sei S Randfläche von R und R Randfläche von M. Dann ist auch S Randfläche von M und $E(R) \subseteq E(M)$.

Beweis a) Betrachte zu einer Randfläche $R \subseteq M$ den Schnitt $M \cap \mathrm{aff}(R)$.

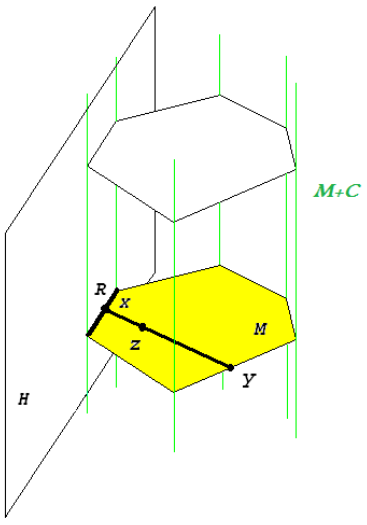
b) Sei $x, y \in M$, $(x,y) \cap S \neq \emptyset$. Dann gilt auch $(x,y) \cap R \neq \emptyset$, da $S \subseteq R$. Wegen der Randeigenschaft von R in M ist dann $x, y \in R$ und die Randeigenschaft von S in R liefert $x, y \in S$. Ecken sind der nulldimensionale Spezialfall. ■

Theorem 3.3.7 (Krein-Milman) *Sei $M \neq \emptyset$, $M \subseteq \mathbb{R}^n$ konvex und kompakt. Dann gilt*

$$M = konv(E(M)).$$

Beweis Induktion über $k = \dim M$, die Behauptung gilt für Punkt ($k = 0$) und Strecke ($k = 1$).

Bei der folgenden Argumentation spielt die Existenz echter Stützebenen H von M mit $M \not\subseteq H$ eine wesentliche Rolle. Daher wird M für $k = \dim M < n$ mit Hilfe des Komplementraums $C := L(\mathrm{aff}(M))^\perp$ zu einem volldimensionalen Zylinder $M + C$ aufgeblasen (im Bild grün).



Anahme: Es sei $z \in M \setminus \mathrm{konv}(E(M)) \neq \emptyset$. Da z keine Ecke ist, liegt es im Inneren einer Strecke $z \in (x,y)$ zwischen Punkten $z \neq x, y \in M$. Dabei können durch Verlängerung dieser Strecke x und y so gewählt werden, dass sie beide auf dem Rand von M liegen. Es sind aber nicht beide eine Ecke, da sonst $z \in \mathrm{konv}(E(M))$ wäre. Sei nun x keine Ecke. Nach Satz 3.2.13 existiert dann eine Stützebene H an $M + C$ mit $x \in H$ und $M + C \subseteq H^\ominus$. H ist insbesondere auch eine Stützebene an M mit $M \not\subseteq H$ und für die Stützmenge $R := H \cap M$ gilt $\dim R < k = \dim M$.

Nach Ind.Voraussetzung ist dann aber $x \in \text{konv}(E(R)) \subseteq \text{konv}(E(M))$, vgl. Satz 3.3.6. Analog zeigt man $y \in \text{konv}(E(M))$. Dies liefert den Widerspruch mit $z \in \text{konv}(E(M))$. ■

3.4 Polyeder, Polytope, Kegel

Theorem 3.3.7 liefert für kompakte, konvexe Mengen eine vollständige, explizite Darstellung mit Hilfe der Ecken. Für unbeschränkte Mengen muss diese Darstellung aber ergänzt werden. Dazu konzentrieren wir uns jetzt auf *Polyeder*. Dieser Begriff wurde schon mehrfach informell für die Lösungsmengen von Ungleichungssystemen benutzt und wird nun zusammen mit einem verwandten Begriff eingeführt. Insbesondere werden auch die Ecken und Kanten des Polyeders über seine algebraische Definition mit Daten aus dem Simplexverfahren identifiziert. Deshalb werden sowohl die zulässigen Polyeder von (LP1) als auch (LP3) betrachtet.

Definition 3.4.1 *Es sei $M \subseteq \mathbb{R}^n$ eine nichtleere Menge.*

a) M heißt Polyeder*, wenn eine Matrix $A \in \mathbb{R}^{m\times n}$ und ein Vektor $b \in \mathbb{R}^m$ existieren mit $M = \{x \in \mathbb{R}^n : Ax \geq b\}$.*

b) M heißt Polytop*, wenn (endlich viele) Punkte $x^{(0)}, \ldots, x^{(k)} \in \mathbb{R}^n$ existieren mit $M = konv(x^{(0)}, \ldots, x^{(k)})$. Wenn die Punkte $x^{(0)}, \ldots, x^{(k)}$ dabei affin linear unabhängig sind, nennt man M einen k-*Simplex*.*

Polyeder und Polytope sind natürlich konvex. Beim Polyeder treten insbesondere in Satz 3.2.15 nur endlich viele (höchstens m) Halbräume auf. Ein Polytop $M = \text{konv}(x^{(0)}, \ldots, x^{(k)})$ ist nach Satz 3.2.7 kompakt, da die Eckenmenge $E(M) \subseteq \{x^{(0)}, \ldots, x^{(k)}\}$ kompakt ist. In einem k-Simplex S hat jeder Punkt $z \in S$ eine eindeutige Darstellung

$$z = \sum_{j=0}^{k} \lambda_j x^{(j)}, \ (\lambda_j) \in \Delta_{k+1}.$$

Die zugehörigen λ_j sind die *baryzentrischen* Koordinaten von z in S, und $\bar{x} = \frac{1}{k+1} \sum_{j=0}^{k} x^{(j)}$ der *Schwerpunkt* von S.

Nach Theorem 3.3.7 ist ein Polytop durch seine Ecken explizit darstellbar. Im kompakten Fall gilt das auch für Polyeder, die zulässigen Bereiche von (LP):

Satz 3.4.2 *Ein nichtleeres, beschränktes Polyeder ist ein Polytop.*

Der Satz folgt direkt aus Theorem 3.3.7, wenn man weiß, dass jedes Polyeder nur endlich viele Ecken hat. Diese Tatsache wiederum folgt elementar aus dem jetzt hergeleiteten Zusammenhang (Satz 3.4.3) zwischen den Ecken von $X = \{x : Ax \geq b\}$ und ihrer algebraischen Charakterisierung durch die *regulären $n \times n$-Untermatrizen* von A. Da es überhaupt nur $\binom{m}{n}$ quadratische $n \times n$-Untermatrizen gibt, ist diese Zahl auch eine obere Schranke für die der Ecken.

Dabei spielen reguläre Untermatrizen $A^{(L)} \in \mathbb{R}^{n\times n}$, $L \subseteq \{1,\dots,m\}$, $n \le m$, bei (LP1) bzw $A_J \in \mathbb{R}^{m\times m}$, $J \subseteq \{1,\dots,n\}$, $n \ge m$, bei (LP3) eine entscheidende Rolle (zur Defin.vgl. §2.3).

Satz 3.4.3 *a) Das Polyeder $X = \{x : Ax \ge b\}$ zu (LP1) sei durch $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, gegeben und es sei $z \in X$. Dann ist z genau dann Ecke, wenn es eine reguläre $n \times n$-Untermatrix $A^{(L)}$, $L \subseteq \{1,\dots,m\}$, $|L| = n$, gibt mit $A^{(L)}z = b_L$.*

b) Das Polyeder $X = \{x : Ax = b, x \ge 0\}$ zu (LP3) sei durch $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, gegeben und es sei $z \in X$. Dann ist z genau dann Ecke, wenn $z \ge 0$ eine zulässige Basislösung ist, $rang(A_{J(z)}) = |J(z)|$.

Bemerkung: a) Wenn die Matrix A bei (LP1) nicht vollen Spaltenrang hat, also ein nichttrivialer Kern existiert, besitzt das Polyeder überhaupt keine Ecken, da mit $Ay = 0$, $y \ne 0$, und $x \in X$ auch $x + ty \in X\ \forall t \in \mathbb{R}$ gilt. Tatsächlich ist dann der Linealraum $L(X) = kern(A)$.

b) Bei (LP1) definiert das Teilsystem $A^{(L)}z = b_L$ aus "straffen" Bedingungen eindeutig den Schnittpunkt der n Hyperebenen $H(a^{(i)}, b_i)$, $i \in L$. Für eine Ecke z müssen aber auch die übrigen Zulässigkeitsbedingungen $A^{(K)}z \ge b_K$ mit $K = \{1,\dots,m\} \setminus L$ erfüllt sein. Diese sind i.d.R. "locker", $A^{(K)}z > b_K$.

c) Die Aussage zu (LP3) kann wegen Satz 2.3.2 analog zum ersten Teil von Satz 3.4.3 formuliert werden: $z \in X$ ist genau dann Ecke, wenn es eine reguläre $m\times m$-Untermatrix A_J, $J \subseteq \{1,\dots,n\}$, $|J| = m$, gibt mit $A_J z_J = b$. Einzige Zusatzbedingung ist hier $z \ge 0$.

Beweis a) Nach Bemerkung a) ist oBdA $m \ge n$ und $Rang(A) = n$.

"$\Leftarrow$" Für ein $z \in X$ gelte $A^{(L)}z = b_L$ und $z = \frac{1}{2}(x + y)$ mit $x, y \in X$. Dann folgt

$$0 = A^{(L)}z - b_L = \frac{1}{2}A^{(L)}(x+y) - b_L = \frac{1}{2}\big(\underbrace{A^{(L)}x - b_L}_{\ge 0}\big) + \frac{1}{2}\big(\underbrace{A^{(L)}y - b_L}_{\ge 0}\big) \ge 0.$$

Beide Klammern sind also null, wegen der Regularität von $A^{(L)}$ ist daher $x = z = y$, also z Ecke.

"$\Rightarrow$" Es sei $z \in X$ Ecke. Die Ungleichungen des Systems teilt man in straffe und lockere:

$$\left\{\begin{array}{lcl} A^{(L)}z & = & b_L, \\ A^{(K)}z & > & b_K, \end{array}\right. \qquad K + L = \{1,\dots,m\}.$$

Wenn $RangA^{(L)} < n$ wäre, gäbe es ein $u \ne 0$ mit $A^{(L)}u = 0$ und mit $t \in [-\varepsilon, \varepsilon]$, $\varepsilon > 0$, gilt

$$\left\{\begin{array}{lcl} A^{(L)}(z+tu) & = & b_L, \\ A^{(K)}(z+tu) & = & A^{(K)}z + tA^{(K)}u \ge b_K, \end{array}\right. \qquad \text{für } A^{(K)}z - \varepsilon|A^{(K)}u| \ge b_K.$$

Dann ergibt sich aber ein Widerspruch, denn $z = \frac{1}{2}(x^{(-)} + x^{(+)})$ ist echter Mittelpunkt der beiden Punkte $x^{(-)} = z - \varepsilon u \ne z + \varepsilon u = x^{(+)}$.

b) Hier sei $J = J(z)$, also $z_J > 0$, $K := N \setminus J$.

"$\Rightarrow$" Es sei z Ecke und $Rang(A_J) < |J|$. Dann existiert ein $u \ne 0$ mit $A_J u_J = 0$, $u_K = 0$. Wie in Teil a) ist dann $A_J(z_J + tu_J) = b\ \forall t \in \mathbb{R}$ und mit $\varepsilon := \min\{z_j/|u_j| : j \in J,\ u_j \ne 0\}$ erhält man den Widerspruch aus

$$x^{(-)} = z - \varepsilon u \ge 0,\ x^{(+)} = z + \varepsilon u \ge 0, \quad z = \frac{1}{2}(x^{(-)} + x^{(+)}).$$

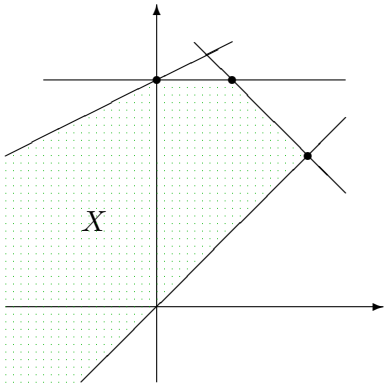
"$\Leftarrow$" Für $Rang A_J = |J|$ sei $z = \frac{1}{2}(x+y)$ mit $x, y \in X$. In den K-Komponenten folgt damit aber

$$0 = z_K = \frac{1}{2}(\underbrace{x_K}_{\geq 0} + \underbrace{y_K}_{\geq 0}) \;\Rightarrow\; x_K = y_K = 0.$$

Damit bleiben die eindeutig lösbaren Systeme $Ax = A_J x_J = b = A_J y_J \;\Rightarrow\; x_J = z_J = y_J$. ■

Beispiel 3.4.4 Bei (LP1) sei $m = 4$, $n = 2$ und

$$A = \begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 0 & -1 \\ 1 & -2 \end{pmatrix},\; b = \begin{pmatrix} 0 \\ -4 \\ -3 \\ -6 \end{pmatrix}$$



Es gibt $\binom{4}{2} = 6$ Indexmengen L mit $|L| = 2$, und da die zugehörigen Untermatrizen regulär sind, auch entsprechend viele Kreuzungspunkte von Hyperebenen (=Geraden). Allerdings sind nur drei davon zulässig, also Ecken von X:

$$1)\; L = \{1,2\}: \quad A^{(L)}x = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \end{pmatrix} = b_L: \quad x^{(1)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

$$2)\; L = \{2,3\}: \quad A^{(L)}x = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -4 \\ -3 \end{pmatrix} = b_L: \quad x^{(2)} = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

$$3)\; L = \{3,4\}: \quad A^{(L)}x = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -3 \\ -6 \end{pmatrix} = b_L: \quad x^{(3)} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}.$$

Das Beispiel zeigt, daß die Ecken hier nicht ausreichen, um die Menge X zu beschreiben. Die Menge enthält zusätzlich bestimmte Richtungen, in denen sie sich unendlich weit trichterförmig ausdehnt. Diese Gestalt läßt sich durch Kegel beschreiben, welche gegenüber konischen Kombinationen (vgl. Defin. 3.1.2) abgeschlossen sind.

Definition 3.4.5 *a) Die nichtleere Menge $K \subseteq \mathbb{R}^n$ heißt* konvexer Kegel*, wenn $\lambda x + \mu y \in K$, $\forall x, y \in K$, $\lambda, \mu \in \mathbb{R}_+$.*

b) Der konvexe Kegel $K \subseteq \mathbb{R}^n$, $K \neq \emptyset$, heißt spitz*, wenn $K \cap (-K) = \{0\}$ ist.*

c) Zu einer beliebigen Menge $M \subseteq \mathbb{R}^n$ ist

$$keg(M) := \bigcup_{k \in \mathbb{N}} \{\sum_{i=1}^{k} \lambda_i x^{(i)} : x^{(i)} \in M, \lambda_i \in \mathbb{R}_+\}$$

der von M erzeugte Kegel*. Ein Kegel K heißt* endlich erzeugt*, wenn $K = keg(b_1, \ldots, b_k)$ ist, $b_1, \ldots, b_k \in \mathbb{R}^n$, d.h.,*

$$K = B \cdot \mathbb{R}^k_+ = \{By : y \in \mathbb{R}^k_+\} \quad \text{mit } B = (b_1, \ldots, b_k) \in \mathbb{R}^{n \times k}. \tag{3.4.1}$$

Bemerkung: a) K konvexer Kegel $\iff K = \mathrm{keg}(K)$.

b) Wenn M schon konvex war, gilt einfach $\mathrm{keg}(M) = \mathbb{R}_+ \cdot M = \{\lambda x : x \in M, \lambda \geq 0\}$. Daher ist für beliebiges M auch $\mathrm{keg}(M) = \mathbb{R}_+ \cdot \mathrm{konv}(M)$.

c) Analog zur Situation bei konvexen Mengen sind Durchschnitte und Linearkombinationen von konvexen Kegeln wieder welche.

d) Die Darstellung (3.4.1) besagt, dass K als lineares Bild des Standard-Kegels $\mathbb{R}^k_+$ darstellbar ist (unter der zu B gehörigen linearen Abbildung).

e) Für einen konvexen Kegel K ist die affine Hülle $\mathrm{aff}(K) = K - K$ und der Linealraum $L(K) = K \cap (-K)$. Spitze Kegel haben also trivialen Linealraum.

Beispiel 3.4.6 a) $\mathbb{R}^n_+$ ist natürlich ein endlich erzeugter konvexer Kegel.

b) Lineare Unterräume $U \subseteq \mathbb{R}^n$ sind endlich erzeugte konvexe Kegel. Mit einer Basismatrix $B \in \mathbb{R}^{n\times l}$, $U = B \cdot \mathbb{R}^l$, läßt sich U auch als Kegel schreiben, $U = (B, -B) \cdot \mathbb{R}^{2l}_+$ (vgl. §1.3, Umformung 2).

Der folgende Kegel hat für die Behandlung von Polyedern zentrale Bedeutung.

Satz 3.4.7 *Gegeben sei das Polyeder $X = \{x : Ax \geq b\}$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Dann ist $O^+(X) := \{x : Ax \geq 0\}$ ein konvexer Kegel. Er wird* Ausdehnungskegel *von X genannt, es gilt*

$$O^+(X) = \{y : \ x + \lambda y \in X \ \forall x \in X, \lambda \in \mathbb{R}_+\}. \tag{3.4.2}$$

Beweis a) Für $y^{(j)} \in O^+(X)$ gilt also $Ay^{(j)} \geq 0$. Mit Vorfaktoren $\lambda_j \geq 0$ folgt aus

$$A\Big(\sum_j \lambda_j y^{(j)}\Big) = \sum_j \underbrace{\lambda_j}_{\geq 0} \underbrace{Ay^{(j)}}_{\geq 0} \geq 0$$

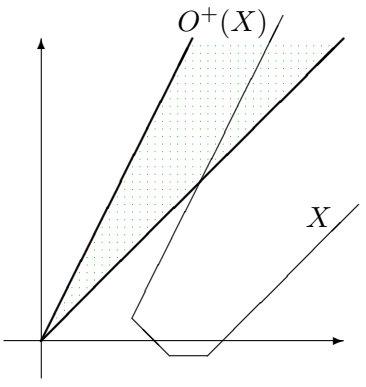
die Kegel-Eigenschaft. Und mit $Ax \geq b$, $Ay \geq 0$, $\lambda \geq 0$ gilt auch $A(x + \lambda y) = Ax + \lambda Ay \geq Ax \geq b$.

b) Sei $x \in X$, für $y \in \mathbb{R}^n$ und $\lambda > 0$ gelte

$$A(x + \lambda y) = Ax + \lambda Ay \geq b \ \Rightarrow \ Ay \geq \frac{1}{\lambda}(\underbrace{b - Ax}_{\leq 0}) \to 0 \ (\lambda \to \infty).$$

Das bedeutet $Ay \geq 0$. ∎

Die Formel (3.4.2) läßt sich als Definition des Kegels $O^+(X)$ für beliebige konvexe Mengen verstehen. Dieser Kegel enthält alle Richtungen, in die sich X unendlich weit ausdehnt. Bei Polyedern ist $O^+(X)$ insbesondere die Lösungsmenge des *homogenen* Ungleichungssystems analog zur Situation bei Linearen *Gleichungs*systemen.



