Operations Research

Universität Bielefeld

WS 2024/2025

Inhaltsverzeichnis

Kapitel 1

Einführung

1.1 Operations Research

'*Operations Research*' (OR) als Disziplin unter diesem Namen war ab Mitte der 1930er Jahre als Methodologie militärischer Strategien enstanden, vor allem in Grossbritannien und den USA: Durch die Entwicklung der Radartechnik hatte man neue Möglichkeiten erlangt, feindliche Flugverbände im Voraus zu erkennen und den Einsatz der eigenen Luftstreitkräfte sowie den Schiffsverkehr optimal zu planen. Später hat sich die OR in vor allem in der Ökonomie verbreitet, z.B. unter dem Titel '*Unternehmensforschung*'.

Aus der Sicht der Mathematik kann man sicher sagen, dass OR schon damals einfach eine Melange aus verschiedenen lange bekannten Techniken war (lineare Algebra, Konvexitätsprobleme, Kombinatorik, Lagrange-Mechanik, Transportprobleme, etc.), gut abgestimmt auf die jeweilige Anwendung. Die Disziplin definiert sich also eher *nicht* durch ganz bestimmte Methoden, sondern durch ihr *Ziel*. (Zum Beispiel ist es müssig zu diskutieren, ob eine Methode zu OR gehört oder in die lineare Optimierung, oft ist beides richtig.)

Das *Ziel der OR* ist die Vorbereitung von möglichst guten Entscheidungen durch Anwendung mathematischer Methoden (stark praxisorientiert).

Die *Hauptaufgabe* der OR ist die Übersetzung eines realen Entscheidungsproblems in ein Optimierungs- bzw. Simulationsmodell.

In der Praxis sind das eher feedback-loops (Erkennen des Problems - Bestimmung von Zielen und Handlungsmöglichkeiten - mathematische Modellbildung - Datensammlung - Lösungsfindung - Bewertung der Lösung - und letzlich Akzeptanz oder Verwerfen der Lösung, ggf. auch Überarbeitung des Modells).

OR im engeren Sinne bezieht sich auf mathematische Modellierung von Entscheidungsproblemen und Entwicklung von Algorithmen zur Anwendung und Lösung mathematischer Modelle. OR in diesem engeren Sinne hat viele Teilgebiete, z.B.

- Lineare Optimierung (Maximierung unter linearen Nebenbedingungen)
- Nichtlineare Optimierung (z.B. quadratisch)
- Optimaler Transport

- Spieltheorie
- Graphentheorie und Netzplantechnik
- Diskrete und kombinatorische Optimierung
- Dynamische Optimierung
- Warteschlangentheorie
- Simulation
- Tabellenkalkulation.

Wir werden uns hier in der Vorlesung konzentrieren auf

- Lineare Optimierung
- Graphentheorie und Netzplantechnik
- Nichtlineare Optimierung, Dynamische Optimierung (sofern die Zeit reicht).

Das Skript, dem wir hier folgen, ist nicht von Ihrem Vorlesenden erdacht worden, sondern über viele Jahre hinweg an der Fakultät gewachsen und durch viele Hände (bzw. Köpfe) gegangen.

1.2 Literatur

Man kann endlos viel Literatur zum Thema finden - praktische, theoretische, klassischere, neuere. Hier einige Vorschläge:

- K.H. Borgwardt, *Optimierung, Operations Research, Spieltheorie*, Springer Basel AG 2001
- W. Domschke, A. Drexl, R. Klein, A. Scholl, *Einführung in Operations Research*, Springer, 2015.
- W. Hochstättler, *Lineare Optimierung*, Springer, 2012.
- D. Jungnickel, *Graphs, Networks and Algorithms*, Springer, 2013. (Engl. Version von *Graphen, Netzwerke und Algorithmen*, Spektrum, 1994.)
- D. Jungnickel, *Optimierungsmethoden - Eine Einführung*, Springer, 2015.
- H. Peters, *Game Theory, A Multi-Leveled Approach*, Springer 2015
- I. Wegener, *Operations Research*, Universität Dortmund, Skript zur Vorlesung WS 1998/99

Kapitel 2

Lineare Optimierung

2.1 Problemstellung

Beispiele 2.1.1. Ein Landwirt hat 100 ha Land. Davon kann er x_1 ha für den Anbau von Kartoffeln nutzen und x_2 ha für den Anbau von Getreide.

Das *Ziel* ist es, x_1 und x_2 so zu wählen, dass der Gewinn maximal ausfällt.

Nehmen wir mal an, Kosten, benötigte und verfügbare Ressourcen und möglicher Gewinn sind wie folgt:

	Kartoffeln/ha	Getreide/ha	insg. verfügbar
Anbaukosten	100	200	11000
Arbeitstage	1	4	160
Gewinn	400	1200	

(Das sind natürlich wilkürliche fiktive Zahlen, wie ein kurzer Blick in die Infos der Landwirtschaftskammer zeigt, deshalb auch ohne Einheit. Das Modell ist auch viel zu naiv. Aber vom grundlegenden Prinzip her sehen solche Kalkulationen tatsächlich so aus.)