Bemerkung: Für Polyeder $X \neq \emptyset$ gilt offensichtlich

a) $X + O^+(X) = X$.

b) X kompakt $\iff O^+(X) = \{0\}$.

c) $O^+(X)$ ist spitz, wenn $L(X) = kern(A) = \{0\}$.

d) Bedeutung für (LP1), $\min\{c^\mathsf{T}x : x \in X\}$: Für nichttriviales $c \in -O^+(X)$ ist (LP) unbeschränkt, denn da dann mit $\bar{x} \in X$ auch $x = \bar{x} - \lambda c \in X\ \forall \lambda \geq 0$ ist und $c^\mathsf{T}(\bar{x} - \lambda c) = c^\mathsf{T}\bar{x} - \lambda\|c\|^2$, folgt $\inf\{c^\mathsf{T}(\bar{x} - \lambda c) : \lambda \geq 0\} = -\infty$.

Beispiel 3.4.8 Zum Beispiel 3.4.4 ist der Ausdehnungskegel $O^+(X)$ durch das homogene System

$$\begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 0 & -1 \\ 1 & -2 \end{pmatrix} y \geq 0$$

bestimmt. Dieses entspricht den Bedingungen $y_1 \leq y_2 \leq 0$, $y_2 \leq -y_1$, $y_2 \leq y_1/2$. Also kommt nur $y_1 \leq 0$ in Frage und es bleiben nur $y_1 \leq y_2 \leq \frac{1}{2}y_1$. Das sind die Bedingungen zu $A^{(L)}y \geq 0$ mit $L = \{1, 4\}$. Die beiden homogenen Lösungen zu ${a^{(j)}}^\mathsf{T} y^{(j)} = 0$, $j \in L$, erzeugen diesen Kegel

$$O^+(X) = \mathrm{keg}\{y^{(1)}, y^{(4)}\} = \mathrm{keg}\{\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \end{pmatrix}\}.$$

Im zentralen Dekompositionssatz wird der Ausdehnungskegel benötigt, um Theorem 3.3.7 für unbeschränkte Polyeder zu ergänzen. Bisher ist aber nur die *implizite* Beschreibung von $O^+(X)$ aus Satz 3.4.7 durch das homogene Ungleichungssystem bekannt, unklar ist auch, ob eine endliche Erzeugermenge für ihn existiert.

Satz 3.4.9 *Der konvexe Kegel $K := \{x \in \mathbb{R}^n : Ax \geq 0\}$, $A \in \mathbb{R}^{m\times n}$, ist endlich erzeugt.*

Beweis Der Nachweis, dass $K := \{x : Ax \geq 0\}$ endlich erzeugt ist, wird über die Behauptung geführt, dass mit einem linearen Unterraum $U \subseteq \mathbb{R}^n$ auch der Schnitt $U \cap \mathbb{R}^n_+$ endlich erzeugt ist.

a) Spezialfall: $Y := Kern(B) \cap \mathbb{R}^m_+ = \{x : Bx = 0,\ x \geq 0\}$ ist endlich erzeugt. Durch eine *Homogenisierung* betrachtet man den kompakten Schnitt

$$M := Y \cap H(\mathbb{1}, 1) = Kern(B) \cap \mathbb{R}^m_+ \cap H(\mathbb{1}, 1) = Kern(B) \cap \Delta_m.$$

Dabei ist Δ_m kompakt, also auch M ($\neq \emptyset$ oBdA). Daher ist M ein Polytop (Satz 3.4.2), ist also Hülle seiner endlichen Eckenmenge $E(M)$, $M = \mathrm{konv}(E(M))$. Durch Streckung von M bekommt man Y zurück: $Y = \mathbb{R}_+M = \mathrm{keg}(E(M))$ ist endlich erzeugt.

b) Anwendung für $Y := AK \subseteq \mathbb{R}^m_+$: Jeder lineare Unterraum, auch $U := Bild(A) = A\mathbb{R}^n = \{Ax : x \in \mathbb{R}^n\}$, ist Kern einer linearen Abbildung, $U = Kern(B)$. Nach Teil a) ist Y endlich erzeugt, daher existieren $y^{(j)} = Ax^{(j)} \in U$, $j = 1, \ldots, k$, mit

$$Y = U \cap \mathbb{R}^m_+ = \mathrm{keg}(y^{(1)}, \ldots, y^{(k)}) = \mathrm{keg}(Ax^{(1)}, \ldots, Ax^{(k)}).$$

Außerdem sei $Kern(A) = span(z^{(1)}, \ldots, z^{(\ell)})$. Für $x \in K$ ist $y := Ax \in Y$ und es gilt:

$$y = Ax = \sum_{j=1}^{k} \lambda_j Ax^{(j)},\ (\lambda_j) \geq 0 \iff A(\underbrace{x - \sum_{j=1}^{k} \lambda_j x^{(j)}}_{\in Kern(A)}) = 0.$$

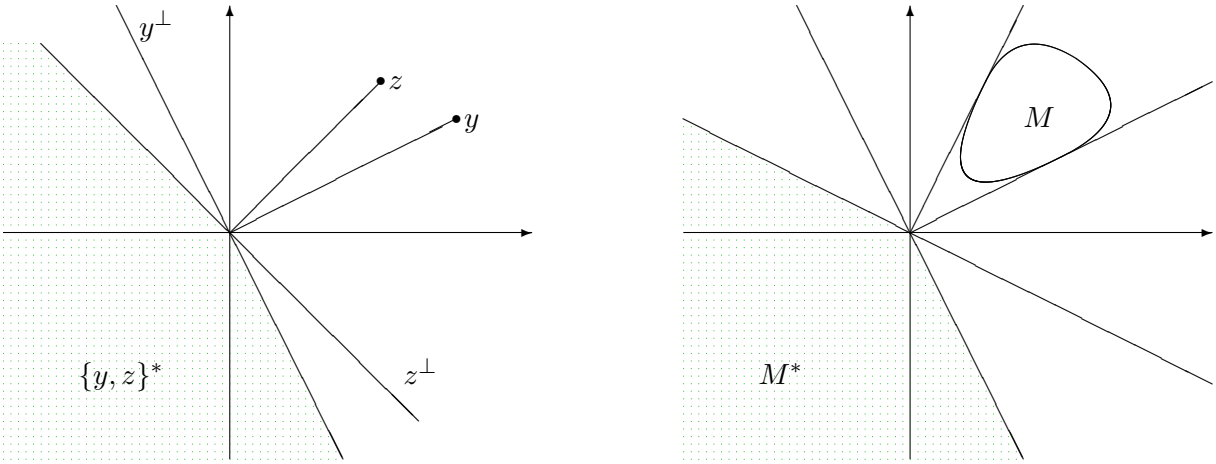
$$\Rightarrow\ x - \sum_{j=1}^{k} \lambda_j x^{(j)} = \sum_{i=1}^{\ell} \mu_i z^{(i)} \iff x \in \mathrm{keg}\Big(x^{(1)}, \ldots, x^{(k)}, z^{(1)}, \ldots, z^{(\ell)}, -z^{(1)}, \ldots, -z^{(\ell)}\Big),$$

denn jedes Kernelement ist Linearkombination der $z^{(i)}$, bzw. konische Kombination der $\pm z^{(i)}$. ■

Bevor die Zerlegung von Polyedern weiter verfolgt wird, wird kurz ein abgeleiteter Kegel studiert, der die Interpretation einiger Ergebnisse erleichtert.

Definition 3.4.10 *Der* Polarkegel *(duale Kegel) zu einer nichtleeren Menge $M \subseteq \mathbb{R}^n$ ist*

$$M^* := \{x \in \mathbb{R}^n : y^\mathsf{T} x \le 0 \,\forall y \in M\} = \bigcap_{y \in M} H^{\ominus}(y, 0).$$



Bemerkung: a) Für einen linearen Unterraum $U \subseteq \mathbb{R}^n$ ist $U^* = U^\perp$.

b) Für $M \neq \emptyset$ gilt $M^* = \big(\mathrm{keg}(M)\big)^*$ und $M \subseteq M^{**} := (M^*)^*$.

c) Der Definition nach entspricht der Polyeder-Kegel $K = \{x : Ax \ge 0\} = O^+(X)$ gerade dem Polarkegel zu den negativen Zeilen von A, $K = \{-a^{(1)}, \ldots, -a^{(m)}\}^* = (-A^\mathsf{T} \cdot \mathbb{R}^m_+)^*$.

Bemerkung b) kann für die hier interessierenden Kegel präzisiert werden (o.Bew.).

Satz 3.4.11 *Für einen endlich erzeugten konvexen Kegel K gilt $K^{**} = K$.*

Also ist für $K = \{x : Ax \ge 0\}$ der Polarkegel $K^* = -A^\mathsf{T} \cdot \mathbb{R}^m_+$ und beide daher endlich erzeugt. Mit diesem Satz kann die obige Bemerkung d) zur Unbeschränkheit von (LP1) präzisiert werden. Für $\bar{x} \in X$, $y \in O^+(X)$ ist auf dem Strahl $\{x = \bar{x} + \lambda y : \lambda \ge 0\} \subseteq X$ der Wert der Zielfunktion $c^\mathsf{T} x = c^\mathsf{T} \bar{x} + \lambda c^\mathsf{T} y$ genau dann (nach unten) beschränkt, wenn $c^\mathsf{T} y \ge 0$ gilt. Beschränkheit erfordert also $c^\mathsf{T} v \ge 0 \,\forall v \in O^+(X)$. Dies heißt aber gerade, dass $-c$ im Polarkegel $(O^+(X))^* = -A^\mathsf{T} \cdot \mathbb{R}^m_+$ liegt. Dieses Ergebnis *(LP1) beschränkt* $\iff$ $c \in A^\mathsf{T} \cdot \mathbb{R}^m_+$ wird in der Dualitätstheorie wieder auftauchen.

3.5 Der Dekompositionssatz für Polyeder

Zur Ergänzung der Polyeder-Zerlegung muss auch der Ausdehnungskegel berücksichtigt werden. Bei der endlichen Darstellung von Polyeder-Kegeln, vgl. Satz 3.4.9, kann eine Minimalmenge erforderlicher Richtungen identifiziert werden, die Kanten des Kegels. Daher wird jetzt das dem Satz 3.4.3 (Eckendarstellung) entsprechende Resultat für die Kanten der der zu (LP1) bzw. (LP3) gehörenden zulässigen Mengen formuliert. Bei (LP1) wird die Aussage wegen des Dekompositionssatzes auf den Ausdehnungskegel beschränkt. Bei (LP3) wird dagegen konkret gezeigt, dass der *elementare Strahl* (2.3.7) gerade eine Polyeder-Kante darstellt, wenn er eine positive (evtl. unendliche) Länge hat. Letzteres ist an den Vorzeichen des Vektors $w_\ell^{(J)} = A_J^{-1} a_\ell$ erkennbar. Damit wird der Zusammenhang zu den Daten des Simplexverfahrens hergestellt.

Satz 3.5.1 *a) Es sei $A \in \mathbb{R}^{m\times n}$, gegeben. Zu $y \in \{x : Ax \geq 0\} \setminus \{0\}$ ist keg(y) genau dann Kante, wenn eine Untermatrix $A^{(L)}$ maximalen Ranges $|L| = n-1$ existiert mit $A^{(L)}y = 0$.*

b) Das Polyeder $X = \{x : Ax = b,\, x \geq 0\}$ zu (LP3) sei durch $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, gegeben und es sei $z \in X$ Ecke mit Basis A_J. Für $\ell \in K = \{1,\dots,n\} \setminus J$ und den Spaltenvektor $w_\ell = A_J^{-1} a_\ell$ der Matrix W_K aus (2.3.6) gelte

$$J^+(w_\ell) \subseteq J(z), \quad \text{d.h. } w_{i\ell} > 0 \Rightarrow z_i > 0 \ \forall i \in J.$$

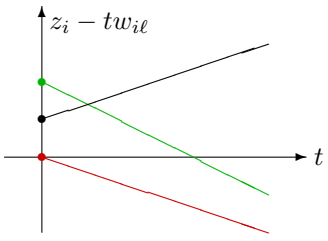
Dann ist $\{z - tw_\ell : t \geq 0\} \cap X$ Kante von X.

Beweis a) ist analog zum Beweis von Satz 3.4.3a), wegen des eindimensionalen Kerns ist der Vorfaktor bei ty frei.
b) Nach (2.3.7) wird das Gleichungssystem $Ax(t) = b$ durch jeden Punkt des Strahls

$$x(t) = z - tw_\ell = \begin{pmatrix} A_J^{-1}(b - ta_\ell) \\ te_\ell^{(K)} \end{pmatrix}$$

erfüllt. Zu prüfen ist das Vorzeichen $x(t) \geq 0$, $\forall t \in [0,\varepsilon]$, $\varepsilon > 0$. Dabei ist der Fall

- $w_{i\ell} \leq 0$: keine Einschränkung an t,
- $z_i > 0, w_{i\ell} > 0$: erfüllbar mit $t > 0$,
- $z_i = 0, w_{i\ell} > 0$: unerfüllbar für $t > 0$.



Nach Voraussetzung tritt der letzte (rote) Fall nicht auf und die Kante hat daher eine positive Länge $t_\ell > 0$, vgl. (2.3.10).
ZZ $S := \{x(t) : t \in [0,\infty)\} \cap X$ ist Kante. Für festes $t > 0$ ist $J(x(t)) \subseteq J \cup \{\ell\}$ mit $\ell \in K$. Nun sei $x(t) = \frac{1}{2}(u+v)$ mit $u, v \in X$.
Wie früher folgt daraus $J(u), J(v) \subseteq J \cup \{\ell\}$, denn für einen Index $k \in K$ gilt, wenn

$$\begin{aligned} k = \ell: &\quad x_\ell(t) = t = \tfrac{1}{2}(u_\ell + v_\ell) &&\Rightarrow\ u_\ell = \alpha t, v_\ell = (2-\alpha)t,\ \alpha \in [0,2], \\ k \neq \ell: &\quad x_k(t) = 0 = \tfrac{1}{2}(u_k + v_k) &&\Rightarrow\ u_k = v_k = 0. \end{aligned}$$

Mit der Basisdarstellung (2.3.5) überträgt sich das auf die J-Komponenten:

$$u_J = z_J - \alpha t A_J^{-1} a_\ell \in S, \quad v_J = z_J - (2-\alpha) t A_J^{-1} a_\ell \in S,$$

somit kann $x(t)$ nur aus Elementen von S konvex kombiniert werden, S ist daher Kante. ∎

Nur spitze Kegel besitzen Ecken. Eine wichtige Schlußweise in spitzen Kegeln K ist, dass für die Null nur die triviale konische Kombination möglich ist,

$$\sum_{i=1}^{k} \lambda_i y^{(i)} = 0, \text{ mit } y^{(i)} \in K, \ \lambda_i \geq 0 \quad \Rightarrow \quad (\lambda_i) = 0.$$

Denn andernfalls wäre für $\lambda_j > 0$ mit $y^{(j)}$ auch $-y^{(j)} = \sum_{i \neq j} (\lambda_i/\lambda_j) y^{(i)} \in K$ und K hätte nichttrivialen Linealraum, da $\text{keg}(y^{(j)}, -y^{(j)}) = span(y^{(j)}) \subseteq L(K)$.

Satz 3.5.2 *Wenn der konvexe Kegel $K := \{x : Ax \geq 0\}$, $A \in \mathbb{R}^{m \times n}$, spitz ist, kann K durch die Richtungen seiner Kanten erzeugt werden.*

Beweis Nach Satz 3.4.8 ist $K = \text{keg}(y^{(1)}, \ldots, y^{(k)})$ darstellbar. Diese Darstellung sei oBdA minimal, also kein $y^{(i)}$ als konische Kombination der anderen darstellbar. Alle $y \in K \setminus \{0\}$ besitzen eine konische Darstellung $y = \sum_{i=1}^{k} \lambda_i y^{(i)}$, $\lambda_i \geq 0$. Wenn dabei mindestens zwei $\lambda_i > 0$ sind, ist $\text{keg}\{y\}$ keine Kante. ZZ Strahl $S := \text{keg}(y^{(j)}) = \{\alpha y^{(j)} : \alpha \geq 0\}$ ist Kante von K. Dazu wird für $\alpha > 0$ eine beliebige Konvexkombination betrachtet mit $x = \sum_{i=1}^{k} \mu_i y^{(i)}$, $z = \sum_{i=1}^{k} \nu_i y^{(i)}$, $(\mu_i), (\nu_i) \in \mathbb{R}^k_+$, und $\lambda \in (0,1)$ für

$$\begin{aligned} \alpha y^{(j)} &= \lambda x + (1-\lambda) z = \sum_{i=1}^{k} \Big(\lambda \mu_i + (1-\lambda)\nu_i\Big) y^{(i)} = \sum_{i=1}^{k} \lambda_i y^{(i)} \\ \Rightarrow \ (\alpha - \lambda_j) y^{(j)} &= \sum_{i \neq j} \lambda_i y^{(i)}, \quad \lambda_i = \lambda \mu_i + (1-\lambda)\nu_i \geq 0. \end{aligned}$$

Fall $\alpha - \lambda_j > 0$: Division durch $\alpha - \lambda_j > 0$ ergibt konische Kombination von $y^{(j)}$, Widerspruch zur Minimalannahme.

$\lambda_j - \alpha > 0$: $0 = (\lambda_j - \alpha) y^{(j)} + \sum_{i \neq j} \lambda_i y^{(i)}$ ist nichttriviale konische Kombination der Null, die aber n.V. nur trivial möglich ist,

$\lambda_j - \alpha = 0$ und $\lambda_i = 0$ für $i \neq j$ ist die einzige mögliche Situation

$\Rightarrow$ $\alpha y^{(j)} = \lambda_j y^{(j)}$ ist nur durch $x, z \in S$ selbst darstellbar, also ist S Kante. ∎

Für das folgende Theorem wird $\text{keg}(\emptyset) := \{0\}$ verabredet.

Theorem 3.5.3 (Dekompositionssatz) *Es sei $X := \{x \in \mathbb{R}^n : Ax \geq b\} \neq \emptyset$ das durch $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ bestimmte Polyeder, und $L(X) = \{0\}$. Dann ist X die Summe eines Polytops und eines endlich erzeugten Kegels. Mit den Ecken $x^{(i)}$, $i = 1, \ldots, k$, von X und Kantenrichtungen $y^{(j)}$, $j = 1, \ldots, \ell$, von $O^+(X)$ gilt*

$$\begin{aligned} X &= konv\big(E(X)\big) + O^+(X) \\ &= konv\big(x^{(1)}, \ldots, x^{(k)}\big) + keg\big(y^{(1)}, \ldots, y^{(\ell)}\big). \end{aligned}$$

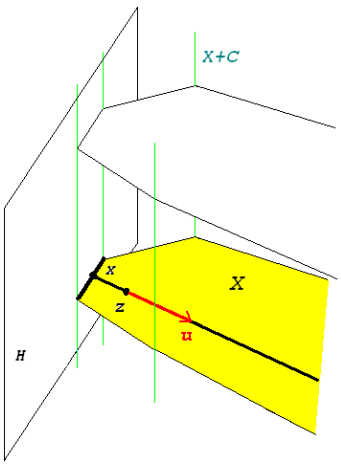
Beweis Der Beweis verläuft analog zum Satz von Krein-Milman unter Einbeziehung des Ausdehnungskegels durch Induktion über $q = \dim X$. Für die Existenz nichttrivialer Stützebenen wird bei Bedarf wieder mit dem volldimensionalen konvexen Polyeder $X + C$, $C = L(\text{aff}(X))^\perp$, gearbeitet. Für Punkt oder Strecke/Strahl gilt die Aussage mit $q \leq 1$. Nun sei $z \in X$ beliebig. Dann gilt einer der Fälle

a) $z \in Rd(X+C)$: nach S. 3.2.13 existiert eine Stützebene H mit $z \in H \cap X$, $(X+C) \subseteq H^{\ominus}$ und $X \not\subseteq H$. Dann ist $\dim(X \cap H) < q$ und die Behauptung folgt aus der I.V.

b) z liegt im Inneren von $X+C$. Dann existiert eine Gerade $G := \{z + tu : t \in \mathbb{R}\}$ durch z, die ein Stück weit in X verläuft, $G \cap (X \setminus \{z\}) \neq \emptyset$. Dabei ist $u \in L(\mathrm{aff}(X))$. Wegen $L(X) = \{0\}$ kann G nicht vollständig zu X gehören, $G \not\subseteq X$, und schneidet daher den Rand von $X + C$.

b1) Es gibt zwei Schnittpunkte x, y mit dem Rand und $z = \lambda x + (1-\lambda)y$, $\lambda \in (0,1)$. Für x und y trifft Fall a) zu.

b2) Es gibt einen Schnittpunkt x mit dem Rand und $x + tu \in X$ $\forall t \geq 0$. Dann ist $u \in O^+(X) = \mathrm{keg}(y^{(1)}, \ldots, y^{(\ell)})$ nach Satz 3.5.2 und zeigt die Behauptung, denn für x trifft wieder Fall a) zu.



■

Der Dekompositionssatz verallgemeinert den Satz über Lösungsmengen von Linearen *Gleichungs*systemen, verwendet aber mehrere spezielle inhomogene Lösung**en** $E(X)$ und die allgemeine homogene Lösung im Kegel $O^+(X)$.

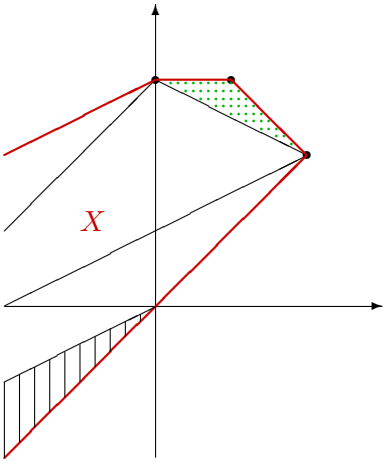
$$\begin{array}{lllllll} \text{LGS} & Ax = b: & X = & \{\hat{x}\} & + & Kern(A) \\ \text{UGlS} & Ax \geq b: & X = & \mathrm{konv}(E(X)) & + & O^+(X). \end{array}$$

Beispiel 3.5.4 Zusammenfassung der Beispiele 3.4.4/8: das Polyeder $X := \{x : Ax \geq b\}$ mit

$$A = \begin{pmatrix} -1 & 1 \\ -1 & -1 \\ 0 & -1 \\ 1 & -2 \end{pmatrix}, \; b = \begin{pmatrix} 0 \\ -4 \\ -3 \\ -6 \end{pmatrix}$$

läßt sich darstellen in der Form

$$X = \mathrm{konv}\{\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}1\\3\end{pmatrix}, \begin{pmatrix}0\\3\end{pmatrix}\} + \mathrm{keg}\{\begin{pmatrix}-1\\-1\end{pmatrix}, \begin{pmatrix}-2\\-1\end{pmatrix}\}.$$



Im Bild zeigt der punktierte Teil das Polytop $\mathrm{konv}(E(X))$, unten ist schraffiert der Ausdehnungskegel $O^+(X)$ eingezeichnet, welcher im Theorem an jeden Punkt des Polytops "angeheftet" wird. Die zwei extremalen verschobenen Kegel sind ebenfalls angedeutet.

Bedeutung für das Simplex-Verfahren: Der Dekompositions-Satz 3.5.3 ist die Arbeitsgrundlage für das Simplexverfahren. Da das Minimum der linearen Zielfunktion von (LP), wenn es existiert, auch auf den Ecken angenommen wird, müssen daher nur diese untersucht werden. Und Satz 3.4.3 bestätigt, dass diese gerade durch Basislösungen gegeben sind. Um zusätzlich die Beschränktheit sicherzustellen, sind auch diejenigen Kanten des Polyeders, auf denen die Zielfunktion wächst, auf endliche Länge zu prüfen. Satz 3.5.1 stellt hierfür die Verbindung zum Simplexverfahren her.

3.6 Existenzsätze für Ungleichungssysteme

Die bisherigen Sätze bezogen sich naturgemäß auf den Fall nichtleerer zulässiger Bereiche X. Kriterien für die Gültigkeit dieser Voraussetzung, d.h., die Lösbarkeit der Ungleichungssysteme, werden jetzt als weitere Anwendung der Trennungssätze aus §3.2 hergeleitet. Grundlage ist das folgende Lemma von Farkas, es bildet insbesondere auch die Basis für die wichtige Dualitätstheorie linearer Programme. Die klassische Form orientiert sich an (LP3):

Satz 3.6.1 (Farkas) *Mit $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ gilt*

$$\{x \in \mathbb{R}^n : Ax = b,\, x \geq 0\} \neq \emptyset \iff \Big(y^\mathsf{T} A \leq 0^\mathsf{T} \Rightarrow y^\mathsf{T} b \leq 0 \,\forall y \in \mathbb{R}^m\Big). \tag{3.6.1}$$

Beweis "⇒" Wenn ein $\hat{x} \geq 0$ existiert mit $A\hat{x} = b$ ergibt sich direkt

$$y^\mathsf{T} A \leq 0 \;\Rightarrow\; y^\mathsf{T} b = (y^\mathsf{T} A)\underbrace{\hat{x}}_{\geq 0} \leq 0.$$

"⇐" Nun gelte die Folgerung "$y^\mathsf{T} A \leq 0^\mathsf{T} \Rightarrow y^\mathsf{T} b \leq 0\,\forall y \in \mathbb{R}^m$", die Lösungsmenge sei aber leer. Dann liegt also b nicht im abgeschlossenen Kegel $K := A\mathbb{R}^n_+ = \{Ax : x \geq 0\} = \mathrm{keg}\{a_1, \ldots, a_n\}$. Nach Satz 3.2.13 existiert daher eine strikt trennende Hyperebene $H(q, \alpha)$ mit

$$K \subseteq H^-(q,\alpha) \quad \text{und} \quad b \in H^+(q,\alpha) \subseteq H^+(q,0).$$

Denn wegen $0 \in K$ ist dabei $0 < \alpha$ und daher $q^\mathsf{T} b > \alpha > 0$. Für alle Strahlen $y^{(j)} := \lambda a_j = \lambda A e_j$, $\lambda > 0$, $j \in N$, gilt natürlich $y^{(j)} \in K \subseteq H^-$, also

$$q^\mathsf{T} y^{(j)} = \lambda q^\mathsf{T} a_j < \alpha \;\Rightarrow\; q^\mathsf{T} a_j \leq \inf_{\lambda>0} \frac{\alpha}{\lambda} = 0 \quad \forall j = 1, \ldots, n.$$

Damit ist aber $q^\mathsf{T} A \leq 0$ und n.V. $q^\mathsf{T} b \leq 0$, also $b \in H^\ominus(q,0)$ im Widerspruch zu $b \in H^+(q,0)$. ∎

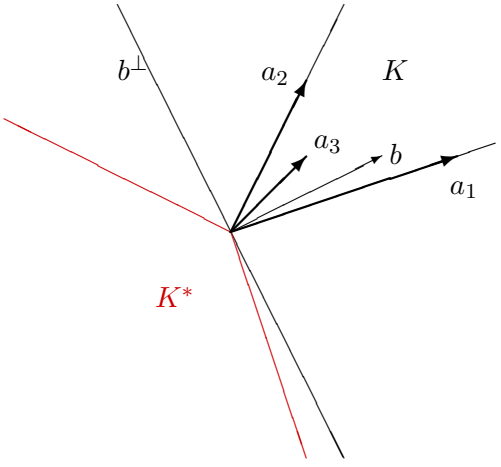
Geometrische Interpretation: Die Lösbarkeit des Systems auf der linken Seite bedeutet, dass b als konische Kombination der Spalten von A ausgedrückt werden kann, $b \in A\mathbb{R}^n_+ =: K$. Die rechte Seite von (3.6.1) heißt, dass $y \in H^\ominus(b,0) = \{b\}^*$ gilt für jeden Vektor $y \in \{a_1, \ldots, a_n\}^*$ aus dem Polarkegel $K^* = (A\mathbb{R}^n_+)^*$. Also entspricht (3.6.1) der einfachen Aussage:

$$b \in A \cdot \mathbb{R}^n_+ = \mathrm{keg}\{a_1, \ldots, a_n\} \iff \{a_1, \ldots, a_n\}^* \subseteq \{b\}^* = H^\ominus(b,0).$$

Beispiel 3.6.2 Bei

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

ist $a_3 = \frac{1}{5}a_1 + \frac{2}{5}a_2$, also $K := A\mathbb{R}^3_+ = \mathrm{keg}\{a_1, a_2\}$. Daher ist der Polarkegel $K^* = \{y : 3y_1 + y_2 \leq 0,\, y_1 + 2y_2 \leq 0\}$, und ist darstellbar als $K^* = \mathrm{keg}\{y^{(1)}, y^{(2)}\}$ mit $y^{(1)} = \binom{1}{-3}$, $y^{(2)} = \binom{-2}{1}$. Es liegt $K^* \subseteq \{b\}^*$, wenn alle Erzeugenden $y^{(i)}$ dies tun. Also gilt $b \in K \iff b^\mathsf{T} y^{(i)} \leq 0$, $i = 1, 2$.



Analoge Lösbarkeitssätze gibt es auch für die allgemeine Standardform.

Satz 3.6.3 *Mit $A_{ij} \in \mathbb{R}^{m_i \times n_j}$, $b_i \in \mathbb{R}^{m_i}$, $i,j = 1,2$, sind äquivalent:*

$$\exists\, x_1 \in \mathbb{R}^{n_1},\, x_2 \in \mathbb{R}^{n_2} \;\; mit \quad \left\{ \begin{array}{cc} A_{11}x_1 + A_{12}x_2 & \geq b_1 \\ A_{21}x_1 + A_{22}x_2 & = b_2 \\ x_1 & \geq 0 \end{array} \right.$$

und

$$\forall\, y_1 \in \mathbb{R}^{m_1},\, y_2 \in \mathbb{R}^{m_2} \;\; mit \left\{ \begin{array}{cc} y_1^\mathsf{T} A_{11} + y_2^\mathsf{T} A_{21} & \leq 0^\mathsf{T} \\ y_1^\mathsf{T} A_{12} + y_2^\mathsf{T} A_{22} & = 0^\mathsf{T} \\ y_1 & \geq 0 \end{array} \right\} \Rightarrow y_1^\mathsf{T} b_1 + y_2^\mathsf{T} b_2 \leq 0.$$

Beweis Umformung mit Schlupfvariablen $z \geq 0$ und der Zerlegung $x_2 = x_2^+ - x_2^-$, $x_2^\pm \geq 0$, ergibt die Form $A\bar{x} = b$, $\bar{x} \geq 0$ mit

$$A = \begin{pmatrix} A_{11} & A_{12} & -A_{12} & -I \\ A_{21} & A_{22} & -A_{22} & 0 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \bar{x} = \begin{pmatrix} x_1 \\ x_2^+ \\ x_2^- \\ z \end{pmatrix}$$

Dieses ist genau dann lösbar, wenn die Folgerung gilt:

$$0 \geq A^\mathsf{T} y = \begin{pmatrix} A_{11}^\mathsf{T} y_1 + A_{21}^\mathsf{T} y_2 \\ A_{12}^\mathsf{T} y_1 + A_{22}^\mathsf{T} y_2 \\ -A_{12}^\mathsf{T} y_1 - A_{22}^\mathsf{T} y_2 \\ -y_1 \end{pmatrix} \;\Rightarrow\; y^\mathsf{T} b = y_1^\mathsf{T} b_1 + y_2^\mathsf{T} b_2 \leq 0.$$

Die mittleren Ungleichungen bedeuten natürlich $A_{12}^\mathsf{T} y_1 + A_{22}^\mathsf{T} y_2 = 0$. ∎

Die anderen Formen der Standardprogramme sind darin als Spezialfälle enthalten, als Übersicht:

$$\begin{array}{lrcl}
\text{(LP1)} & \{x \in \mathbb{R}^n :\, Ax \geq b\} \neq \emptyset & \iff & \{y^\mathsf{T} A = 0^\mathsf{T} \Rightarrow y^\mathsf{T} b \leq 0 \;\forall y \in \mathbb{R}^m_+\} \\
\text{(LP2)} & \{x \in \mathbb{R}^n :\, Ax \geq b,\, x \geq 0\} \neq \emptyset & \iff & \{y^\mathsf{T} A \leq 0^\mathsf{T} \Rightarrow y^\mathsf{T} b \leq 0 \;\forall y \in \mathbb{R}^m_+\} \\
\text{(LP3)} & \{x \in \mathbb{R}^n :\, Ax = b,\, x \geq 0\} \neq \emptyset & \iff & \{y^\mathsf{T} A \leq 0^\mathsf{T} \Rightarrow y^\mathsf{T} b \leq 0 \;\forall y \in \mathbb{R}^m\} \\
\text{(LGS)} & \{x \in \mathbb{R}^n :\, Ax = b\} \neq \emptyset & \iff & \{y^\mathsf{T} A = 0^\mathsf{T} \Rightarrow y^\mathsf{T} b = 0 \;\forall y \in \mathbb{R}^m\}
\end{array}$$

Als vierte Variante wurden Gleichungssysteme aufgenommen. Das Lösbarkeitskriterium dort ist bekanntlich $b \in (A \cdot \mathbb{R}^n) = kern(A^\mathsf{T})^\perp$ und wird oft als Fredholm-Alternative formuliert. Auch die obigen Kriterien können als Alternativsätze formuliert werden, z.B.:

$$\begin{array}{llllll}
\text{(LGS)} & \text{Entweder ist} & Ax = b & \text{lösbar, oder} & y^\mathsf{T} A = 0^\mathsf{T}, & y^\mathsf{T} b = 1 \\
\text{(LP1)} & \text{Entweder ist} & Ax \geq b & \text{lösbar, oder} & y^\mathsf{T} A = 0^\mathsf{T},\, y \geq 0, & y^\mathsf{T} b = 1 \\
\text{(LP3)} & \text{Entweder ist} & Ax = b,\, x \geq 0 & \text{lösbar, oder} & y^\mathsf{T} A \leq 0^\mathsf{T}, & y^\mathsf{T} b = 1
\end{array}$$

Die Merkregeln für den Zusammenhang zwischen den Alternativsystemen entsprechen denen bei der Dualität und werden dort formuliert.

4 Duale Programme

4.1 Optimalitätskriterien

Im letzten Abschnitt konnte die Lösbarkeit eines Ungleichungssystems mit Eigenschaften eines davon abgeleiteten Systems in Beziehung gesetzt werden. Dieser Zusammenhang kann auf vollständige Lineare Programme durch Betrachtung ihrer dualen Versionen ausgeweitet werden. Als wichtige Arbeitshilfe für die Praxis werden dabei Kriterien für die *Optimalität* eines zulässigen Punktes x hergeleitet, die (etwa durch einen Auftraggeber) *effektiv nachprüfbar* sind, da sie nur wenige Berechnungsschritte erfordern ("Einsetzen").

Ansatzpunkt ist eine Standardmethode bei Extremalproblemen mit Nebenbedingungen, die Verwendung von *Lagrange-Multiplikatoren*. Beim Problem (LP1) hat man m Nebenbedingungen $Ax - b \geq 0$, verwendet dazu also Multiplikatoren $y \in \mathbb{R}^m$ und bildet die Lagrangefunktion

$$\phi(x, y) = c^\mathsf{T} x + y^\mathsf{T}(b - Ax) = y^\mathsf{T} b + (c^\mathsf{T} - y^\mathsf{T} A)x.$$

Die rechte Version zeigt, dass ϕ auch als Lagrangefunktion eines Extremalproblems für y, des *dualen* Problems, interpretiert werden kann. Beim Umgang damit sind aber auch Vorzeichenbedingungen zu berücksichtigen. Der Vollständigkeit halber wird die duale Form (LP*) zunächst zum allgemeinen *primalen* Programm (LP) angegeben.

$$\text{(LP)} \qquad \left.\begin{array}{rll} \min\ c_1^\mathsf{T} x_1 + c_2^\mathsf{T} x_2 & & \\ A_{11}x_1 + A_{12}x_2 & \geq b_1 & \\ A_{21}x_1 + A_{22}x_2 & = b_2 & \\ x_1 & \geq 0 & \end{array}\right\} \qquad \left\{\begin{array}{rll} \max\ b_1^\mathsf{T} y_1 + b_2^\mathsf{T} y_2 & & \\ A_{11}^\mathsf{T} y_1 + A_{21}^\mathsf{T} y_2 & \leq c_1 & \\ A_{12}^\mathsf{T} y_1 + A_{22}^\mathsf{T} y_2 & = c_2 & \\ y_1 & \geq 0 & \end{array}\right. \qquad \text{(LP*)}$$

In der Regel betrachtet man aber eine der Standardformen (LP1..3), für diese ist

(LP1)	$\min\ c^\mathsf{T} x$ $Ax \geq b$	$\max\ b^\mathsf{T} y$ $A^\mathsf{T} y = c$ $y \geq 0$ (LP1*)
(LP2)	$\min\ c^\mathsf{T} x$ $Ax \geq b$ $x \geq 0$	$\max\ b^\mathsf{T} y$ $A^\mathsf{T} y \leq c$ $y \geq 0$ (LP2*)
(LP3)	$\min\ c^\mathsf{T} x$ $Ax = b$ $x \geq 0$	$\max\ b^\mathsf{T} y$ $A^\mathsf{T} y \leq c$ (LP3*)

Die Übersicht zeigt jetzt den Grund, warum die Form (LP2) überhaupt betrachtet wird. Es ist dasjenige Programm, bei dem das duale i.w. die gleiche Gestalt hat. Die Übergänge (LP) $\rightarrow$ (LP*) und (LP*) $\rightarrow$ (LP**)=(LP) sind symmetrisch. Die Begründung für die Details der dualen Form liefern die im Anschluß folgenden Sätze, der Übergang geschieht nach folgenden Merkregeln:

1. Aus einem Minimum-Problem wird ein Maximierungsproblem,

2. die Koeffizientenmatrix wird transponiert,
3. der Gradientenvektor der Zielfunktion wird mit der rechten Seite des (Un-) Gleichungssystems getauscht,
4. Ungleichungsrestriktionen werden ausgetauscht durch vorzeichenbeschränkte Variable, Gleichungen durch freie Variable und umgekehrt.