Man erhält folgende Restriktionen:

$$\begin{aligned} x_1, x_2 &\geq 0 \quad \text{(Fläche immer nichtnegativ)} \\ 100\,x_1 + 200\,x_2 &\leq 11000 \quad \text{(wegen Anbaukosten)} \\ x_1 + 4\,x_2 &\leq 160 \quad \text{(wegen Arbeitstagen)} \\ x_1 + x_2 &\leq 100 \quad \text{(wegen max. verfügbarer Fläche).} \end{aligned}$$

Der entsprechende Gewinn ist

$$f(x_1, x_2) = 400\,x_1 + 1200\,x_2.$$

Wir formalisieren diesen Typ von Aufgabenstellung abstrakter. Zur Erinnerung: Eine lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ nennt man eine *Linearform* auf $\mathbb{R}^n$. Zu jeder Linearform f auf $\mathbb{R}^n$ gibt es $a_1, ..., a_n \in \mathbb{R}$, sodass

$$f(x) = a_1x_1 + ... + a_nx_n, \quad \text{für alle } x = (x_1, ..., x_n) \in \mathbb{R}^n.$$

Die Menge aller Linearformen auf $\mathbb{R}^n$ bildet einen n-dimensionalen Vektorraum (den *Dualraum* zu $\mathbb{R}^n$).

Definition 2.1.2. Bei einem *allgemeinen linearen Optimierungsproblem* sind Zahlen $m, n \in \mathbb{N}$ sowie Linearformen $f_1, ..., f_m$ und f auf $\mathbb{R}^n$ und Zahlen $c_1, ..., c_m \in \mathbb{R}$ gegeben. Die Menge

$$P := \{x \in \mathbb{R}^n : f_i(x) \leq c_i \quad \text{für alle } i = 1, ..., m\}$$

heisst *zulässiger Bereich*, die Bedingungen $f_i \leq c_i$, $i = 1, ..., m$ nennt man die *Restriktionen*. Die Linearform f nennt man die *Zielfunktion*.

Ein Punkt $y \in P$ heisst *Lösung* des Problems, falls er eine Maximalstelle der Zielfunktion f auf P ist, also falls

$$f(y) \geq f(x) \quad \text{für alle } x \in P.$$

Bemerkung 2.1.3.

(i) Möchte man statt einer Maximalstelle für f eine Minimalstelle finden, so kann man einfach $-f$ betrachten.

(ii) Restriktionen der Form $f_i = c_i$ kann man ganz einfach umschreiben als $f_i \leq c_i$ und $-f_i \leq -c_i$.

(iii) Umgekehrt kann man Restriktionen in Ungleichungsform durch Einführung einer zusätzlichen reellen Variablen in Restriktionen mit Gleichheit umwandeln: Man hat

$$f_i(x) \leq c_i \quad \text{genau dann, wenn} \quad f_i(x) + \widetilde{x} = c_i \text{ und } \widetilde{x} \geq 0.$$

Das ist trivial, aber in der Praxis oft sinnvoll, z.B. kann man anstelle der Restriktion

'Arbeitszeit $\leq$ verfügbare Zeit'

die Restriktion

'Arbeitszeit + Freizeit = verfügbare Zeit'

in das Modell aufnehmen.

2.2 Polyeder

Wir schauen uns einige Grundbegriffe zu Konvexität an.

Definition 2.2.1. Eine Teilmenge $A \subset \mathbb{R}^n$ heisst *konvex*, falls für alle $x, y \in A$ die Verbindungsstrecke

$$[x, y] := \{\lambda x + (1 - \lambda)y : \ 0 \leq \lambda \leq 1\}$$

in A enthalten ist.

Wir nutzen auch die Schreibweisen

$$]x, y[:= [x, y] \setminus \{x, y\}, \quad]x, y] := [x, y] \setminus \{x\}, \quad \text{und} \quad [x, y[:= [x, y] \setminus \{y\}.$$

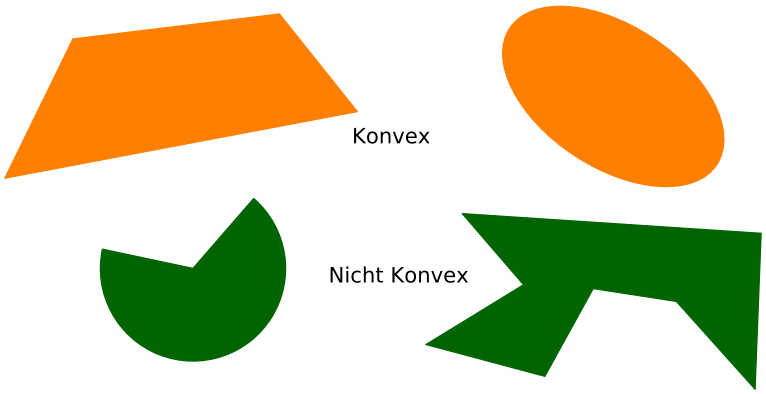


Abbildung 2.1: Konvexe und nicht konvexe Mengen

Beispiel 2.2.2. Sei f eine Linearform auf $\mathbb{R}^n$, $f \neq 0$ und sei $c \in \mathbb{R}$. Dann ist die Menge

$$\{f \leq c\} := \{x \in \mathbb{R}^n : \ f(x) \leq c\}$$

konvex:

Sind $x, y \in \{f \leq c\}$ und $0 \leq \lambda \leq 1$, so hat man $f(x) \leq c$ und $f(y) \leq c$ und damit, wegen der Linearität von f,

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y) \leq \lambda c + (1-\lambda)c = c,$$

also $\lambda x + (1-\lambda)y \in \{f \leq c\}$.