Für die Zielfunktionen in zulässigen Punkten von primalem und dualem Programm gibt es einen grundlegenden Zusammenhang:

Satz 4.1.1 *Der Vektor $x^\mathsf{T} = (x_1^\mathsf{T}, x_2^\mathsf{T})$ sei zulässig für (LP) und $y^\mathsf{T} = (y_1^\mathsf{T}, y_2^\mathsf{T})$ zulässig für (LP*). Dann gilt für die Zielfunktionen $c^\mathsf{T}x = c_1^\mathsf{T}x_1 + c_2^\mathsf{T}x_2$ und $b^\mathsf{T}y = b_1^\mathsf{T}y_1 + b_2^\mathsf{T}y_2$ die Beziehung*

$$c^\mathsf{T}x \geq b^\mathsf{T}y.$$

Bei Gleichheit, $c^\mathsf{T}\hat{x} = b^\mathsf{T}\hat{y}$, ist $\hat{x}$ optimal *für (LP) und $\hat{y}$* optimal *für (LP*).*

Beweis Für primal zulässige $x \in X \subseteq \mathbb{R}^n$ bzw. dual zulässige $y \in Y \subseteq \mathbb{R}^m$ gilt

$$\begin{aligned} y^\mathsf{T}b &= y_1^\mathsf{T}b_1 + y_2^\mathsf{T}b_2 \leq \underbrace{y_1^\mathsf{T}}_{\geq 0}(A_{11}x_1 + A_{12}x_2) + y_2^\mathsf{T}(A_{21}x_1 + A_{22}x_2) \\ &= (y_1^\mathsf{T}A_{11} + y_2^\mathsf{T}A_{21})\underbrace{x_1}_{\geq 0} + (y_1^\mathsf{T}A_{12} + y_2^\mathsf{T}A_{22})x_2 \leq c_1^\mathsf{T}x_1 + c_2^\mathsf{T}x_2 = c^\mathsf{T}x. \end{aligned}$$

Für Punkte $\hat{x} \in X$ und $\hat{y} \in Y$ mit $c^\mathsf{T}\hat{x} = b^\mathsf{T}\hat{y}$ ist dann insbesondere auch $c^\mathsf{T}x \geq b^\mathsf{T}\hat{y} = c^\mathsf{T}\hat{x}\,\forall x \in X$ und $b^\mathsf{T}y \leq c^\mathsf{T}\hat{x} = b^\mathsf{T}\hat{y}\,\forall y \in Y$, also $\hat{x}, \hat{y}$ extremal. ∎

Anwendung Bei Kenntnis von zulässigen Punkten $\hat{x}, \hat{y}$ ist die Prüfung auf Optimalität, "$c^\mathsf{T}\hat{x} = b^\mathsf{T}\hat{y}$?", trivial (z.B., für Auftraggeber). Und trivialerweise erhält man mit jedem dual zulässige y aus $b^\mathsf{T}y$ eine untere Schranke für den Optimalwert bei (LP).

Einzelne Eigenschaften der Programme haben eine bestimmte Bedeutung für das dazu duale. Es sei daran erinnert, dass mit der *Lösung* eines Programms eine Optimallösung gemeint ist. Ein Problem mit nichtleerem zulässigem Bereich nennt man *konsistent*, ansonsten *inkonsistent*. Die folgenden Sätze werden jeweils nur für dasjenige Standardprogramm (LPi) bewiesen, dessen Form sich dazu anbietet. Sie gelten aber natürlich für (LP). In den folgenden Beweisen spielt das Lemma von Farkas eine zentrale Rolle.

Satz 4.1.2 *Die Probleme (LP) und (LP*) seien beide konsistent. Dann existieren auch Lösungen für beide Programme.*

Beweis Der Nachweis erfolgt für das symmetrisch aufgebaute Programm (LP2). Mit dem Satz von Farkas, (3.6.1) ist die Voraussetzung $X \neq \emptyset$, $Y \neq \emptyset$ äquivalent mit

$$\begin{cases} \forall u \geq 0 \text{ mit} \quad u^\mathsf{T}A \leq 0 \quad \Rightarrow\ u^\mathsf{T}b \leq 0, \\ \forall v \geq 0 \text{ mit} \quad -Av \leq 0 \quad \Rightarrow\ -v^\mathsf{T}c \leq 0. \end{cases} \tag{4.1.1}$$

Und mit Satz 4.1.1 entspricht die Behauptung der Lösbarkeit des Systems

$$\begin{pmatrix} A & 0 \\ 0 & -A^\mathsf{T} \\ -c^\mathsf{T} & b^\mathsf{T} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \geq \begin{pmatrix} b \\ -c \\ 0 \end{pmatrix}, \ \begin{pmatrix} x \\ y \end{pmatrix} \geq 0. \tag{4.1.2}$$

Man beachte, dass dabei in der letzten Zeile wegen Satz 4.1.1 nur Gleichheit in Frage kommt. Nach Farkas, (3.6.1) ist diese Lösbarkeit äquivalent mit

$$\forall u \geq 0, v \geq 0, \lambda \geq 0 \text{ mit } \left\{ \begin{array}{c} u^\mathsf{T} A \leq \lambda c^\mathsf{T} \\ Av \geq \lambda b \end{array} \right\} \Rightarrow \ u^\mathsf{T} b \leq v^\mathsf{T} c. \tag{4.1.3}$$

Wenn dabei $\lambda = 0$ ist, entspricht dies der Voraussetzung (4.1.1), ist deshalb erfüllt und zeigt die Lösbarkeit. Im Fall $\lambda > 0$ kann man aber die Folgerung von (4.1.3) direkt aus den Prämissen von (4.1.3) schließen, indem man diese mit $v, u \geq 0$ multipliziert:

$$\underbrace{\lambda}_{>0} c^\mathsf{T} v \geq (u^\mathsf{T} A)v = u^\mathsf{T}(Av) \geq \underbrace{\lambda}_{>0} u^\mathsf{T} b.$$

Also gilt auch im Fall $\lambda > 0$: $u^\mathsf{T} b \leq v^\mathsf{T} c$ in (4.1.3) und zeigt die Lösbarkeit von (4.1.2). ∎

Man beachte, dass in (4.1.2) die Lösung von (LP) und (LP*) auf ein reines Ungleichungssystem zurückgeführt wurde.

Der folgende Satz nutzt die Tatsache aus, dass in einer Lösung von Problem (LP1) nur ein Teil der Restriktionen *straff* sind, vgl. Satz 3.4.3. Im Beweis wird ein Zusammenhang zwischen den Lösungen von Primal- und Dual-Problem konstruiert, der weitergehende Bedeutung hat.

Satz 4.1.3 *Es sei $\hat{x} \in \mathbb{R}^n$ eine Lösung von (LP1) und $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$.*

a) Mit $L \subseteq \{1, \ldots, m\}$, $K = \{1, \ldots, m\} \setminus L$ gelte dabei

$$A^{(L)}\hat{x} = b_L, \quad A^{(K)}\hat{x} > b_K.$$

Dann ist $\hat{x}$ auch Lösung des reduzierten Programms $\min\{c^\mathsf{T} x : \ A^{(L)}x \geq b_L\}$.

b) Dann hat das duale Programm (LP1) eine Lösung.*

Beweis a) Da im reduzierten Programm weniger Restriktionen gelten, hat es keinen größeren Wert als (LP1). Nun sei angenommen, es besitze eine Lösung $\bar{x}$ mit Wert $c^\mathsf{T}\bar{x} < c^\mathsf{T}\hat{x}$. Damit werden die Punkte $x(\lambda) := \lambda\bar{x} + (1-\lambda)\hat{x} = \hat{x} + \lambda(\bar{x} - \hat{x})$, $\lambda \in [0, 1]$, betrachtet. Diese erfüllen die L-Restriktionen, denn

$$A^{(L)}x(\lambda) = \lambda A^{(L)}\bar{x} + (1-\lambda)A^{(L)}\hat{x} \geq (\lambda + 1 - \lambda)b_L = b_L.$$

Wegen des Spielraums in den lockeren Restriktionen gibt es aber ein $\epsilon > 0$ so, dass auch noch

$$A^{(K)}x(\epsilon) = A^{(K)}\hat{x} + \epsilon A^{(K)}(\bar{x} - \hat{x}) \geq b_K$$

gilt, also ist $x(\epsilon)$ zulässig bei (LP1). Nach Annahme ist dort aber die Zielfunktion

$$c^\mathsf{T} x(\epsilon) = c^\mathsf{T}\hat{x} + \epsilon \underbrace{(c^\mathsf{T}\bar{x} - c^\mathsf{T}\hat{x})}_{<0} < c^\mathsf{T}\hat{x}$$

echt kleiner und widerspricht der Voraussetzung über $\hat{x}$.

b) Für ein beliebiges zulässiges Element x des reduzierten Programms gilt nach Teil a) $A^{(L)}x \geq b_L = A^{(L)}\hat{x}$ und $c^\mathsf{T}x \geq c^\mathsf{T}\hat{x}$, also die Folgerung

$$A^{(L)}(\hat{x} - x) \leq 0 \;\Rightarrow\; c^\mathsf{T}(\hat{x} - x) \leq 0 \quad \forall x \in \mathbb{R}^n.$$

Nach Satz 3.6.3 (Farkas) ist daher die Menge $Y_L := \{y_L : y_L^\mathsf{T}A^{(L)} = c^\mathsf{T},\, y_L \geq 0\} \neq \emptyset$. Daraus folgt aber sofort, dass der zulässige Bereich $Y := \{y \in \mathbb{R}^m : y^\mathsf{T}A = c^\mathsf{T}, y \geq 0\}$ von (LP1*) ebenfalls nicht leer ist. Denn mit $y_L \in Y_L$ liegt $y^\mathsf{T} := (y_L^\mathsf{T}, y_K^\mathsf{T})$, $y_K := 0_K$ in Y, es gilt

$$y^\mathsf{T}A = y_L^\mathsf{T}A^{(L)} + 0_K^\mathsf{T}A^{(K)} = c^\mathsf{T}, \quad \text{sowie} \quad y^\mathsf{T}b = y_L^\mathsf{T}b_L + 0_K^\mathsf{T}b_K = y_L^\mathsf{T}A^{(L)}\hat{x} = c^\mathsf{T}\hat{x}. \tag{4.1.4}$$

Da die Zielfunktionen gleiche Werte haben, ist nach Satz 4.1.1 jedes solche y optimal bei (LP1*). ■

Im Beweis wurde also mit den straffen Restriktionen eine duale Lösung konstruiert. Wenn die zugehörige Untermatrix $A^{(L)}$ maximalen Rang hat, besteht Y_L aus genau einem Punkt y_L, der durch Nullen zu einer Lösung $y^\mathsf{T} = (y_L^\mathsf{T}, 0_K^\mathsf{T})$ von (LP1*) ergänzt werden kann.

Theorem 4.1.4 (Dualitätssatz)

Das Lineare Programm (LP) ist genau dann lösbar, wenn (LP) lösbar ist.*

Beweis Der Beweis wird bei (LP1) geführt, im Satz 4.1.3 wurde dazu schon die Lösbarkeit von (LP1*) bei Lösbarkeit von (LP1) gezeigt. Umgekehrt sei (LP1*) lösbar, also auch das äquivalente Programm

$$\min(-b^\mathsf{T}y) : \begin{pmatrix} A^\mathsf{T} \\ -A^\mathsf{T} \\ I \end{pmatrix} y \geq \begin{pmatrix} c \\ -c \\ 0 \end{pmatrix} =: d.$$

Dieses hat die Standardform (LP1) und nach Satz 4.1.3 ist dann dessen Dual auch lösbar, also existiert $\hat{z}^\mathsf{T} = (z_-^\mathsf{T}, z_+^\mathsf{T}, u^\mathsf{T}) \geq 0^\mathsf{T}$ mit

$$\max d^\mathsf{T}z = d^\mathsf{T}\hat{z} = c^\mathsf{T}(z_- - z_+) \text{ mit } \; -b^\mathsf{T} = (z_-^\mathsf{T}, z_+^\mathsf{T}, u^\mathsf{T}) \begin{pmatrix} A^\mathsf{T} \\ -A^\mathsf{T} \\ I \end{pmatrix} = z_-^\mathsf{T}A^\mathsf{T} - z_+^\mathsf{T}A^\mathsf{T} + u^\mathsf{T}.$$

Der Vektor $\hat{x} := z_+ - z_- \in \mathbb{R}^n$ erfüllt also $Ax = A(z_+ - z_-) = b + u \geq b$ und ist Maximalstelle von $-c^\mathsf{T}x$, also Lösung von (LP1). ■

Wenn beide Probleme inkonsistent sind, ist die Situation klar. Andernfalls gilt:

Satz 4.1.5 *Wenn nur eines der Programme (LP) oder (LP*) zulässige Punkte hat, dann ist dessen Zielfunktion unbeschränkt.*

Beweis Ist (LP1*) inkonsistent, also $\{y : A^\mathsf{T}y = c, y \geq 0\} = \emptyset$, gibt es aufgrund der Farkas-Alternative in §3.6 ein $u \in \mathbb{R}^n$ mit

$$-u^\mathsf{T}A^\mathsf{T} \leq 0^\mathsf{T} \text{ und } (-u^\mathsf{T})c = 1 \iff Au \geq 0,\; c^\mathsf{T}u = -1.$$

Dann ist $u \in O^+(X)$ und mit beliebigem zulässigem x ist auch $x + tu$, $t \geq 0$, zulässig: $A(x + tu) = Ax + tAu \geq Ax \geq b$. Die Zielfunktion aber ist unbeschränkt, $c^\mathsf{T}(x+tu) = c^\mathsf{T}x - t \to -\infty\,(t \to \infty)$. ■

Die Beschränktheit von (LP1) wurde schon am Ende von §3.4 behandelt, dort wurde das Kriterium $c \in A^\mathsf{T}\mathbb{R}^m_+$ über Polarkegel hergeleitet. Es entspricht gerade der Lösbarkeit des Systems $A^\mathsf{T}y = c$, $y \geq 0$.

Insgesamt ergibt sich folgende Situation:

Zusammenfassung	(LP) hat zulässige Punkte	(LP) inkonsistent
(LP*) hat zulässige Punkte	(LP) und (LP*) lösbar	(LP*) unbeschränkt
(LP*) inkonsistent	(LP) unbeschränkt	keine Lösungen

4.2 Komplementarität

Zur Vorbereitung des Dualitätssatzes wurde in Satz 4.1.3 i.w. die Konstruktion einer dualen Optimallösung aus der primalen durchgeführt. Ansatzpunkt war die Erkenntnis, dass in Optimallösungen bestimmte Restriktionen *straff* sind, d.h., Gleichheit gilt. Eine analoge Formulierung bzw. Schlußweise verwendet dazu die folgende *strukturelle Orthogonalität* bei nicht-negativen Vektoren:

$$u, v \geq 0,\ u^\mathsf{T}v = 0 \ \Rightarrow\ \forall i:\ \Big\{u_i = 0 \text{ oder } v_i = 0\Big\}$$

Satz 4.2.1 (Komplementarität)

a) Es sei x zulässig für (LP1), y für (LP1). Beide Punkte sind genau dann optimal, wenn gilt*

$$y^\mathsf{T}(Ax - b) = 0, \quad \textit{d.h., für } i = 1, \ldots, m: \begin{cases} y_i > 0 \Rightarrow {a^{(i)}}^\mathsf{T} x = b_i \\ {a^{(i)}}^\mathsf{T} x > b_i \Rightarrow y_i = 0 \end{cases}.$$

b) Es sei x zulässig für (LP) und y für (LP). Beide Punkte sind genau dann optimal, wenn gilt*

$$y^\mathsf{T}(Ax - b) = 0 \quad \textit{und} \quad (c^\mathsf{T} - y^\mathsf{T}A)x = 0. \tag{4.2.1}$$

Beweis Für zulässige x, y gilt beim allgemeinen Problem

$$\begin{aligned} y^\mathsf{T}(Ax - b) &= y_1^\mathsf{T}(A_{11}x_1 + A_{12}x_2 - b_1) + y_2^\mathsf{T}(A_{21}x_1 + A_{22}x_2 - b_2) \ \geq 0 \\ (c^\mathsf{T} - y^\mathsf{T}A)x &= (c_1^\mathsf{T} - y_1^\mathsf{T}A_{11} - y_2^\mathsf{T}A_{21})x_1 + (c_2^\mathsf{T} - y_1^\mathsf{T}A_{12} - y_2^\mathsf{T}A_{22})x_2 \geq 0 \end{aligned}$$

Addition der beiden Formeln liefert

$$0 \leq y^\mathsf{T}(Ax - b) + (c^\mathsf{T} - y^\mathsf{T}A)x = y^\mathsf{T}Ax - y^\mathsf{T}b + c^\mathsf{T}x - y^\mathsf{T}Ax = c^\mathsf{T}x - y^\mathsf{T}b,$$

und die Differenz verschwindet nach Satz 4.1.1 genau dann, wenn x und y optimal sind. ■

Anmerkung: In Teil b) des Satzes wurde zur einfacheren Darstellung eine etwas verkürzte Schreibweise gewählt. Die Anteile der Gleichungsrestriktionen an den Innenprodukten verschwinden von vorneherein. In den restlichen bedeutet (4.2.1) ausführlich

$$y_1^\mathsf{T}(A_{11}x_1 + A_{12}x_2 - b_1) = 0, \quad (c_1^\mathsf{T} - y_1^\mathsf{T}A_{11} - y_2^\mathsf{T}A_{21})x_1 = 0.$$

Damit markieren die nichtverschwindenen Komponenten von y_1 die straffen Restriktionen von (LP) und die nichttrivialen bei x_1 die straffen bei (LP*).

Man redet im Zusammenhang mit Satz 4.2.1 auch von *komplementärem Schlupf*. Denn die Ungleichungen in (LP) und (LP*) können durch Einführung von Schlupfvariablen $u_1 \geq 0, v_1 \geq 0$ zu Gleichungsrestriktionen gemacht werden, $A_{11}x_1 + A_{12}x_2 - u_1 = b_1$, $A_{11}^{\mathsf{T}}y_1 + A_{21}^{\mathsf{T}}y_2 + v_1 = c_1$. Damit entspricht die Bedingung (4.2.1) einfach der Aussage

$$y_1^{\mathsf{T}}u_1 = 0, \quad v_1^{\mathsf{T}}x_1 = 0,$$

dass je Komponente die Schlupfvariable im $\left\{\begin{smallmatrix}\text{primalen}\\ \text{dualen}\end{smallmatrix}\right.$ Problem oder die Variable im $\left\{\begin{smallmatrix}\text{dualen}\\ \text{primalen}\end{smallmatrix}\right.$ Problem verschwindet.

Schattenpreise: Außer den Existenzaussagen zu Lösungen können aus dem dualen Problem auch *quantitative* Angaben zum Primalproblem abgeleitet werden. Die Größe b enthält in (LP1) die unteren Grenzen für die einzelnen Restriktionen (Resourcen bei Produktionsplanung in §1.2), die einer Verringerung der Kosten $c^{\mathsf{T}}x$ im Wege stehen. In einem Lösungs-Paar $\hat{x}, \hat{y}$ wird die Aufteilung der Restriktionen wie in Satz 4.1.3 benutzt,

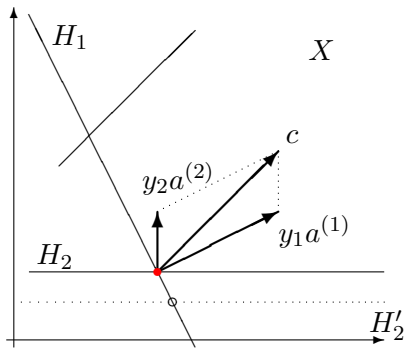
$$A^{(L)}\hat{x} = b_L, \; A^{(K)}\hat{x} > b_K, \quad L \cup K = \{1, \ldots, m\}.$$

Die Restriktionen zu L sind also straff, die zu K locker und aus dem Komplementaritätssatz folgt $\hat{y}_K = 0$. Für die Zielfunktion gilt damit $W := c^{\mathsf{T}}\hat{x} = b^{\mathsf{T}}\hat{y} = b_L^{\mathsf{T}}\hat{y}_L$. Für eine Verringerung der Kosten ist es sicher nicht sinnvoll, lockere Restriktionen aus K weiter zu lockern. In dem dualen Wert $b^{\mathsf{T}}\hat{y}$ kommt das dadurch zum Ausdruck, dass eine Verkleinerung von b_K wegen $\hat{y}_K = 0$ keine Auswirkung hätte. Dagegen stellen die straffen Restriktionen aus L *Flaschenhälse* dar. Bei einer kleinen Verringerung $b_L \to b_L - \bar{b}_L \leq b_L$ ($\|\bar{b}_L\| \leq \epsilon$) bleibt die zugehörige Lösung $\hat{x} - \bar{x}$ in der Regel (z.B., im generischen Fall $|L| = n$, $A^{(L)}$ regulär) weiterhin zulässig mit $A^{(K)}(\hat{x} - \bar{x}) \geq b_K$, und die Zielfunktion verändert sich gemäß

$$c^{\mathsf{T}}(\hat{x} - \bar{x}) = (b_L - \bar{b}_L)^{\mathsf{T}}\hat{y}_L = W - \bar{b}_L^{\mathsf{T}}\hat{y}_L. \tag{4.2.2}$$

Also gibt die Komponente $\hat{y}_i$ für $i \in L$ an, welche direkte Auswirkung eine Verkleinerung der Schranke b_i auf den Zielwert hätte.