Geometrisch ist die Situation wie folgt: Die Menge

$$\{f = c\} := \{x \in \mathbb{R}^n : \ f(x) = c\}$$

ist eine Hyperebene im Raum $\mathbb{R}^n$, welche letzteren in zwei Halbräume 'teilt', nämlich in $\{f \leq c\}$ und $\{f \geq c\}$ (mit offensichtlicher Bedeutung).

Wir formalisieren das.

Definition 2.2.3. Ein *abgeschlossener Halbraum* in $\mathbb{R}^n$ ist eine Teilmenge, die sich in der Form $\{f \leq c\}$ darstellen lässt mit einer Linearform f auf $\mathbb{R}^n$ und einer Konstanten $c \in \mathbb{R}$.

Definition 2.2.4. Ein endlicher Durchschnitt abgeschlossener Halbräume in $\mathbb{R}^n$ heisst ein *Polyeder*.

Lemma 2.2.5. *Sei $I \neq \emptyset$ und $\{A_i\}_{i \in I}$ eine Familie konvexer Mengen $A_i \subset \mathbb{R}^n$. Dann ist auch*

$$\bigcap_{i \in I} A_i$$

konvex.

Beweis. Falls $x, y \in \bigcap_{i \in I} A_i$, so hat man $x, y \in A_i$ für alle $i \in I$, und weil die A_i alle konvex sind, $[x, y] \subset A_i$, $i \in I$. Damit folgt aber $[x, y] \subset \bigcap_{i \in I} A_i$. □

Korollar 2.2.6. *Jedes Polyeder ist konvex und abgeschlossen.*

Beispiel 2.2.7. In der Situation von Beispiel 2.1.1 (i) sind die Hyperebenen Geraden, und gemäss den Restriktionen (und der Vernunft) muss der zulässige Bereich das farbige Polyeder sein, dass 'unterhalb' aller drei Geraden liegt und zudem von unten und von links durch die Achsen begrenzt wird, Abbildung 2.2.7

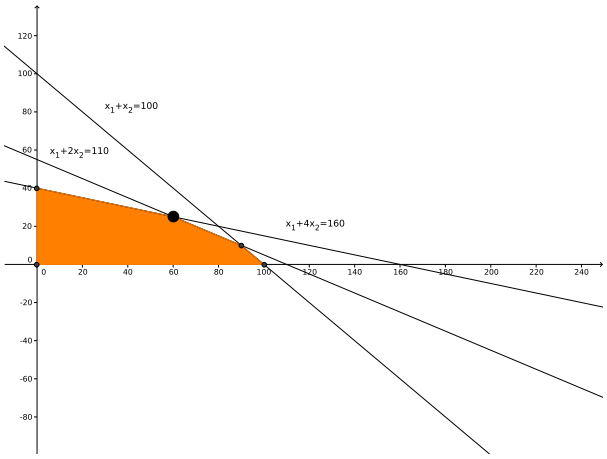


Abbildung 2.2: Der zulässige Bereich aus Beispiel 2.1.1 (i) als Polyeder.

Jetzt erinnern wir uns an folgendes: Ist $A \subset \mathbb{R}^n$ kompakt, so nimmt eine stetige Funktion $f : A \to \mathbb{R}$ auf A ihr Maximum an, d.h. es existiert ein Punkt $y \in A$ sodass $f(y) \geq f(x)$ für alle $x \in A$. (Ein solches y nennt man dann eine Maximalstelle für f auf A.)

Da ein beschränktes Polyeder $P \neq \emptyset$ im $\mathbb{R}^n$ kompakt ist und eine Linearform f auf $\mathbb{R}^n$ stetig, so nimmt $f|_P$ auf P ihr Maximum an, genauer:

Es gibt einen Punkt $y \in P$ sodass $f(y) \geq f(x)$ für alle $x \in P$.

Man kann nun fragen, wo in P man so eine Maximalstelle y finden kann.

Proposition 2.2.8. *Ist P ein beschränktes Polyeder im $\mathbb{R}^n$ und nimmt eine Linearform $f \neq 0$ ihr Maximum auf P in einem Punkt $y \in P$ an, so kann y kein innerer Punkt von P sein.*

Beweis. Nehmen wir an, y sei ein innerer Punkt von P, dann gibt es also eine kleine offene Kugel B_y mit Mittelpunkt y, sodass $B_y \in P$. Mit geeigneten $a_1, ..., a_n \in \mathbb{R}^n$ gilt $f(x) = a_1x_1 + ... + a_nx_n$ für alle $x \in \mathbb{R}^n$, insbesondere ist also f in y stetig partiell differenzierbar. Wenn f in y ihr Maximum auf P annimmt, so hat f in y ein lokales Maximum. Dann hat man aber $a_i = \frac{\partial f}{\partial x_i}(y) = 0$ für alle i, was $f = 0$ implizieren würde, Widerspruch. □

Diese Beobachtung motiviert uns, den Rand ∂P eines Polyeders P genauer anzuschauen. Man kann sogar ziemlich abstrakt arbeiten, mit der Intuition 'Polyeder' als Leitmotiv.

Definition 2.2.9. Sei $A \subset \mathbb{R}^n$ konvex.

(i) Eine Teilmenge S von A ist eine *Seite von A*, falls S konvex ist und für zwei beliebige Punkte $x, y \in A$ folgendes gilt: Gibt es ein $0 < \lambda < 1$, sodass $\lambda x + (1 - \lambda)y \in S$, dann hat man $x, y \in S$.

(ii) Ein Punkt $x \in A$ heisst *extremal*, falls $\{x\}$ eine Seite von A ist.

Wir schreiben A_e für die Menge der Extremalpunkte von A.

Wie üblich schreiben wir $B(x, r) = \{y \in \mathbb{R}^n : |x - y| < r\}$ für die offene Kugel mit Radius $r > 0$ und Mittelpunkt x. Wir schreiben $\overline{B(x, r)}$ oder $\mathrm{cl}(B(x, r))$ für ihren Abschluss, d.h. die abgeschlossene Kugel mit Radius $r > 0$ und Mittelpunkt x. Die Sphäre $\partial B(x, r)$ mit Radius $r > 0$ und Mittelpunkt x bezeichnen wir auch mit $S(x, r)$.

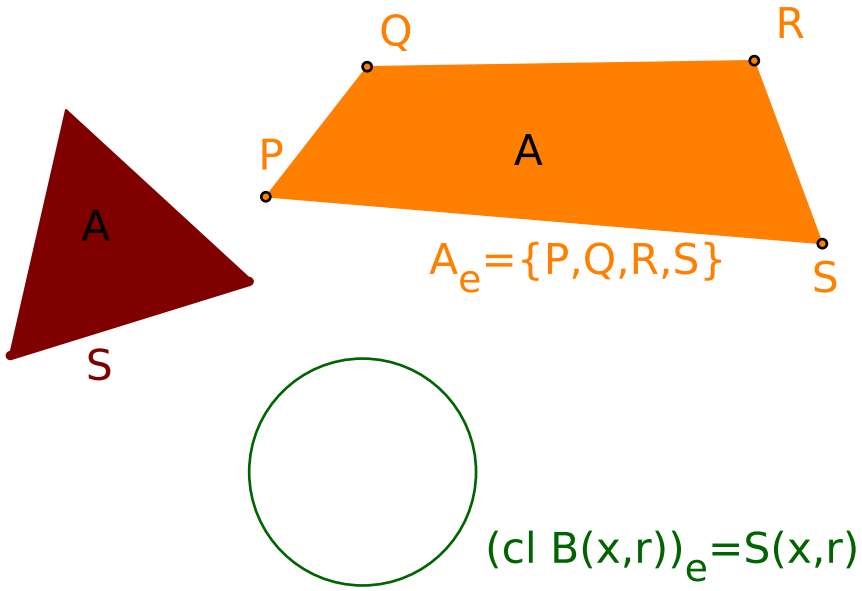


Abbildung 2.3: Seiten und Extremalpunkte

Beispiele 2.2.10. Die Verbindungsgerade zwischen zwei Eckpunkten eines abgeschlossenen Dreiecks ist eine Seite. Jeder Eckpunkt eines Polyeders ist ein Extremalpunkt. Die Sphäre $S(x,r)$ ist die Menge der Extremalpunkte der abgeschlossenen Kugel $\overline{B(x,r)}$.

Bemerkung 2.2.11.

(i) Für jede konvexe Menge A sind $\emptyset$ und A (triviale) Seiten.

(ii) Ist S eine Seite von A, so ist auch $A \setminus S$ konvex: Sind $x, y \in A \setminus S$ so muss man $\lambda x + (1-\lambda)y \in A \setminus S$ haben für alle $\lambda \in [0,1]$, sonst ergäbe sich ein Widerspruch zur Definition einer Seite.

Lemma 2.2.12. *Sei $A \subset \mathbb{R}^n$ konvex. Die folgenden Aussagen sind äquivalent:*

(i) x ist ein Extremalpunkt von A.

(ii) x ist nur trivial konvex kombinierbar aus Punkten von A, d.h. falls $x = \lambda y + (1-\lambda)z$ ist mit $y, z \in A$ und $0 < \lambda < 1$, so muss $y = z = x$ sein.

(iii) $A \setminus \{x\}$ konvex.

Beweis. Die Äquivalenz of (i) und (ii) folgt direkt aus der Definition, ebenso, dass Aussage (ii) Aussage (iii) impliziert. Aus (iii) folgt aber auch (ii): Seien $y, z \in A$, $\lambda \in [0,1]$ mit $\lambda y + (1-\lambda)z = x$. Da $A \setminus \{x\}$ konvex ist, muss $y = x$ sein oder $z = x$. Sagen wir o.B.d.A. dass $y = x$. Falls auch $z = x$, dann muss man nichts mehr zeigen. Falls nicht, so ist $\lambda < 1$ und $(1-\lambda)z = (1-\lambda)x$, also auch in diesem Falle $z = x$. □

Für Linearformen beobachten wir nun folgenden Effekt.