Geometrische Interpretation Die nichttrivialen Werte $\hat{y}_L$ der dualen Variablen erfüllen die Bedingungen $\hat{y}_L^{\mathsf{T}}A^{(L)} = c^{\mathsf{T}}$, $\hat{y}_L \geq 0$. Geometrisch bedeutet das, dass der Zielgradient c konische Kombination der L-Zeilen von A ist, also in dem davon erzeugten Kegel liegt, $c \in \text{keg}\{a^{(j)} : j \in L\}$. Dies ist auch geometrisch klar, denn da die $a^{(j)}$ die nach innen (!) zeigenden Normalen auf den Randflächen H_j des Polyeders X sind, würde andernfalls das Minimum überhaupt nicht in $\hat{x}$ (roter Punkt) angenommen. Verringert



man im Bild ($J = \{1, 2\}$) den Wert b_2 etwas, entspricht die neue Nebenbedingung der gestrichelten Ebene H_2' und der Optimalpunkt bewegt sich mit (offener Kreis). Der Wert $c^\mathsf{T} x$ ändert sich aber nicht im gleichen Ausmaß, nur proportional zu y_2, da $a^{(2)}$ im Bild nur einen kleineren Anteil an c hat.

Ökonomische Interpretation Man nennt die Komponenten $\hat{y}_i$ der dualen Variablen auch *Schattenpreise*, da ihr Wert angibt, bei welchem Preis sich für den Nutzer eine Verkleinerung von b_i lohnt, da die Änderung der Kostenfunktion $c^\mathsf{T} x$ nach (4.2.2) gerade $-\hat{y}_i$ multipliziert mit der Änderung $\bar{b}_i$ ist. Diese Interpretation läßt sich anhand der Beispiele aus §1.2 erläutern.

Beispiel 4.2.2 Die *Produktionsplanung* ist ein Maximierungsproblem, wobei c_j der Gewinn für das Produkt P_j und b_i der Umfang der begrenzten Resource R_i ist. Mit einer Lösung y des dualen Programms

$$\min b^\mathsf{T} y, \quad \sum_{i=1}^{m} y_i a_{ij} \geq c_j, \ j = 1, \ldots, n, \ y \geq 0,$$

kann y_i als innerer oder Schattenpreis der Resource R_i interpretiert werden. Nach der Vorüberlegung darf die (Vergrößerung der) Resource R_i höchstens diesen Preis y_i kosten, damit beim Verkauf ein Zugewinn bleibt. Das duale Programm bestimmt diese Preise so, dass der *innere Gesamtpreis* der verwendeten Resourcen $\sum_i b_i y_i = c^\mathsf{T} x$ beim Verkauf der Produkte (x_j) exakt erzielt wird. Dabei unterschreitet der innere Einzelpreis $\sum_i y_i a_{ij}$ von Produkt P_j nicht den beim Verkauf erzielten äußeren Preis c_j. Die Folgerungen des Komplementaritätssatzes

$$\Big\{ \sum_{j=1}^{n} a_{ij} x_j < b_i \Rightarrow y_i = 0 \Big\}, \qquad \Big\{ \sum_{i=1}^{m} y_i a_{ij} > c_j \Rightarrow x_j = 0 \Big\}$$

können so interpretiert werden:

- Eine Resource, die nicht ausgeschöpft wird, ist im Überfluß vorhanden und bekommt den inneren Preis null.
- Ein Produkt, dessen innerer Preis höher als der außen erzielbare ist, wird nicht hergestellt.

Beispiel 4.2.3 Beim *Transportproblem* aus §1.2 war s_i die Kapazität von Produzent P_i und r_j der Bedarf von Abnehmer V_j. Für die Formulierung mit Ungleichungen $\min\{\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} : \sum_{j=1}^n x_{ij} \leq s_i, \sum_{i=1}^m x_{ij} \geq r_j, x_{ij} \geq 0\}$ hat das duale Problem die Form

$$\max \Big(\sum_{j=1}^{n} v_j r_j - \sum_{i=1}^{m} u_i s_i \Big) : \ v_j - u_i \leq c_{ij}, \ u_i, v_j \geq 0.$$

Interpretiert man u_i als Herstellungspreis bei P_i und v_j als Abnahmepreis bei V_j, bedeutet diese Form, dass zwar der Gesamtgewinn $\sum v_j r_j - \sum u_i s_i$ maximiert wird, aber die Gewinnspannen $v_j - u_i$ im Einzelfall nicht über den Transportkosten c_{ij} liegen.

5 Dualität beim Simplexverfahren

Die Dualitätsaussagen aus §4 liefern wichtige Hintergrundinformation zu den Eigenschaften eines linearen Programms. Tatsächlich kann zwischen den Daten des Simplexverfahrens zum Primalproblem (LP3) und dessen Dualprogramm (LP3*) ein direkter Zusammenhang hergestellt werden, der zusätzliche Möglichkeiten bei der Implementierung von Simplexverfahren eröffnet. Bei (LP3) sind die beiden Programme

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

zueinander dual. Im Simplexverfahren aus §2.4 wird ein Hilfsvektor $y^\mathsf{T} = c_J^\mathsf{T}A_J^{-1}$ berechnet. Wenn A_J Basis zu einer (Optimal-) Lösung $\hat{x}$ ist, gilt damit für den Vektor γ der reduzierten Kosten die Ungleichung

$$0 \leq \gamma^\mathsf{T} = c^\mathsf{T} - c_J^\mathsf{T}A_J^{-1}A = c^\mathsf{T} - y^\mathsf{T}A, \quad \text{d.h.} \quad y^\mathsf{T}A \leq c^\mathsf{T}. \tag{5.0.1}$$

Also ist dieser Vektor y eine dual zulässige Lösung. Wenn man die Lagrangefunktion $\phi = c^\mathsf{T}x + y^\mathsf{T}(b - Ax)$ aus der Einleitung von §4.1 betrachtet, ist der Kostenvektor gerade deren Gradient bezgl. x, $\gamma^\mathsf{T} = \nabla_x\phi(x, y) = c^\mathsf{T} - y^\mathsf{T}A$. Wegen $\gamma_J = 0$ sind die J-Ungleichungen straff, $y^\mathsf{T}A_J = c_J^\mathsf{T}$, was genau der Aussage des Komplementaritätssatzes $0 = (y^\mathsf{T}A - c^\mathsf{T})\hat{x} = 0$ entspricht. Damit stimmen auch die Zielfunktionen $y^\mathsf{T}b = c_J^\mathsf{T}A_J^{-1}b = c^\mathsf{T}\hat{x}$ überein und der Vektor y ist daher sogar (Optimal-) Lösung von (LP3*).

5.1 Duales Simplexverfahren

Vollkommen unabhängig von der Zulässigkeit des primalen Vektors $A_J^{-1}b$ gehört zu jeder Basis, die (5.0.1) erfüllt, ein dual zulässiger Vektor y.

Definition 5.1.1 *Eine Basis A_J heißt* dual zulässig*, wenn (5.0.1) gilt mit $y^\mathsf{T} = c_J^\mathsf{T}A_J^{-1}$, sie heißt* primal zulässig*, wenn $\bar{x}_J = A_J^{-1}b \geq 0$, und* optimal*, wenn sie primal und dual zulässig ist.*

Beim dualen Simplexverfahren arbeitet man mit den gleichen Basen A_J wie in §2.4, startet aber mit einer dual zulässigen Basis. In Bezug auf das Primal-Problem ist der zugehörige Vektor $\bar{x}_J = A_J^{-1}b$ zwar "optimal", aber i.A. nicht zulässig. Beim Basisaustausch werden daher negative Komponenten $\bar{x}_p < 0$ eliminiert.

Mit dieser Variante gewinnt man zusätzliche Wahlmöglichkeiten der Verfahrensgestaltung. Z.B. gehört beim Problem (LP2),

$$\min\{c^\mathsf{T}x : Ax - z = b, x \geq 0, z \geq 0\},$$

das hier durch Schlupfvariablen ergänzt wurde, mit $D = (A, -I)$ zu $J = \{n+1, \ldots, n+m\}$ die Basis D_J mit $c_J = 0$. Die Basislösungen sind $(\bar{x}, \bar{z}) = (0, -b)$ und $\bar{y} = 0$. Daher

$$\text{ist die Basis } D_J = -I_m \begin{cases} \text{primal} & \text{zulässig für } b \leq 0, \\ \text{dual} & \text{zulässig für } c \geq 0. \end{cases}$$

Im zweiten Fall läßt sich die Anlaufrechnung also durch Verwendung des jetzt entwickelten dualen Simplexverfahrens einsparen.

Zur Herleitung sei jetzt also A_J eine dual zulässige Basis mit

$$y^\mathsf{T} = c_J^\mathsf{T} A_J^{-1},\ \gamma^\mathsf{T} = c^\mathsf{T} - y^\mathsf{T} A \geq 0,\ \bar{x}_J = A_J^{-1} b, \quad K = \{1, \ldots, n\} \setminus J.$$

Ist nun $\bar{x}_p < 0$ für ein $p \in J$, so ist die duale Zielfunktion

$$y^\mathsf{T} b = c_J^\mathsf{T} A_J^{-1} b = c_J^\mathsf{T} \bar{x}_J$$

noch nicht maximal. Der negative "duale Schattenpreis" $x_p < 0$ zeigt an, dass durch eine virtuelle Verkleinerung von c_p, $p \in J$, eine Vergrößerung dieser Zielfunktion $y \mapsto y^\mathsf{T} b$ erfolgen kann. Unter Inkaufnahme zusätzlichen Schlupfs in der Ungleichung $c_p - y^\mathsf{T} a_p \geq 0$ betrachtet man analog zu (2.3.7) daher den Strahl

$$y(\lambda)^\mathsf{T} := (c - \lambda e_p)_J^\mathsf{T} A_J^{-1} = y^\mathsf{T} - \lambda (e_p)_J^\mathsf{T} A_J^{-1}, \quad \lambda \geq 0. \tag{5.1.1}$$

Für die duale Zielfunktion gilt dort tatsächlich

$$y(\lambda)^\mathsf{T} b = y^\mathsf{T} b - \lambda (e_p)_J^\mathsf{T} A_J^{-1} b = y^\mathsf{T} b - \lambda \bar{x}_p > y^\mathsf{T} b \quad \text{für } \lambda > 0.$$

Allerdings muß dabei, wieder analog zu (2.3.10), die duale Zulässigkeit von $y(\lambda)$ geprüft werden. Es ist zu fordern

$$0^\mathsf{T} \overset{!}{\leq} c^\mathsf{T} - y(\lambda)^\mathsf{T} A = c^\mathsf{T} - y^\mathsf{T} A + \lambda (e_p)_J^\mathsf{T} A_J^{-1} A = \gamma^\mathsf{T} + \lambda u_p^\mathsf{T}, \quad u_p^\mathsf{T} := (e_p)_J^\mathsf{T} A_J^{-1} A.$$

Wegen $\gamma_J = 0$ ist diese Bedingung für Indizes aus J automatisch erfüllt, $\gamma_J^\mathsf{T} + \lambda (e_p)_J^\mathsf{T} A_J^{-1} A_J = \lambda (\delta_{pj})_{j \in J} \geq 0^\mathsf{T}$. Auch ist für $u_p \geq 0$ zu erkennen, dass λ beliebig groß werden darf. In diesem Fall ist (LP3*) unbeschränkt und (LP3) inkonsistent, vgl. §4.1. Nur für negative Komponenten von $u_p = (u_{pj})_j$ ergeben sich Einschränkungen und führen zum maximal zulässigen Wert

$$\lambda_p := \min\{\frac{\gamma_j}{-u_{pj}} : u_{pj} < 0,\ j \in K\} = \frac{\gamma_\ell}{-u_{p\ell}}. \tag{5.1.2}$$

Wenn das Minimum, wie angegeben, im Index $\ell \in K$ angenommen wird, wird die entsprechende Ungleichung *straff*,

$$0 = \gamma_\ell + \lambda_p u_{p\ell} = c_\ell - y^\mathsf{T} a_\ell + \lambda_p (e_p)_J^\mathsf{T} A_J^{-1} a_\ell = c_\ell - y(\lambda_p)^\mathsf{T} a_\ell.$$

Der Index ℓ wandert also in die Stützmenge J der straffen Ungleichungen bei (LP3*), vgl. Satz 3.4.3. Umgekehrt ist für $\lambda_p > 0$ in der Ungleichung zu $p \in J$ nach Konstruktion das Gegenteil der Fall, $0 < c_p - y(\lambda_p)^\mathsf{T} a_p = \lambda_p$. Daher ist $y(\lambda_p)$ die duale Basislösung zur Basis

$$A_{J'}, \quad J' = J \setminus \{p\} \cup \{\ell\}.$$

Analog zu Satz 2.3.5 läßt sich zeigen, dass $A_{J'}$ wegen $u_{p\ell} < 0$ tatsächlich regulär ist. Die obigen Überlegungen werden zusammengefaßt zum folgenden Algorithmus:

Duales Simplex-Verfahren

Eingabe:	Dual zulässige Basis A_J, $J \subseteq \{1,\dots,n\}$	
Schritt 1	$x_J := A_J^{-1}b$, $y^\mathsf{T} := c_J^\mathsf{T} A_J^{-1}$, $K := \{1,\dots,n\} \setminus J$,	
2	suche $x_p < 0$ unter x_i, $i \in J$.	
3	wenn $x_i \geq 0\, \forall i \in J$:	**STOP**, Optimum!
4	$u_{pj} := (e_p)_J^\mathsf{T} A_J^{-1} a_j$, $j \in K$, wenn $u_{pj} \geq 0\, \forall j \in K$:	**STOP**, (LP3) inkonsistent!
5	$\gamma_j := c_j - y^\mathsf{T} a_j$, $j \in K$, suche $\ell \in K$: $-\gamma_\ell / u_{p\ell} = \min\{-\gamma_j/u_{pj} :\ u_{pj} < 0, j \in K\} = \lambda_p$	
6	$J := J \setminus \{p\} \cup \{\ell\}$, weiter mit 1	

Zur Durchführung sind wie beim Primalverfahren drei Gleichungssysteme zu lösen, etwa mit einer fortlaufend angepaßten LR-Zerlegung von A_J. Dies sind zunächst wieder die drei Systeme $A_J x_J = b$, $y^\mathsf{T} A_J = c_J^\mathsf{T}$, und $f^\mathsf{T} A_J = (e_p)_J^\mathsf{T}$. Der Aufwand dafür liegt wieder bei $O(m^2)$ einschließlich der LR-Anpassung. Dann sind folgende Innenprodukte zu berechnen

$$u_{pj} = f^\mathsf{T} a_j,\ j \in K, \quad \text{sowie } c_j - y^\mathsf{T} a_j, \text{ für } u_{p_j} < 0.$$

Hierfür sind zwischen $2m(n-m)$ und $4m(n-m)$ Operationen nötig, dieser Anteil ist also etwa doppelt so groß wie beim primalen Verfahren aus §2.4. Bei vorhandener Wahlmöglichkeit hat das Primalverfahren also einen Effizienzvorteil.

Beispiel 5.1.2 Für das Problem

$$\left.\begin{array}{rl} \min\ 2x_1 + x_2 + 3x_3 & \\ x_1 + x_2 + x_3 & \geq 1 \\ 2x_1 - x_2 + 2x_3 & \leq -2 \\ x_1 + 2x_2 - 2x_3 & \geq 1 \\ x_i & \geq 0 \end{array}\right\} \iff \left\{\begin{array}{rlll} \min\ 2x_1 + x_2 + 3x_3 & & & \\ -x_1 - x_2 - x_3 & +x_4 & = -1 \\ 2x_1 - x_2 + 2x_3 & +x_5 & = -2 \\ -x_1 - 2x_2 + 2x_3 & +x_6 & = -1 \\ & x_i & \geq 0 \end{array}\right.$$

gehört zu $J = \{4,5,6\}$ eine dual, aber nicht primal zulässige Basis. Das duale Simplexverfahren führt hier mit den folgenden Daten in 2 Schritten zum Ziel:

B-1 $\boxed{1.}$ $J = \{4,5,6\}$, $A_J = I$, $\bar{x}_J^\mathsf{T} = (-1,-2,-1)$, $y = 0$, $y^\mathsf{T} b = 0$.

$\boxed{2.}$ wähle $p = 4$, $u^\mathsf{T} = (e_4)_J^\mathsf{T} A_J^{-1} A = e_1^\mathsf{T} A = (-1,-1,-1,1,0,0)$, $(\gamma_1,\gamma_2,\gamma_3) = (2,1,3)$; $\lambda_p = \min\{2,1,3\} = 1$ angenommen in $\ell = 2$.

B-2 $\boxed{1.}$ $J = \{2,5,6\}$, $A_J^{-1} = A_J = \begin{pmatrix} -1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} = B$, $x_J = A_J^{-1} b = (1,-1,1)^\mathsf{T}$, $y^\mathsf{T} = (-1,0,0)$, $y^\mathsf{T} b = -b_1 = 1 = c^\mathsf{T} \bar{x}$.

$\boxed{2.}$ wähle $p = 5$, $u^\mathsf{T} = (e_5)_J^\mathsf{T} A_J^{-1} A = e_2^\mathsf{T} B A = (3,0,3,-1,1,0)$, $\gamma_4 = 1$, $\lambda_p = -\gamma_4/u_{p4} = 1$ mit $\ell = 4$.

B-3 $\boxed{1.}$ $J = \{2,4,6\}$, $A_J^{-1} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$, $\bar{x}_J = A_J^{-1} b = \underbrace{(2,1,3)^\mathsf{T}}_{>0}$ **optimal**, $y^\mathsf{T} b = c^\mathsf{T} \bar{x} = 2$.