Satz 2.2.13. *Sei $A \subset \mathbb{R}^n$ konvex, f eine Linearform auf $\mathbb{R}^n$ und $\alpha \in \mathbb{R}$ mit $f \leq \alpha$ auf A. Dann ist*

$$S := A \cap \{f = \alpha\}$$

eine Seite von A. Ist A ein Polyeder, so ist auch S ein Polyeder.

Beweis. Da $A \subset \{f \leq \alpha\}$ gilt, hat man

$$S = A \cap \{f \geq \alpha\} = A \cap \{-f \leq -\alpha\},$$

und da der abgeschlossene Halbraum $\{-f \leq -\alpha\}$ konvex ist, ist mit Lemma 2.2.5 auch S konvex. Falls A ein Polyeder ist, folgt mit Definition 2.2.4, dass auch S ein Polyeder ist.

Um zu zeigen, dass S eine Seite von A ist, nehmen wir an, dass $y, z \in A$ und $0 < \lambda < 1$ sind und $x := \lambda y + (1-\lambda)z$ ein Element von S ist. Dann folgt, dass

$$\alpha = f(x) = f(\lambda y + (1-\lambda)z) = \lambda f(y) + (1-\lambda) f(z).$$

Es ist klar, dass $f(y) \leq \alpha$ und $f(z) \leq \alpha$ gilt. Wäre nun z.B. $f(y) < \alpha$, könnte die vorangegangene Gleichheit nicht mehr gelten. Ebenso, falls $f(z) < \alpha$. Also hat man $f(y) = f(z) = \alpha$ und damit $y, z \in S$. □

Korollar 2.2.14. *Ist P der zulässige Bereich eines allgemeinen linearen Optimierungsproblems und f seine Zielfunktion, dann nimmt $f|_P$ ihr Maximum auf einer Seite von P an.*

Wir betrachten als nächstes die Extremalpunkte von Seiten.

Satz 2.2.15. *Sei S die Seite einer konvexen Menge $A \subset \mathbb{R}^n$. Dann gilt*

$$S_e = A_e \cap S.$$

Beweis. Man hat $A_e \cap S \subset S_e$: Ist $x \in A_e \cap S$, so ist nach Lemma 2.2.12 die Menge $A \setminus \{x\}$ konvex. Also ist auch

$$S \setminus \{x\} = (A \setminus \{x\}) \cap S$$

konvex, und somit ist $x \in S_e$ wegen Lemma 2.2.12.

Man hat $S_e \subset A_e \cap S$: Sei $x \in S_e$. Sind $y, x \in A$ und $0 < \lambda < 1$ so, dass $x = \lambda y + (1-\lambda)z$, dann folgt, da S eine Seite von A ist, $y, z \in S$. Weil aber x ein Extremalpunkt von S ist, muss dann $y = z = x$ gelten. Daraus folgt, dass $x \in A_e$ ist. □

Wir betrachten einen Mechanismus, der Konvexität erzeugt.

Definition 2.2.16. Für $A \subset \mathbb{R}^n$ heisst

$$k(A) := \bigcap_{A \subset B,\ B \text{ konvex}} B$$

die *konvexe Hülle* von A.

Bemerkung 2.2.17.

(i) $k(A)$ ist wohldefiniert für alle $A \subset \mathbb{R}^n$: Mit $B = \mathbb{R}^n$ existiert stets eine konvexe Menge mit $A \subset B$.

(ii) Nach Lemma 2.2.5 ist $k(A)$ konvex, und nach Konstruktion ist $k(A)$ die kleinste konvexe Menge, die A enthält.

(iii) A ist genau dann konvex, wenn $k(A) = A$: Ein konvexes A ist dann die kleinste konvexe Obermenge von sich selbst. Falls $k(A) = A$, so ist A konvex wegen (ii).

(iv) $k(A)$ ist genau dann beschränkt, wenn A beschränkt ist: Dass die Beschränktheit von $k(A)$ die von A impliziert ist klar wegen $k(A) \supset A$. Ist A beschränkt, dann existiert $r > 0$ sodass $A \subset B(0, r)$, und da die Kugel $B(0, r)$ konvex ist, kommt sie als eine der Mengen B im Durchschnitt vor, und somit folgt $k(A) \subset B(0, r)$.

Man kann $k(A)$ explizit ausdrücken.

Satz 2.2.18. *Sei $A \subset \mathbb{R}^n$. Dann gilt*

$$k(A) = \left\{ \sum_{i=1}^{m} \lambda_i a_i : \ m \in \mathbb{N}, \ a_1, ..., a_m \in A, \ \lambda_1, ..., \lambda_m \geq 0, \ \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Für endlich viele Punkte $a_1, ..., a_m \subset \mathbb{R}^n$ und $\lambda_1, ..., \lambda_m \geq 0$ mit $\sum_{i=1}^{m} \lambda_i = 1$ nennt man

$$\sum_{i=1}^{m} \lambda_i a_i$$

eine *Konvexkombination* der Punkte $a_1, ..., a_m$. Satz 2.2.18 sagt also, dass die konvexe Hülle $k(A)$ die Menge aller Konvexkombinationen von endlich vielen Punkten aus A ist.

Beweis. Sei

$$B := \left\{ \sum_{i=1}^{m} \lambda_i a_i : \ m \in \mathbb{N}, \ a_1, ..., a_m \in A, \ \lambda_1, ..., \lambda_m \geq 0, \ \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

Der Satz behauptet, dass $k(A) = B$. Wir zeigen zunächst, dass $A \subset B$ gilt und B konvex ist. Aus diesen beiden Fakten zusammen folgt, dass $k(A) \subset B$, denn $k(A)$ ist ja die kleinste Menge mit diesen beiden Eigenschaften. Die Inklusion $A \subset B$ ist trivial, denn jedes $a \in A$ selbst ist ja Konvexkombi von endlich vielen Punkten (nämlich einem) aus A. Um die Konvexität von B zu sehen, seien

$$x = \sum_{i=1}^{m} \lambda_i a_i \quad \text{und} \quad y = \sum_{j=1}^{l} \mu_j b_j$$

zwei beliebige Punkte aus B. Hier sind $a_i, b_j \in A$ und die λ_i und μ_j sind alle nichtnegativ und summieren sich jeweils zu eins. O.B.d.A. dürfen wir annehmen, dass $l = m$ (sonst 'füllen wir mit Nullen auf') und $b_i = a_i$ für $i = 1, ..., m$ (denn $\{a_1, ..., a_m, b_1, ..., b_l\}$ ist endliche Menge von Punkten aus A, und man könnte die einfach relabeln mit $\{c_1, ..., c_k\}$, wobei $k \leq m + l$ wäre). Ist $0 \leq \lambda \leq 1$, so hat man

$$\lambda x + (1 - \lambda) y = \lambda \sum_{i=1}^{m} \lambda_i a_i + (1 - \lambda) \sum_{i=1}^{m} \mu_i a_i = \sum_{i=1}^{m} (\lambda \lambda_i + (1 - \lambda) \mu_i) a_i,$$

und es gilt

$$\sum_{i=1}^{m} (\lambda \lambda_i + (1 - \lambda) \mu_i) = \lambda \sum_{i=1}^{m} \lambda_i + (1 - \lambda) \sum_{i=1}^{m} \mu_i = \lambda + (1 - \lambda) = 1,$$

also ist auch $\lambda x + (1 - \lambda) y$ Konvexkombi von endlich vielen Punkten aus A. Also ist B konvex.

Um die gewünschte Gleichheit $k(A) = B$ zu zeigen, müssen wir nun noch $B \subset k(A)$ beweisen, also, dass ein beliebiges Element $x = \sum_{i=1}^{m} \lambda_i a_i$ aus B auch ein Element von

$k(A)$ ist. Das machen wir induktiv: Für $m = 1$ ist $\lambda_1 = 1$ und $x = a_1 \in A \subset k(A)$. Nehmen wir an, jede Konvexkombi $\sum_{i=1}^m \lambda_i a_i \in B$ ist in $k(A)$. Für den Induktionsschritt müssen wir zeigen dann, dass auch jede Konvexkombi $\sum_{i=1}^{m+1} \lambda_i a_i \in B$ in $k(A)$ ist. O.E. hat man $\lambda_{m+1} < 1$ (denn sonst müssten $\lambda_1 = ... = \lambda_m = 0$ sein, dieser Fall ist schon klar). Nun ist

$$\sum_{i=1}^{m+1} \lambda_i a_i = (1 - \lambda_{m+1}) \left(\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} a_i \right) + \lambda_{m+1} a_{m+1},$$

und weil $\frac{\lambda_i}{1-\lambda_{m+1}} \geq 0$ und $\sum_{i=1}^m \frac{\lambda_i}{1-\lambda_{m+1}} = 1$ (Frage: Warum ?), folgt

$$\sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} a_i \quad \in \quad k(A)$$

nach Induktionsvoraussetzung. Weil $a_{m+1} \in A \subset k(A)$ bedeutet das aber, dass $\sum_{i=1}^{m+1} \lambda_i a_i$ Konvexkombi zweier Elemente aus $k(A)$ ist, und wegen Konvexität von $k(A)$ somit in $k(A)$. □

Wir wissen bereits, dass $k(A)$ genau dann beschränkt ist, wenn A beschränkt ist. Insofern ist in der folgenden Beobachtung der Aspekt der Abgeschlossenheit der interessante Teil.

Korollar 2.2.19. *Für jede endliche Menge $A \subset \mathbb{R}^n$ ist $k(A)$ kompakt.*

Natürlich ist eine endliche Menge insbesondere beschränkt. (Frage: Wie zeigt man das ?) Die Kompaktheit (also insbesondere die Abgeschlossenheit) der konvexen Hülle einer beschränkten Menge ist nicht selbstverständlich.

Beispiel 2.2.20. Man hat $k(B(0,1)) = B(0,1)$, eine offene Menge.