Auch beim dualen Verfahren besteht die Gefahr des Kreisens, wenn das Minimum bei (5.1.2) nicht in einem einzigen Index ℓ angenommen wird. Diese Gefahr läßt sich auch hier wieder durch *kleinste Index*-Regeln ausschalten. Diese lauten in Schritt 2 und 5:

2	bestimme $p \in J: \quad p = \min\{i \in J : x_i < 0\}$
5	bestimme $\ell \in K: \quad \ell = \min\{j \in K : -\gamma_j/u_{pj} = \lambda_p\}$

5.2 Problem-Modifikationen

In die Formulierung praktischer Probleme gehen oft Daten ein, deren Wert nicht genau bekannt oder vorhersehbar ist (z.B., die Preis- oder Zinsentwicklung bei einer Produktions- oder Finanzplanung). Dann ist es klug, auch Varianten des Ausgangsproblems zu lösen ("was passiert, wenn der Euro über 1.40 Dollar steigt?"), etwa in Abhängigkeit von einem künstlichen Parameter $t \in \mathbb{R}$ ("parametrische Optimierung"). Oft will man auch unerwünschte Lösungen nachträglich durch weitere Restriktionen ausschließen, etwa nicht-ganzzahlige in der ganzzahligen Optimierung. In diesen Fällen kann man durch eine geschickte Kombination aus primalem und dualem Simplexverfahren eine bekannte Lösung dem veränderten Problem anpassen. Wir betrachten vier Situationen, Ausgangspunkt sei jeweils eine bekannte (Optimal-) Lösung $\hat{x}$ mit Basis A_J.

- Änderung der Zielfunktion c. Die Untersuchung einer parametrischen Änderung $c(t) = c + t\tilde{c}$, $t \geq 0$, (zur Vereinfachung) ist vorteilhaft, da Änderungen der Ausgangssituation dann schrittweise eintreten. Es sei daher

$$W(t) := \min\{(c + t\tilde{c})^\mathsf{T} x : \; Ax = b,\, x \geq 0\}.$$

  Die Lösung $\hat{x}$ zu $t = 0$ ist auch primal zulässig für $t \neq 0$. Der Kostenvektor ist allerdings

$$\gamma(t)^\mathsf{T} = c(t)^\mathsf{T} - c_J(t)^\mathsf{T} A_J^{-1} A = \gamma(0)^\mathsf{T} + t\tilde{\gamma}^\mathsf{T}, \quad \tilde{\gamma}^\mathsf{T} = \tilde{c}^\mathsf{T} - \tilde{c}_J^\mathsf{T} A_J^{-1} A.$$

  Da $\hat{x}$ optimal in $t = 0$ war, ist $\gamma(0) \geq 0$ und $\hat{x}$ bleibt solange optimal, wie

$$\gamma(t) = \gamma(0) + t\tilde{\gamma} \geq 0 \iff t \leq \min\{\frac{\gamma_j(0)}{-\tilde{\gamma}_j} : \; \tilde{\gamma}_j < 0, j \in K\} =: t_{\max},$$

  ($\gamma_J(t) \equiv 0$ gilt weiterhin). Wenn $t_{\max} > 0$ ist, ist $\hat{x}$ für $t \in [0, t_{\max}]$ optimal und daher $W(t) = W(0) + t\tilde{c}^\mathsf{T}\hat{x}$ dort linear (insgesamt ist $W(t)$ stückweise linear). Bei Vergrößerung von t über $t_{\max}$ hinaus verliert $\hat{x}$ seine Optimalität und im reduzierten Kostenvektor tauchen negative Komponenten auf. Ausgehend von der primal zulässigen Basis A_J kann mit dem *primalen* Verfahren aus §2.4 nachoptimiert werden.

- Änderung des (Resourcen-) Vektors $b(t) = b + t\tilde{b}$, wieder parametrisiert mit $t \geq 0$. Also sei

$$W(t) := \min\{c^\mathsf{T} x : \; Ax = b + t\tilde{b},\, x \geq 0\}.$$

  Dann löst $x(t)$ mit den Nichtbasisvariablen $x_K(t) = 0$ und der Basislösung

$$x_J(t) = A_J^{-1}(b + t\tilde{b}) = \hat{x}_J + t\xi_J, \quad \xi_J := A_J^{-1}\tilde{b},$$

immer noch das Gleichungssystem $Ax = b + t\tilde{b}$. Dabei ist $x(t)$ primal zulässig, solange

$$\hat{x}_J + t\xi_J \geq 0 \iff t \leq \min\{\frac{\hat{x}_i}{-\xi_i} : \xi_i < 0, i \in J\} =: t_{\max}.$$

Wenn $\hat{x}$ nicht ausgeartet ist, ist $t_{\max} > 0$ und die Zielfunktion $W(t) = W(0) + tc_J^\mathsf{T}\xi_J = W(0)+ty^\mathsf{T}\tilde{b}$ ($y =$ Schattenpreise!) im Intervall $[0, t_{\max}]$ also wieder linear. Der Kostenvektor γ ist hier unabhängig von t, da er nur von c und A abhängt. Wenn jetzt also t über $t_{\max}$ hinaus vergrößert wird, bleibt $x(t)$ immer noch dual zulässig, verliert aber seine primale Zulässigkeit. Ausgehend von der dual zulässigen Basis A_J kann jetzt mit dem *dualen* Simplexverfahren aus §5.1 nachoptimiert werden.

- Einführung zusätzlicher Ungleichungen, etwa ${a^{(m+1)}}^\mathsf{T} x \geq b_{m+1}$. Das Programm (LP3) wird also erweitert um die Gleichung ${a^{(m+1)}}^\mathsf{T} x - x_{n+1} = b_{m+1}$, $x_{n+1} \geq 0$, in der Zielfunktion ist $c_{n+1} = 0$. Mit der entsprechend erweiterten Matrix $\tilde{A}$ und $J' := J \cup \{n+1\}$ ist

$$\tilde{A}_{J'} = \begin{pmatrix} A_J & 0 \\ {a_J^{(m+1)}}^\mathsf{T} & -1 \end{pmatrix} \quad \Rightarrow \quad (\tilde{A}_{J'})^{-1} = \begin{pmatrix} A_J^{-1} & 0 \\ {a_J^{(m+1)}}^\mathsf{T} A_J^{-1} & -1 \end{pmatrix}. \tag{5.2.1}$$

Wegen $c_{n+1} = 0$ liefert die letzte Zeile keinen Beitrag zum erweiterten Kostenvektor $(c^\mathsf{T}, 0) - c_J^\mathsf{T} A_J^{-1}(A, 0) = (\gamma^\mathsf{T}, 0) \geq 0$ und der ergänzte Vektor $(\hat{x}^\mathsf{T}, \bar{x}_{n+1})^\mathsf{T}$ bleibt weiterhin dual zulässig, allerdings nicht mehr primal zulässig für

$$\bar{x}_{n+1} = {a^{(m+1)}}^\mathsf{T}\hat{x} - b_{m+1} < 0.$$

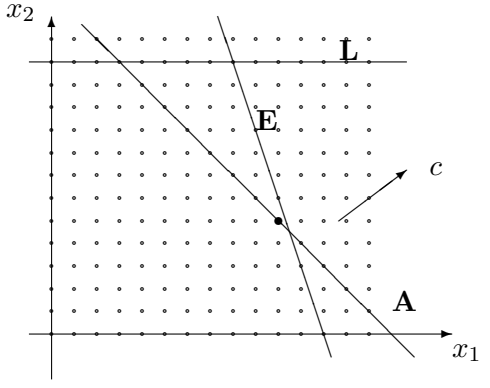
Dies ist also mit $p = n + 1$ wieder ein Fall für das *duale* Simplexverfahren aus §5.1.

Dieser Fall hat eine große Bedeutung in der ganzzahligen und nichtlinearen Optimierung. Dort werden lineare (Hilfs-) Programme gelöst und schrittweise unerwünschte Lösungen durch *Schnittebenen*, d.h. zusätzliche Ungleichungen eliminiert.

Beispiel 5.2.1 Im Einführungsbeispiel 1.2.1 zur Produktionsplanung

$$\begin{array}{lrlll} \min & -4x_1 - 3x_2 & & & \\ \mathbf{A}: & x_1 + x_2 & +x_3 & = 15, & \\ \mathbf{L}: & x_2 & +x_4 & = 12, & \\ \mathbf{E}: & 3x_1 + x_2 & +x_5 & = 36, & x_i \geq 0, \end{array}$$

wurde die Schranke für Resource **A** auf $b_1 = 15$ geändert, mit $J = \{1, 2, 4\}$ lautet die Lösung dann $\hat{x}_J^\mathsf{T} = (10.5, 4.5, 7.5)$, $W = -55.5$. Wenn nur ganze Einheiten produziert werden, ist diese Lösung unbrauchbar. Eine Rundung dieser Werte ist auch keine Hilfe, da die Zulässigkeit dann nicht gesichert ist. Mit Hilfe der zusätzlichen Ungleichung $2x_1 + x_2 \leq 25$ kann diese Ecke des zulässigen Bereichs abgeschnitten werden.



Die Konstruktion solcher Ungleichungen wird in der ganzzahligen Optimierung behandelt. Im erweiterten Problem ist jetzt $\hat{x}_6 = 25 - 2\hat{x}_1 - \hat{x}_2 = -1/2 < 0$, $J' = \{1,2,4,6\}$. Mit

$$A_J^{-1} = \frac{1}{2}\begin{pmatrix} -1 & 0 & 1 \\ 3 & 0 & -1 \\ -3 & 2 & 1 \end{pmatrix}$$

wird u_p zu $p = 6$ aus der letzten Zeile von (5.2.1) berechnet, wegen des Schlupfes $+x_6$ aber mit anderem Vorzeiechen $u_p^{\mathsf{T}} = \big(-{a_J^{(4)}}^{\mathsf{T}} A_J^{-1}, 1\big)\tilde{A} = (-\frac{1}{2}, 0, -\frac{1}{2}, 1)\tilde{A} = (0, 0, -\frac{1}{2}, 0, -\frac{1}{2}, 1)$. Der (alte) Kostenvektor ist $\gamma^{\mathsf{T}} = (0, 0, \frac{5}{2}, 0, \frac{1}{2})$ und führt auf $\lambda_6 = 1$ bei $\ell = 5$. Zu den neuen Basisindizes $J'' = \{1,2,4,5\}$ gehört die ganzzahlige Lösung $x^{\mathsf{T}} = (10,5,0,7,1,0)$ mit $W = -55$.

- Einführung einer zusätzlichen Variablen x_{n+1}. Es sei $\tilde{A} = (A, a_{n+1})$, $\tilde{c}^{\mathsf{T}} = (c^{\mathsf{T}}, c_{n+1})$. Der Vektor $(\hat{x}^{\mathsf{T}}, 0)$ ist dann auch primal zulässig beim erweiterten Problem. In Bezug auf Optimalität ist mit der dualen Lösung $y^{\mathsf{T}} = c_J^{\mathsf{T}} A_J^{-1}$ nur der Wert γ_{n+1} zu prüfen. Für $\gamma_{n+1} \geq 0$ bleibt der erweiterte Punkt opimal. Für

$$\gamma_{n+1} = c_{n+1} - y^{\mathsf{T}} a_{n+1} < 0$$

kann wieder das *primale* Verfahren aus §2.4 mit der primal zulässigen Basis $\tilde{A}_J = A_J$ angewendet werden.

Beispiel 5.2.2 Das Einführungsbeispiel 1.2.1 zur Produktionsplanung hatte die Form

$$\begin{array}{lrll} \min & -4x_1 - 3x_2 & & \\ \mathbf{A}: & x_1 + x_2 & +x_3 & = 16, \\ \mathbf{L}: & x_2 & +x_4 & = 12, \\ \mathbf{E}: & 3x_1 + x_2 & +x_5 & = 36,\ x_i \geq 0, \end{array}$$

und die Lösung $\hat{x}^{\mathsf{T}} = (10,6,0,6,0)$ zu $J = \{1,2,4\}$ mit $c^{\mathsf{T}}\hat{x} = -58$. Die Ungleichungen zu Arbeitsaufwand ($\hat{x}_3 = 0$) und Energiebedarf ($\hat{x}_5 = 0$) sind straff, die Schattenpreise der dualen Lösung $y^{\mathsf{T}} = (-\frac{5}{2}, 0, -\frac{1}{2})$ zeigen, dass der Wert verringert werden kann, wenn eine Erhöhung von Arbeitsleistung nicht mehr als $-y_1 = \frac{5}{2}$ bzw. der Energiekosten um mehr als $-y_3 = \frac{1}{2}$ pro Einheit kostet. Nun werde angenommen, dass zusätzliche Energie zu einem Preis von $c_6 > 0$ erhältlich ist. Am besten kauft man zusätzliche Energie nicht blind, sondern erweitert das Problem um den zusätzlichen Energieanteil $x_6 \geq 0$. Die geänderte Bedingung **E**: $3x_1 + x_2 \leq 36 + x_6$ führt zur Restriktion

$$\mathbf{E}:\ 3x_1 + x_2 + x_5 - x_6 = 36,\ \text{sowie } c^{\mathsf{T}}x = -4x_1 - 3x_2 + c_6 x_6.$$

Also ist $a_6 = -e_3$ und $\gamma_6 = c_6 - y^{\mathsf{T}} a_6 = c_6 - \frac{1}{2}$. Für $c_6 < \frac{1}{2}$ sind die Kosten γ_6 negativ. Die inverse Basismatrix ist die aus Beisp. 5.2.1. Ein Austauschschritt mit $\ell = 6$, $w_\ell^{(J)} = A_J^{-1} a_6 = -A_J^{-1} e_3 = \frac{1}{2}(-1,1,-1)^{\mathsf{T}}$ ergibt $p = j_2 = 2$ und führt zur neuen Lösung $(16,0,0,12,0,12)^{\mathsf{T}}$ mit $J' = \{1,5,6\}$ und Zielfunktionswert $-64 + 12c_6$ (< -58 für $c_6 < \frac{1}{2}$).

Praktischer Ausblick

Professionelle Computerprogramme ("Dynamische Simplex-Verfahren") bringen beim allgemeinen Problem (LP) beide Varianten des Simplexverfahrens adaptiv zum Einsatz, teilweise auch als Ersatz für eine Anlaufrechnung. Ansatzweise sei das am Programm (LP) ohne freie Variable erläutert, d.h. bei

$$\min\{c^\mathsf{T}x : \ Ax = b,\, Mx \geq d,\, x \geq 0\}. \tag{LP}$$

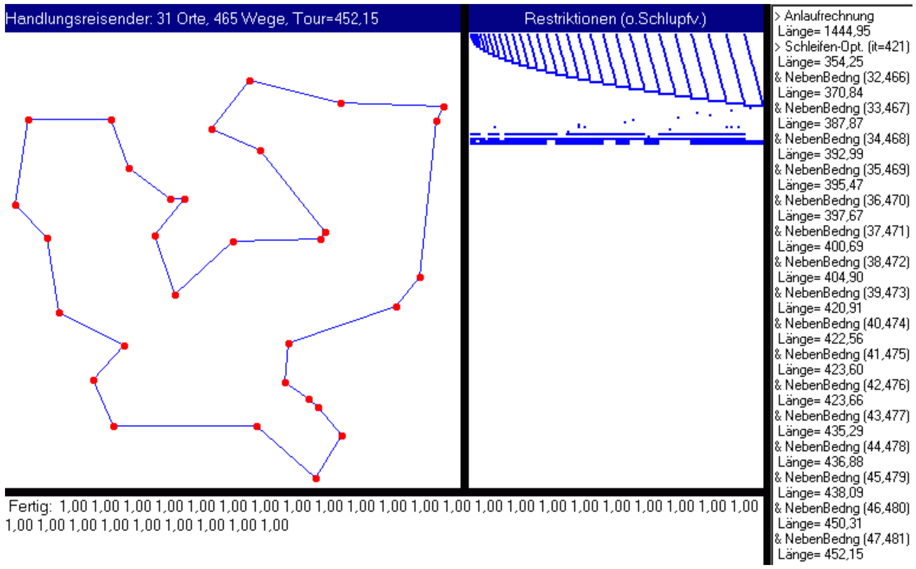
Dabei seien $A \in \mathbb{R}^{m\times n}$ und $M \in \mathbb{R}^{\mu\times n}$ sehr große Matrizen. Um dennoch mit annehmbaren Aufwand arbeiten zu können, betrachtet man Teilprobleme, in denen nur ein Teil der Variablen und ein Teil der Ungleichungen aktiviert ist ([Padberg]). Mit $P \subseteq \{1,\ldots,n\}$, $L \subseteq \{1,\ldots,\mu\}$ sind das Probleme der Form

$$\min\{c_P^\mathsf{T}x_P : \ A_Px_P = b,\, M_P^{(L)}x_P \geq d_L,\, x_P \geq 0\}, \tag{LP$_P^L$}$$

nur die Gleichungsrestriktionen werden also alle berücksichtigt. Schrittweise werden nun solche Teilprobleme gelöst und danach durch Suche nach negativen Kosten $\gamma_j < 0$ neue Variable mit Index $j \notin P$, oder verletzte Ungleichungen $\notin L$ aktiviert. Für die Einheitsvektoren zu den Schlupfvariablen der Ungleichungen $M_P^{(L)}x_P \geq d_L$ wird natürlich kein Speicherplatz reserviert, sie werden bei Bedarf erzeugt. Die Anpassung der Lösung der neuen Teilprobleme kann, wie gerade besprochen, mit dem primalen bzw. dualen Verfahren durchgeführt werden. Umgekehrt können Variable zu $j \in P$ (für $\gamma_j \gg 0$) bzw. Ungleichungen aus L auch wieder deaktiviert werden, wenn Kosten oder Schlupfvariable bestimmte Schwellenwerte unter- bzw. überschreiten. Sehr große Probleme können insbesondere dann so gelöst werden, wenn die Suche zur Aktivierung *algorithmisch* erfolgen kann. Dies ist z.B. bei Schnittebenenverfahren der Fall.

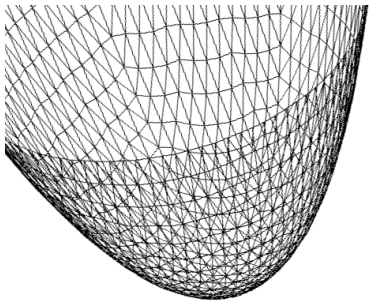
Ähnliches gilt beim TSP, wo die $\cong 2^n$ Ungleichungen (1.2.2) sicherstellen, dass die Tour zusammenhängend ist. Für eine vorliegende Näherungslösung x kann eine verletzte Ungleichung (1.2.2) graphentheoretisch durch Bestimmung eines sogenannten *minimalen Schnitts* generiert werden, was mit einem polynomiellen Aufwand geschehen kann. Oft sind Lösungen des relaxierten Problems (1.2.3) ganzzahlig, andernfalls müssen zusätzlich Schnittebenen eingeführt werden.

Beispiel 5.2.3 Anwendung der Verfahren auf das (TSP), Start mit den Gleichheitsrestriktionen (1.2.1). Diese Tour besteht i.d.R. aus vielen kleinen Schleifen. Anschließend wird jeweils eine kurze Schleife gesucht (kein minimaler Schnitt!) und eine Ungleichung (1.2.2), welche diese ausschließt, in (LP$_P^L$) aufgenommen. In einigen Fällen führt dies zum Erfolg, etwa im gezeigten Beispiel. Das Problem mit 31 Orten hat 465 Wege (d.h. $n = 465$ Variable, $m = 31$ Gleichungen). Anschließend werden 16 zusätzliche Ungleichungen (von $\cong 2^{31} \cong 10^{10}$ möglichen) generiert, bis eine zusammenhängende (und sogar ganzzahlige) Lösung erreicht ist. Das Bild zeigt die Tour und in der Mitte oben die Struktur der Matrix im Ungleichungssystem.