Beweis. Sei $A = \{a_1, ..., a_m\}$. Dann ist $k(A)$ das Bild der kompakten Menge

$$\{(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m_+ : \sum_{i=1}^{m} \lambda_i = 1\}$$

unter der stetigen Abbildung $\varphi((\lambda_1, ..., \lambda_m)) := \sum_{i=1}^m \lambda_i a_i$. Wie üblich ist hier $\mathbb{R}^m_+ = \{(x_1, ..., x_m) \in \mathbb{R}^m : x_i \geq 0,\ i = 1, ..., m\}$. □

Der folgende Satz ist eine Babyversion eines der wichtigsten Sätze zu konvexen Mengen. Er sagt, dass man ein beschränktes Polyeder stets aus seinen Extremalpunkten rekonstruieren kann durch Bildung der konvexen Hülle.

Satz 2.2.21 (Satz von Krein-Milman für Polyeder)**.** *Sei $P \subset \mathbb{R}^n$ ein beschränktes Polyeder. Dann gilt*

$$P = k(P_e).$$

Beweis. O.B.d.A. können wir annehmen, dass $P \neq \emptyset$. Da $P_e \subset P$ ist und P konvex, hat man sofort $k(P_e) \subset P$. Die umgekehrte Inklusion $P \subset k(P_e)$ zeigen wir induktiv. Für $n = 1$ ist P ein Intervall $[x, y]$, somit $P_e = \{x, y\}$ und daher

$$k(P_e) = k(\{x, y\}) = \{\lambda x + (1 - \lambda)y : \ 0 \leq \lambda \leq 1\} = [x, y] = P.$$

Nehmen wir an, die Behauptung gilt für n. Sei P ein beschränktes Polyeder in $\mathbb{R}^{n+1}$, sei $x \in P$ und sei $v \in \mathbb{R}^{n+1} \setminus \{0\}$. Da P beschränkt ist, ist

$$t_1 := \sup\{t \geq 0 : \ x + tv \in P\}$$

endlich, und da P abgeschlossen ist, hat man

$$x_1 := x + t_1 v \in P$$

(Frage: Warum genau ist das so ?). Ganz analog sieht man, dass

$$t_2 := \sup\{t \geq 0 : \ x - tv \in P\}$$

endlich ist und

$$x_2 := x - t_2 v \in P.$$

Umstellen nach v ergibt

$$\frac{1}{t_1}(x_1 - x) = v = \frac{1}{t_2}(x - x_2),$$

und Auflösen nach x zeigt, dass x eine Konvexkombination aus x_1 und x_2 ist,

$$x = \frac{t_2}{t_1 + t_2} x_1 + \frac{t_1}{t_1 + t_2} x_2 \quad \in \quad k(\{x_1, x_2\}).$$

Da P ein Polyeder ist, gibt es Linearformen $f_1, ..., f_m \neq 0$ auf $\mathbb{R}^{n+1}$ und Zahlen $c_1, .., c_m \in \mathbb{R}$ sodass

$$P = \bigcap_{i=1}^{m} \{f_i \leq c_i\}.$$

Somit ist $x + tv \in P$ genau dann, wenn $f_i(x) + t f_i(v) = f_i(x + tv) \leq c_i$ für alle $i = 1, ..., m$. Wegen der Definition von t_1 gibt es dann aber ein $i \in \{1, ..., m\}$ sodass

$$f_i(x_1) = c_i$$

(sonst Widerspruch), d.h. x_1 ist Element der Seite $P_i := P \cap \{f_i = c_i\}$ von P, und P_i ist selbst ein Polyeder wegen Satz 2.2.13. Weil die Mengen $\{f_i = c_i\}$ eine Hyperebene im $\mathbb{R}^{n+1}$ ist, ist P_i ein Polyeder im $\mathbb{R}^n$ (Reduktion auf n Koordinaten, siehe Übung). Daher gilt nach Induktionsvoraussetzung, dass

$$x_1 \in P_i = k((P_i)_e) = k(P_e \cap P_i) \subset k(P_e),$$

hier haben wir Satz 2.2.15 benutzt. Analog sieht man, dass $x_2 \in k(P_e)$. Nun folgt

$$x \in k(\{x_1, x_2\}) \subset k(k(P_e)) = k(P_e).$$

Da $x \in P$ beliebig war, zeigt das $P \subset k(P_e)$, die Behauptung für $n + 1$. □

Man kann diese Beobachtung noch verbessern und die Anzahl der Extremalpunkte in den Konvexkombinationen in $P = k(P_e)$ dimensionsabhängig beschränken.

Satz 2.2.22 (Satz von Carathéodory)**.** *Sei $P \subset \mathbb{R}^n$ ein beschränktes Polyeder. Dann ist jeder Punkt $x \in P$ Konvexkombination von höchstens $n + 1$ Punkten aus P_e.*

Beweis. Für $n = 1$ ist das klar. Nehmen wir an, die Behauptung gilt für n. Sei $P \subset \mathbb{R}^{n+1}$ ein beschränktes Polyeder und o.E. $x \in P \setminus P_e$. Wähle beliebiges $x' \in P_e$ und setze im vorangegangenen Beweis $v = x' - x$. Dann ist

$$x_1 = x' \in P_e, \quad \text{d.h. } t_1 = 1,$$

denn: Für $0 < t < 1$ hätte man

$$x + tv = x + t(x' - x) = tx' + (1 - t)x \in P,$$

deshalb muss $t_1 \geq 1$ sein (mit t_1 definiert wie im vorangegangenen Beweis), und falls $x + tv \in P$ wäre für ein $t > 1$, dann ergäbe sich