Handlungsreisender: 31 Orte, 465 Wege, Tour=452,15
Restriktionen (o.Schlupfv.)
> Anlaufrechnung
Länge= 1444,95
> Schleifen-Opt. (it=421)
Länge= 354,25
& NebenBedng (32,466)
Länge= 370,84
& NebenBedng (33,467)
Länge= 387,87
& NebenBedng (34,468)
Länge= 392,99
& NebenBedng (35,469)
Länge= 395,47
& NebenBedng (36,470)
Länge= 397,67
& NebenBedng (37,471)
Länge= 400,69
& NebenBedng (38,472)
Länge= 404,90
& NebenBedng (39,473)
Länge= 420,91
& NebenBedng (40,474)
Länge= 422,56
& NebenBedng (41,475)
Länge= 423,60
& NebenBedng (42,476)
Länge= 423,66
& NebenBedng (43,477)
Länge= 435,29
& NebenBedng (44,478)
Länge= 436,88
& NebenBedng (45,479)
Länge= 438,09
& NebenBedng (46,480)
Länge= 450,31
& NebenBedng (47,481)
Länge= 452,15
Fertig: 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00
1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00

6 Innere-Punkt-Methoden

Das Simplex-Verfahren startet mit einer Ecke des zulässigen Polyeders X und wandert dann zu Nachbar-Ecken mit fallender Zielfunktion. Insbesondere bewegt sich das Verfahren ausschließlich auf dem Rand des Polyeders. Obwohl das Verfahren in einer endlichen Zahl von Schritten endet, kann dies in einigen (Ausnahme-?) Fällen bei hohen Dimensionen wegen der großen Eckenzahl zu sehr langen Laufzeiten des Verfahrens führen. Ein alternativer Zugang sind neuere Verfahren, die eine Iterationsfolge konstruieren, welche sich durch das Innere des Polyeders auf die optimale Ecke zu bewegt.



6.1 Der zentrale Pfad

Betrachtet man mit dem primalen Programm, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$,

$$(LP3) \quad \min c^\mathsf{T} x : \ Ax = b, \ x \geq 0$$

gleichzeitig dessen duales $\max\{y^\mathsf{T} b : \ A^\mathsf{T} y \leq c\}$ und führt dabei Schlupfvariablen z ein,

$$(LP3^*) \quad \max y^\mathsf{T} b : \ A^\mathsf{T} y + z = c, \ z \geq 0,$$

dann kann man deren Lösung wegen des Komplementaritäts-Satzes 4.2.1,

$$\underbrace{(c^\mathsf{T} - y^\mathsf{T} A)}_{\geq 0} \underbrace{x}_{\geq 0} = z^\mathsf{T} x = 0$$

("komplementärer Schlupf") auch als ein reines Un-Gleichungssystem schreiben,

$$\begin{array}{rll} Ax & = b, & x \geq 0, \\ A^\mathsf{T} y + z & = c, & z \geq 0, \\ z^\mathsf{T} x & = 0. & \end{array} \tag{6.1.1}$$

Man beachte, dass dabei (nur) die letzte Bedingung nichtlinear ist und tatsächlich wegen der Nichtnegativität eine strukturelle Orthogonalität darstellt, $x_j z_j = 0$, $j = 1, \ldots, n$. Dies läßt sich mit Hilfe der Diagonalmatrizen

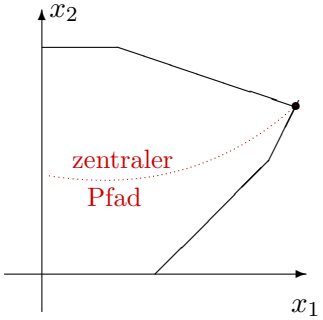
$$Z = \operatorname{diag}(z_j) = \begin{pmatrix} z_1 & & \\ & \ddots & \\ & & z_n \end{pmatrix}, \quad X = \operatorname{diag}(x_j) = \begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{pmatrix}$$

besser in der Form $Zx = Xz = 0$ zum Ausdruck bringen. Wenn man zwei beliebige zulässige Punkte $x^{(0)}$ von $(LP3)$ und $y^{(0)}$ von $(LP3^*)$ hat, gilt $0 \leq {z^{(0)}}^\mathsf{T} x^{(0)} =: n\mu_0$. Für $x^{(0)} > 0$ und $z^{(0)} > 0$ (komponentenweise) ist sogar ${z^{(0)}}^\mathsf{T} x^{(0)} = n\mu_0 > 0$.

Formuliert man die Bedingung $x^\mathsf{T} z = n\mu > 0$ für jede Komponente einheitlich zu $x_j^\mathsf{T} z_j = \mu$, $j = 1, \ldots, n$, bzw $Xz = \mu \mathbb{1}$ mit $\mathbb{1} = (1, \ldots, 1)^\mathsf{T}$ bekommt man folgendes Problem, das den Innere-Punkte-Verfahren zugrunde liegt:

$$F_\mu(x, y, z) := \begin{pmatrix} Ax - b \\ A^\mathsf{T} y + z - c \\ Xz - \mu \mathbb{1} \end{pmatrix} \stackrel{!}{=} 0, \quad x > 0, z > 0. \tag{6.1.2}$$

Da unter geeigneten Voraussetzungen zu jedem $\mu > 0$ eine Lösung existiert, bildet die Menge dieser Punkte einen stetigen(!) *zentralen Pfad* $\big(x(\mu), y(\mu), z(\mu)\big)$, der bei Variation des Parameters μ durchlaufen wird. Mit zulässigen Lösungen $x^{(0)}$, $y^{(0)}$ wie oben kennt man insbesondere einen (Start-) Punkt $\big(x(\mu_0), y(\mu_0), z(\mu_0)\big)$ mit $\mu_0 > 0$ und kann versuchen, diesen bis zum Ziel, der Lösung $\big(x(0), y(0), z(0)\big)$ zu verfolgen.



Das Problem (6.1.2) läßt sich durch folgende Umformungen anders interpretieren. Aus den beiden letzen Gleichungen $A^\mathsf{T} y + z - c = 0$, $x_j z_j = \mu \forall j$ eliminiert man z und x,

$$z_j = c_j - a_j^\mathsf{T} y, \quad x_j = \frac{\mu}{z_j} = \frac{\mu}{c_j - a_j^\mathsf{T} y}, \quad j = 1, \ldots, n, \tag{6.1.3}$$

und reduziert das Problem (6.1.2) dadurch auf einen Satz von nichtlinearen Gleichungen

$$g_i(y) := b_i - \mu \sum_{j=1}^{n} \frac{a_{ij}}{c_j - a_j^\mathsf{T} y} \stackrel{!}{=} 0, \quad i = 1, \ldots, m. \tag{6.1.4}$$

Diese Umformung ist sehr hilfreich, denn g ist der Gradient der folgenden Funktion.

Definition 6.1.1 *Die Menge der dual strikt zulässigen Punkte*

$$\hat{Y} := \{y \in \mathbb{R}^m : \; A^\mathsf{T} y < c\}$$

sei nicht leer. Für $\mu > 0$ wird dort die (duale) Barrierefunktion $b_\mu : \hat{Y} \mapsto \mathbb{R}^m$ *definiert durch*

$$b_\mu(y) := b^\mathsf{T} y + \mu \sum_{j=1}^{n} \log(c_j - a_j^\mathsf{T} y).$$

Der Name *Barrierefunktion* veranschaulicht die Gestalt von b_μ. Wenn sich y dem Rand von $\hat{Y}$ nähert, also $0 < c_j - a_j^\mathsf{T} y \to 0$ geht für ein j, geht der Summand $\mu \log(c_j - a_j^\mathsf{T} y) \to -\infty$ und baut eine unüberwindliche Barriere (Graben) am Rand auf. Im Inneren von $\hat{Y}$ ist b_μ aber beliebig oft differenzierbar. Zum Zusammenhang mit g gilt tatsächlich

$$\frac{\partial b_\mu}{\partial y_i} = b_i - \mu \sum_{j=1}^{n} \frac{a_{ij}}{c_j - a_j^\mathsf{T} y} = g_i(y), \quad i = 1, \ldots, m.$$

Die zweiten Ableitungen sind

$$\frac{\partial^2 b_\mu}{\partial y_i \partial y_k} = -\mu \sum_{j=1}^{n} \frac{a_{ij} a_{kj}}{(a_j^\mathsf{T} y - c_j)^2}, \quad 1 \le i, k \le m.$$

Durch Einführung der nicht-negativen Diagonalmatrix $N := \mathrm{diag}((a_j^\mathsf{T} y - c_j)^2) \ge 0$ läßt sich die Hesse-Matrix dieser 2. Ableitungen einfach darstellen als

$$H_\mu(y) = -\mu A N^{-1} A^\mathsf{T}. \tag{6.1.5}$$

Für $y \in \hat{Y}$ ist $(c_j - a_j^\mathsf{T} y)^2 > 0$, $j = 1, \ldots, m$ und daher N positiv definit, also ist $-\mu A N^{-1} A^\mathsf{T}$ negativ definit, wenn A vollen Rang besitzt. Daher ist die Funktion b_μ überall in $\hat{Y}$ streng konkav.

Satz 6.1.2 *Es gelte Rang$(A) = m$, die Menge $\hat{Y}$ sei nichtleer und beschränkt. Dann besitzt das Problem*

$$\max b_\mu(y) : \ y \in \hat{Y}$$

für jedes $\mu > 0$ genau eine Lösung $y(\mu)$, die mit (6.1.3) Komponente einer Lösung von (6.1.2), $F_\mu = 0$, ist. Diese vollständige Lösung $(x(\mu), y(\mu), z(\mu))$ heißt zentraler Pfad *des primal-dualen Problems (6.1.1).*

Beweis Nach Voraussetzung existiert ein $y^{(0)} \in \hat{Y}$. Da $\hat{Y}$ beschränkt ist n.V., ist die Niveaumenge $M := \{y \in \hat{Y} : \ b_\mu(y) \ge b_\mu(y^{(0)})\}$ kompakt und b_μ dort stetig. Also existiert eine Maximalstelle $\hat{y} \in \hat{Y}$. In dieser verschwindet der Gradient, $g(\hat{y}) = 0$. Nach dem Satz von Taylor gibt es daher zu jedem $y \ne \hat{y}$, $y \in \hat{Y}$ eine Zwischenstelle $\eta \in \hat{Y}$ so, dass

$$b_\mu(y) = b_\mu(\hat{y}) + \underbrace{g(\hat{y})^\mathsf{T}}_{=0}(y - \hat{y}) + \underbrace{(y - \hat{y})^\mathsf{T} H_\mu(\eta)(y - \hat{y})}_{<0} < b_\mu(\hat{y}).$$

Denn nach (6.1.5) ist tasächlich $(y - \hat{y})^\mathsf{T} H_\mu(\eta)(y - \hat{y}) = -\mu \sum_j (a_j^\mathsf{T}(y - \hat{y})/(c_j - a_j^\mathsf{T}\eta))^2 < 0$. Also ist $\hat{y}$ globale Maximalstelle von b_μ. ∎

Ausgehend von einem primal-dual zulässigen Paar, das als Startpunkt $(x(\mu_0), y(\mu_0), z(\mu_0))$ mit einem geeignet gewählten μ_0 dienen kann, kann man den zentralen Pfad schrittweise in Richtung auf die gesuchte Lösung $(x(0), y(0), z(0))$ verfolgen. Dazu löst man jeweils das nichtlineare Problem (6.1.2) an einer Stelle μ_k und verwendet den bekannten Wert an der Stelle μ_{k-1} als Ausgangspunkt. Eine Standardmethode zur Lösung nichtlinearer Gleichungssystem ist das Newtonverfahren

6.2 Newtonverfahren zur Pfadverfolgung

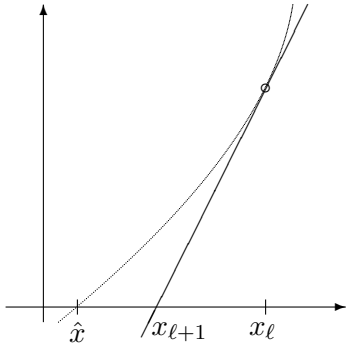
Für eine einzelne nichtlineare Gleichung $f(x) = 0$, $f : \mathbb{R} \to \mathbb{R}$, besteht das Newton-Verfahren darin, dass man die Funktion f durch ihre Tangente an einer Stelle x_ℓ ersetzt und deren Nullstelle als neue, bessere Näherung betrachtet. Dieses Vorgehen wederholt man iterativ:

$$f(z) = \underbrace{0 \stackrel{!}{=} f(x_\ell) + f'(x_\ell)(z - x_\ell)}_{\text{Tangente}} + \dots$$

Die Nullstelle der Tangente liefert das *Newton-Verfahren*

$$x_{\ell+1} = x_\ell - \frac{f(x_\ell)}{f'(x_\ell)}, \quad \ell = 0, 1, \dots, \tag{6.2.1}$$

dessen Durchführung offensichtlich eine nicht verschwindende Ableitung $f'(x_\ell) \neq 0$ erfordert. Wenn man mit einem genügend guten Startwert x_0 beginnt, ist die Konvergenz sehr schnell in einer Umgebung der Nullstelle $\hat{x}$.



Satz 6.2.1 *Die Funktion $f \in C^2[a,b]$ besitze eine Nullstelle $\hat{x} \in (a,b)$, es gelte $|f'(x)| \geq m_1 > 0$, $|f''(x)| \leq M_2$ $\forall x \in [a,b]$. Dann konvergiert das Newtonverfahren (6.2.1) für jeden Startwert $x_0 \in [a,b]$ mit $|x_0 - \hat{x}| < r := \min\{2m_1/M_2, b - \hat{x}, \hat{x} - a\}$. Dabei gilt $x_\ell \to \hat{x}$ $(\ell \to \infty)$ und die Konvergenz ist quadratisch, d.h.,*

$$|x_{\ell+1} - \hat{x}| \leq \frac{M_2}{2m_1} |x_\ell - \hat{x}|^2, \quad \ell = 0, 1, \dots.$$

Beweis Zunächst gilt für jedes $x_\ell \in [a,b]$ nach dem Satz von Taylor mit einer Zwischenstelle $\xi_\ell \in [a,b]$ die Identität

$$\begin{aligned} |x_{\ell+1} - \hat{x}| &= \Big| x_\ell - \hat{x} - \frac{f(x_\ell) - f(\hat{x})}{f'(x_\ell)} \Big| = \Big| \frac{f(x_\ell) - f(\hat{x}) - f'(x_\ell)(x_\ell - \hat{x})}{f'(x_\ell)} \Big| \\ &= \Big| \frac{f''(\xi_\ell)}{2f'(x_\ell)} \Big| |x_\ell - \hat{x}|^2 \leq \frac{M_2}{2m_1} |x_\ell - \hat{x}|^2. \end{aligned}$$

Dies ist die Ungleichung aus der Behauptung. Für $|x_0 - \hat{x}| < r$ ist dabei $\frac{M_2}{2m_1}|x_0 - \hat{x}| =: q < 1$ und damit folgt induktiv zunächst $|x_\ell - \hat{x}| < r$, $\ell \geq 0$, also auch $x_\ell \in (a,b)$ und daher die Konvergenz

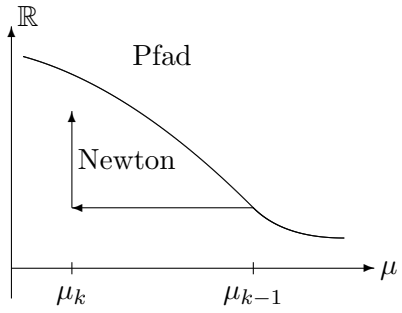
$$|x_\ell - \hat{x}| \leq q|x_{\ell-1} - \hat{x}| \leq \dots \leq q^\ell |x_0 - \hat{x}| \to 0 \, (\ell \to \infty). \qquad \blacksquare$$

Die Voraussetzungen des Satzes bedeuten insbesondere, dass f streng monoton ist in $[a,b]$. Es gibt zwei für den Einsatz des Verfahrens wesentliche Aspekte dieser Konvergenzaussage.

- *Quadratische Konvergenz* bedeutet, dass die Folge (x_ℓ) so schnell gegen $\hat{x}$ konvergiert, dass sich die Zahl der exakten Ziffern pro Schritt ungefähr verdoppelt. Denn etwa aus einem Fehler $|x_\ell - \hat{x}| \cong 10^{-4}$ wird $|x_{\ell+1} - \hat{x}| \cong (10^{-4})^2 = 10^{-8}$. Zum Beispiel bekommt man für die Nullstelle $\sqrt[3]{2}$ von $f(x) = x^3 - 2$ mit $x_0 = 1$ die gezeigte Folge. Die korrekten Ziffern sind unterstrichen.

ℓ	x_ℓ
0	1
1	1.333333333
2	1.263888888
3	1.2599334934499
4	1.2599210500177
5	1.25992104989487

- *Lokale Konvergenz*: Der Bereich günstiger Startwerte ist allerdings durch den Abstand r eingeschränkt, für kleines m_1 oder großes M_2 (zweite Ableitungen) kann diese Umgebung um die Nullstelle sehr klein sein. Dann ist die Konstruktion eines guten Startwerts x_0 der schwierigste Teil beim Einsatz des Newtonverfahrens.
  Gerade hier bietet die Pfad-Verfolgung einen einfachen Ausweg. Wenn das Problem glatt von einem Parameter $\mu \in \mathbb{R}$ abhängt, definiert $f_\mu(x) = 0$ einen ganzen Lösungs-Pfad $x(\mu)$. Kennt man für diesen eine Lösung $x(\mu_{k-1})$, so ist diese an einer benachbarten Stelle μ_k eine gute Näherung für $x(\mu_k)$, wenn die Änderung $|\mu_k - \mu_{k-1}|$ klein genug ist. Offensichtlich reicht es dann auch aus, den neuen Pfadpunkt $x(\mu_k)$ mit dem Newtonverfahren nicht mit maximaler Genauigkeit zu approximieren (z.B. nur ein/zwei Newtonschritte!).



Durch Dämpfung kann der Konvergenzbereich oft vergrößert werden. Dazu betrachtet man den Quotienten $s_\ell := -f(x_\ell)/f'(x_\ell)$ nur als eine Richtungsangabe für den nächsten Schritt, verwendet aber kürzere Schrittweiten, $0 < t_\ell \le 1$,

$$x_{\ell+1} := x_\ell + t_\ell s_\ell = x_\ell - t_\ell \frac{f(x_\ell)}{f'(x_\ell)}, \quad \ell = 0, 1, \ldots.$$

Durch Wahl von t_ℓ erhält man weitergehende Steuerungsmöglichkeiten.

Beispiel 6.2.2 Die Funktion $f(x) = x/\sqrt{1+x^2}$ hat die einzige Nullstelle $\hat{x} = 0$. Mit $f'(x) = (1+x^2)^{-3/2}$ folgt

$$x_{\ell+1} = x_\ell - t_\ell(x_\ell + x_\ell^3) = (1 - t_\ell)x_\ell - t_\ell x_\ell^3.$$

Beim einfachen Newtonverfahren ($t_\ell \equiv 1$) führt die Iteration $x_{\ell+1} = -x_\ell^3$ für $|x_0| < 1$ zu schneller Konvergenz und $|x_0| > 1$ zu Divergenz. Dagegen vergrößert sich für $t < 1$ der Bereich, wo Konvergenz auftritt auf $|x_0| < \sqrt{2/t - 1}$ auf Kosten der quadratischen Konvergenz.

Für nichtlineare *Systeme* von Gleichungen $f(x) = 0$, $f : \mathbb{R}^n \to \mathbb{R}^n$, läßt sich das Konstruktionsprinzip weiter anwenden. Iterationsvektoren werden jetzt wieder oben indiziert. Allerdings stellt die Bedingung des einfachen Newtonverfahrens

$$0 \stackrel{!}{=} f(x^{(\ell)}) + f'(x^{(\ell)})\big(x^{(\ell+1)} - x^{(\ell)}\big)$$

jetzt ein lineares Gleichungssystem für den Schritt $x^{(\ell+1)} - x^{(\ell)}$ dar, wobei die (Jacobi-) Matrix $J_\ell := f'(x^{(\ell)})$ der partiellen Ableitungen von f auch von der aktuellen Näherung abhängt. Ausführlicher lautet hier das gedämpfte *Newton-Verfahren* für Systeme

$$\begin{aligned} J_\ell s^{(\ell)} &:= -f(x^{(\ell)}), \quad \text{mit } J_\ell = f'(x^{(\ell)}) \\ x^{(\ell+1)} &:= x^{(\ell)} + t_\ell s^{(\ell)}. \end{aligned} \tag{6.2.2}$$

Für dieses Verfahren gelten für $t_\ell \equiv 1$ zu Satz 6.2.1 analoge Aussagen, die Konvergenz ist lokal quadratisch, wenn die Jacobi-Matrix regulär ist,

$$\|f'(x)^{-1}\| \le \frac{1}{m_1}$$

in einer hinreichend großen Umgebung der Nullstelle $\hat{x}$. Sehr interessant ist dabei die folgende Eigenschaft der Newtonrichtung $s^{(l)} = -J_\ell^{-1} f(x^{(\ell)})$. Man betrachtet die *Zielfunktion*

$$\varphi(x) := \|f(x)\|_2^2$$

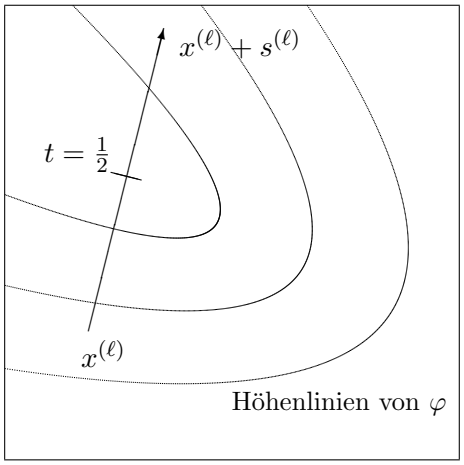
des nichtlinearen Problems, die Nullstellen von f sind gerade ihre globalen Minima, da $\varphi \ge 0$ ist. Der Gradient ist $\varphi'(x) = 2f(x)^\mathsf{T} f'(x)$. Für den Zuwachs von φ in Richtung $s^{(\ell)}$ betrachtet man die Funktion $\psi(t) := \varphi(x^{(\ell)} + ts^{(\ell)})$, $t \ge 0$. In $t = 0$ gilt hierfür

$$\begin{aligned} \psi'(0) = &\quad \varphi'(x^{(\ell)})s^{(\ell)} = -2f(x^{(\ell)})^\mathsf{T} J_\ell J_\ell^{-1} f(x^{(\ell)}) = -2\|f(x^{(\ell)})\|_2^2 \\ = &\quad -2\varphi(x^{(\ell)}) = -2\psi(0) \le 0. \end{aligned} \tag{6.2.3}$$

Wenn also $x^{(\ell)}$ keine Lösung ist, $\psi(0) > 0$, ist $s^{(\ell)}$ eine Abstiegsrichtung der Zielfunktion φ und in einer Umgebung von $t = 0$ existieren Punkte mit $\psi(t) < \psi(0)$. Dies nutzt man in Form einer Liniensuche, man bestimmt t_ℓ (beginnend mit dem Wert eins) so aus

$$t_\ell \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots\} : \quad \varphi(x^{(\ell)} + t_\ell s^{(\ell)}) < \varphi(x^{(\ell)}).$$

Man kann zeigen, dass dabei in der Nähe der Nullstelle immer $t_\ell = 1$ gewählt wird und daher die quadratische Konvergenz erhalten bleibt.



Beispiel 6.2.3 Lineare Bedingungen erzwingt das ungedämpfte Newtonverfahren in einem Schritt. Mit einer Matrix $A \in \mathbb{R}^{m\times n}$ und $G : \mathbb{R}^n \to \mathbb{R}^{n-m}$ sei dazu

$$F(x) = \begin{pmatrix} Ax - b \\ G(x) \end{pmatrix} \quad \Rightarrow \quad F'(x) = \begin{pmatrix} A \\ G'(x) \end{pmatrix}.$$

Löst man den gedämpften Newtonschritt auf nach der neuen Näherung $x^{(\ell+1)}$, ergibt sich $F'(x^{(\ell)})x^{(\ell+1)} = F'(x^{(\ell)})x^{(\ell)} - tF(x^{(\ell)}) \iff$

$$\begin{pmatrix} Ax^{(\ell+1)} - b \\ G'(.)x^{(\ell+1)} \end{pmatrix} = \begin{pmatrix} Ax^{(\ell)} - b - t(Ax^{(\ell)} - b) \\ G'(.)x^{(\ell)} - tG(x^{(\ell)}) \end{pmatrix} = \begin{pmatrix} (1-t)(Ax^{(\ell)} - b) \\ G'(.)x^{(\ell)} - tG(x^{(\ell)}) \end{pmatrix}$$

Im ungedämpften Verfahren mit $1 - t = 0$ erfüllt also schon die nächste Iterierte die lineare Gleichung $Ax^{(\ell+1)} = b$ exakt, und für $0 < t \le 1$ wird der Defekt $Ax^{(\ell)} - b$ um den Faktor $1 - t < 1$ verkleinert.

Diese Prinzipien werden auch beim System (6.1.2) für den zentralen Pfad eingesetzt. Dabei nutzt man die Wahl des Parameters t aber auch dazu, die strengen Zulässigkeitsrestriktionen

einzuhalten. Außerdem muß der Pfadparameter μ gesteuert werden. Für (6.1.2) lautet die Ableitungsmatrix

$$F'_\mu(x,y,z) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^\mathsf{T} & I \\ Z & 0 & X \end{pmatrix} \tag{6.2.4}$$

Man beachte, dass die Jacobimatrix F'_μ nicht direkt von μ abhängt (indirekt über X, Z). Die Regularität erkennt man am homogenen System

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\mathsf{T} & I \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = 0 \iff \begin{cases} A\xi = 0, \\ A^\mathsf{T}\eta = -\zeta, \\ Z\xi + X\zeta = 0. \end{cases}$$

Da Z und X positiv definit sind, kann dieses zu $\xi = -Z^{-1}X\zeta = Z^{-1}XA^\mathsf{T}\eta$ umgeformt werden und reduziert sich am Ende auf

$$AZ^{-1}XA^\mathsf{T}\eta = 0.$$

Da die Matrix $AZ^{-1}XA^\mathsf{T}$ unter den getroffenen Voraussetzungen positiv definit ist, gilt

Satz 6.2.4 *Für $x > 0$, $z > 0$ und Rang$(A) = m$ ist die Matrix (6.2.4) regulär.*

Für den Einsatz der gedämpften Newton-Iteration zur näherungsweisen Pfadverfolgung ist noch die Wahl der Verfahrensparameter festzulegen. Dazu gehören nicht nur die Zahlen t_ℓ und μ_k, sondern auch die Anzahl der Newtonschritte pro Pfadpunkt. Diese wird hier einfach auf eins festgelegt, nach jedem Newtonschritt wird μ_k neu bestimmt. Daher ist immer $\ell = 0$ und alle Größen werden nur noch mit $k-1, k$ indiziert. Hat man (im Newton-Verfahren) eine Approximation $(x^{(k-1)}, y^{(k-1)}, z^{(k-1)})$ akzeptiert, stellt diese eine Näherung für den Pfad $(x(\tilde{\mu}), y(\tilde{\mu}), z(\tilde{\mu}))$ dar in Wert

$$\tilde{\mu} = \frac{1}{n} {z^{(k-1)}}^\mathsf{T} x^{(k-1)}.$$

Damit diese Approximation als Newton-Startpunkt in einem Nachbarpunkt auf dem Pfad dienen kann, definiert man den nächsten Pfadparameter etwas kleiner als $\tilde{\mu}$,

$$\mu_k := \delta_k \tilde{\mu} = \frac{\delta_k}{n} {z^{(k-1)}}^\mathsf{T} x^{(k-1)}, \quad 0 < \delta_k < 1. \tag{6.2.5}$$

Durch die Wahl der Folge δ_k (z.B. konstant 0.1) erzwingt man das Erreichen des Pfad-Endes, $\mu_k \to 0\,(k \to \infty)$. Dann wird für $(x(\mu_k), y(\mu_k), z(\mu_k))$ eine einzelne, gedämpfte Newtoniteration durchgeführt für die negative Suchrichtung $-{s^{(k-1)}}^\mathsf{T} = (\xi^\mathsf{T}, \eta^\mathsf{T}, \zeta^\mathsf{T})$:

$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\mathsf{T} & I \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} Ax^{(k-1)} - b \\ A^\mathsf{T} y^{(k-1)} + z^{(k-1)} - c \\ X_{k-1} z^{(k-1)} - \mu_k \mathbb{1} \end{pmatrix}.$$

Den Dämpfungsparameter t_k wählt man aber so, dass die strikte Zulässigkeit von

$$\begin{pmatrix} x^{(k)} \\ y^{(k)} \\ z^{(k)} \end{pmatrix} := \begin{pmatrix} x^{(k-1)} \\ y^{(k-1)} \\ z^{(k-1)} \end{pmatrix} - t_k \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} \tag{6.2.6}$$

eingehalten wird. Dazu bestimmt man zunächst diejenige Schrittweite λ, welche genau zum Rand der zulässigen Menge führt, geht dann aber nur einen Bruchteil dieses Schritts. Der größte Wert für λ mit $x_j^{(k-1)} - \lambda\xi_j \geq 0$, $z_j^{(k-1)} - \lambda\zeta_j \geq 0\,\forall j$ ist gegeben durch

$$\lambda := \min\left\{ \min_{\xi_j>0} \frac{x_j^{(k-1)}}{\xi_j}, \min_{\zeta_j>0} \frac{z_j^{(k-1)}}{\zeta_j} \right\}$$

(Minima über leere Mengen als ∞ gewertet). In diesem λ ist mindestens eine der Komponenten von (6.2.6) null. Mit t_k bleibt man nun leicht unter diesem Wert und wählt einen Bruchteil

$$t_k := q_k\lambda \quad \text{in (6.2.6)}, \quad 0 < q_k < 1,$$

etwa mit konstantem $q_k = 0.9$.

Eine einfache Version dieses Gesamtverfahrens wird hier zusammengefaßt (Primal-duale Innere-Punkt-Methode):

Ausgehend von $x^{(0)} \in \hat{X}$, $y^{(0)} \in \hat{Y}$, $z^{(0)} := A^\mathsf{T} y^{(0)} - c$. Für $k = 1, 2, \ldots$:

1. Berechne $\mu_k := \frac{\delta_k}{n} z^{(k-1)\mathsf{T}} x^{(k-1)}$,

2. Löse
$$\begin{pmatrix} A & 0 & 0 \\ 0 & A^\mathsf{T} & I \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} Ax^{(k-1)} - b \\ A^\mathsf{T} y^{(k-1)} + z^{(k-1)} - c \\ X_{k-1} z^{(k-1)} - \mu_k \mathbb{1} \end{pmatrix}.$$

3. Berechne
$$t_k := q_k \min\left\{ \min_{\xi_j>0} \frac{x_j^{(k-1)}}{\xi_j}, \min_{\zeta_j>0} \frac{z_j^{(k-1)}}{\zeta_j} \right\}, \quad \begin{pmatrix} x^{(k)} \\ y^{(k)} \\ z^{(k)} \end{pmatrix} := \begin{pmatrix} x^{(k-1)} - t_k\xi \\ y^{(k-1)} - t_k\eta \\ z^{(k-1)} - t_k\zeta \end{pmatrix}$$

4. Stop, falls $\mu_k \leq tol$, andernfalls setze $k := k + 1$ und gehe zu 1.

Bemerkung: Durch eine andere Anordnung der Funktion F_μ kann man übrigens im Newtonschritt eine symmetrische Jacobi-Matrix erreichen in der Form

$$\begin{pmatrix} 0 & A^\mathsf{T} & I \\ A & 0 & 0 \\ I & 0 & Z^{-1}X \end{pmatrix}$$

Diese ist allerdings nicht definit.

Beispiel 6.2.5 Das einfache Innere-Punkte-Verfahren wird auf das Problem (LP3) angewendet mit

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} \frac{5}{2} \\ 3 \end{pmatrix}, \quad c^\mathsf{T} = (-3, -2, 0).$$

a) Test mit konsistenten Startwerten $x^{(0)} := (\frac{1}{2}, 1, \frac{1}{2})^\mathsf{T} \in \hat{X} > 0^\mathsf{T}$, $y^{(0)} = (-1, -2)^\mathsf{T} \in \hat{Y}$, $z^{(0)} = c - A^\mathsf{T} y^{(0)} = (1, 3, 3)^\mathsf{T} > 0^\mathsf{T}$. Konstante Faktoren $\delta_k = 0.1$, $q_k = 0.95$. Mit dem ersten $\tilde{\mu} = \frac{1}{n} {z^{(0)}}^\mathsf{T} x^{(0)} = 5/3$ und $\mu_1 = \frac{1}{6}$ lauten die Näherungen:

k	μ_k	t	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$-y_1^{(k)}$	$-y_2^{(k)}$	$z_1^{(k)}$	$z_2^{(k)}$	$z_3^{(k)}$
1	0.166667	0.9	0.56864	1.06864	0.29406	1.42711	0.40847	0.26271	0.24406	1.83559
2	0.031667	0.9	0.64401	1.14401	0.06796	1.34976	0.34205	0.04157	0.03386	1.69181
3	0.006016	0.95	0.66414	1.16414	0.00757	1.33774	0.33418	0.00965	0.00609	1.67192
4	0.000872	0.95	0.66637	1.16637	0.00089	1.33413	0.33344	0.00169	0.00100	1.66757
5	0.000126	0.95	0.66663	1.16663	0.00012	1.33346	0.33335	0.00026	0.00015	1.66681
6	0.000018	1	0.66666	1.16666	0.00001	1.33335	0.33333	0.00003	0.00002	1.66668

Man prüft leicht nach, dass alle Iterierten konsistente Punkte für (LP3) und (LP3*) sind. Die Größen μ_{k-1} und die Fehler verkleinern sich pro Schritt um einen Faktor $\in [0.1, 0.2]$.

b) Das Verfahren funktioniert auch mit inkonsistenten Startwerten. Denn wegen Beisp.6.2.3 werden auch im gedämpften Newtonverfahren die Defekte der linearen Nebenbedingungen $Ax - b = 0$, $A^\mathsf{T} y + z - c = 0$ immer kleiner, und die Vorzeichenbedingungen wird in Schritt 3 des Verfahrens explizit erzwungen. Mit $x^{(0)} = (1, 1, 1)^\mathsf{T}$, $y^{(0)} = (-2, -2)^\mathsf{T}$, $z^{(0)} = c - A^\mathsf{T} y^{(0)} = (3, 4, 4)$ ist $\mu_1 = \frac{1}{30} {z^{(0)}}^\mathsf{T} x^{(0)} = 11/30$. In diesem Beispiel wird in allen Schritten wegen $\lambda = 1$ immer $t = 0.95$ gewählt.

k	μ_k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$-y_1^{(k)}$	$-y_2^{(k)}$	$z_1^{(k)}$	$z_2^{(k)}$	$z_3^{(k)}$
1	0.366666	0.52795	1.00295	0.516163	2.43081	0.05287	1.91449	0.53655	2.48368
2	0.094362	0.59334	1.09209	0.22498	1.32942	0.36955	0.02839	0.06852	1.69897
3	0.015797	0.65917	1.15911	0.02275	1.34472	0.33412	0.02356	0.01296	1.67884
4	0.002292	0.66581	1.16581	0.00257	1.33534	0.33355	0.00424	0.00245	1.66890
5	0.000332	0.66656	1.16656	0.00032	1.33366	0.33337	0.000682	0.000392	1.66702
6	0.000048	0.66666	1.16666	0.00003	1.33337	0.33334	0.00007	0.00004	1.66670

Die Konvergenzgeschwindigkeit ist vergleichbar zum Fall a), das Verfahren startet aber mit einem größerem Anfangsfehler.

A Symbole, Abkürzungen

Symbole

		Seite
N	$=\{1,2,\ldots,n\}$	
$\mathbb{1}$	Vektor von Einsen $\mathbb{1}=(1,\ldots,1)^\mathsf{T}$	10
$\mathrm{aff}(M)$	affine Hülle einer Menge M	10
$\arg\min$	Argumentwert in einer Minimalstelle	13
$\mathbb{B}$	boolesche Menge $\mathbb{B}=\{0,1\}$	2
$B_r(z)$	Kugel um z mit Radius r	2
Δ_n	Standardsimplex $\Delta_n=\{x\in\mathbb{R}^n:\ \mathbb{1}^\mathsf{T}x=1,\, x\geq 0\}$	30
$E(M)$	Menge der Ecken einer konvexen Menge M	36
e_i	Einheitsvektoren	
$\mathrm{keg}(M)$	konische Hülle einer Menge M, von M erzeugter Kegel	40
$\mathrm{konv}(M)$	konvexe Hülle einer Menge M	30
$H(a,\alpha)$	Hyperebene mit Normalenrichtung $a\neq 0$	28
H^+, H^-	offener Halbraum in (bzw. entgegen) Normalenrichtung	28
$H^\oplus, H^\ominus$	entsprechende abgeschlossene Halbräume	28
$L(M)$	Linealraum einer Menge M	28
(LP1)..(LP3)	Standardformen linearer Programme	8
(LP*), (LPi*)	duale Programme	49
$O^+(M)$	Ausdehnungskegel einer konvexen Menge M	41
p_M	Projektion $p_M:\mathbb{R}^n\to M$ auf eine konvexe Menge M	33
$\mathbb{R}^n_+$	positiver Kegel $\mathbb{R}^n_+=\{x\in\mathbb{R}^n:\ x\geq 0\}$	9

Bezeichnungsweisen

Zu Mengen $M\subseteq\mathbb{R}^n$, Matrizen $A\in\mathbb{R}^{m\times n}$, Vektoren $x,y\in\mathbb{R}^n$ wurden eingeführt:

$\bar{M}, \mathring{M}$	Abschluß und Inneres
M^*	Polarkegel
$[x,y]$	Verbindungsstrecke von Punkten
$J(x), J^\pm(x)$	Stützindizes von x: der x_i, die nicht null, bzw. positiv/negativ
x_J	Teilvektor $x_J=(x_{j_1},\ldots,x_{j_k})^\mathsf{T}$ zu Indexmenge $J=\{j_1,\ldots,j_k\}\subseteq\{1,\ldots,n\}$
$A^{(L)}$	Untermatrix aus Zeilenvektoren ${a^{(i)}}^\mathsf{T}=e_i^\mathsf{T}A$ zu Indexmenge $L\subseteq\{1,\ldots,m\}$
A_J	Untermatrix aus Spaltenvektoren $a_j=Ae_j$ zu Indexmenge $J\subseteq\{1,\ldots,n\}$
H, $\bar{H}$	einfaches und erweitertes Simplex-Tableau, H und $A_J^{-1}A$ enthalten die gleichen Elemente, die Zeilenindizes von H sind aber $1,\ldots,m$

Index