$$x' = \frac{t-1}{t}\, x + \frac{1}{t}\,(x + tv),$$

d.h. x' wäre Konvexkombi zweier Punkte aus P und könnte nicht extremal sein, deshalb muss auch $t_1 \leq 1$ gelten. Der Punkt $x_2 = x - t_2(x' - x)$ (mit t_2 definiert wie im vorangegangenen Beweis), liegt, wie wir (dort) gesehen hatten, auf einer Seite

$$S := P \cap \{f_i = c_i\}$$

von P, diese ist ein Polyeder im $\mathbb{R}^n$. Nach Induktionsvoraussetzung ist x_2 also Konvexkombi von höchstens $n+1$ Punkten aus $S_e = P_e \cap S \subset P_e$. Somit ist

$$x = \frac{t_2}{1+t_2}\, x' + \frac{1}{1+t_2}\, x_2$$

Konvexkombi von höchstens $n+2$ Punkten aus P_e. □

Bemerkung 2.2.23. Satz 2.2.21 und Satz 2.2.22 gelten für beliebige kompakte konvexe Mengen in $\mathbb{R}^n$.

Praktisch relevant ist folgende Konsequenz.

Korollar 2.2.24. *Sei $P \subset \mathbb{R}^n$ ein nichtleeres beschränktes Polyeder und f eine Linearform auf $\mathbb{R}^n$. Dann gibt es ein $x_0 \in P_e$ mit*

$$f(x) \leq f(x_0) \quad \textit{für alle } x \in P.$$

Das heisst, mindestens ein Extremalpunkt von P muss eine Maximalstelle für $f|_P$ sein.

Beweis, erste Version. Sei $\alpha := \sup_{x \in P} f(x)$ und $y_0 \in P$ so, dass $f(y_0) = \alpha$. Dann gibt es $x_1, ..., x_{n+1} \in P_e$ und $\lambda_i \geq 0$ mit $\sum_{i=1}^{n+1} \lambda_i = 1$, sodass

$$y_0 = \sum_{i=1}^{n+1} \lambda_i x_i$$

und daher

$$\alpha = f(y_0) = \sum_{i=1}^{n+1} \lambda_i f(x_i).$$

Das geht aber nur, wenn $f(x_i) = \alpha$ gilt für mindestens ein i. □

Beweis, zweite Version. Sei α wie oben. Dann ist $P' := P \cap \{f = \alpha\}$ (eine Seite von P) ein nichtleeres beschränktes Polyeder und somit $(P')_e \neq \emptyset$. Für $x_0 \in (P')_e = P' \cap P_e \subset P_e$ hat man

$$x_0 \in (P')_e \subset P' = P \cap \{f = \alpha\},$$

also $f(x_0) = \alpha$. □

Wir kommen zu folgendem Fazit.

Bemerkung 2.2.25. Die Menge aller Punkte in P, in denen f ihr Maximum auf P annimmt, ist eine Seite P' von P und (da $P' = k((P')_e) = k(P_e \cap P')$) die konvexe Hülle derjenigen Extremalpunkte von P, in denen das Maximum angenommen wird.

2.3 Bestimmung von Extremalpunkten

Die vorangegange Diskussion motiviert die Frage, wie man Extremalpunkte bestimmen kann.

Im Folgenden seien $f_1, ..., f_m$ nichttriviale Linearformen auf $\mathbb{R}^n$, $c_1, ..., c_m \in \mathbb{R}$ und

$$P = \bigcap_{i=1}^{m} \{f_i \leq c_i\}.$$

Die folgende Beobachtung reduziert das Auffinden von Extremalpunkten auf die Lösung linearer Gleichungssysteme.

Satz 2.3.1.

(i) Jeder Extremalpunkt x von P ist Lösung eines linearen Gleichungssystems

$$f_{i_k}(x) = c_{i_k}, \quad 1 \leq k \leq n, \tag{2.1}$$

mit n Gleichungen und n Unbekannten, wobei $1 \leq i_k \leq m$ und $f_{i_1}, ..., f_{i_n}$ linear unabhängige Linearformen sind. Insbesondere hat man $P_e = \emptyset$ falls $m < n$ und $\#P_e \leq \binom{m}{n}$ falls $m \geq n$.

(ii) Sind f_{i_k}, $1 \leq k \leq n$ linear unabhängig und ist die Lösung x von (2.1) in P enthalten, so ist x auch ein Extremalpunkt von P, d.h. $x \in P_e$.

Die zweite Aussage in (i) kann man sich gut geometrisch vorstellen: Für $n = 1$ braucht man mindestens eine Restriktion, also $m = 1$, um einen Extremalpunkt zu haben; hat man zwei Restriktionen, kann man bestenfalls zwei Extremalpunkte bekommen. Für $n = 2$ sind $m = 1$ Restriktionen zu wenig, um einen Extremalpunkt zu bekommen (denn dann ist das Polyeder ein Halbraum); für $m = 2$ ist (bei guter Konstellation) bereits einen Extremalpunkt möglich.

Ist $n = 2$ und P ein Dreieck, das als Schnitt dreier abgeschlossener Halbräume entsteht, also $m = 3$, so löst jeder Eckpunkt von P ein System (2.1) bestehend aus zwei Gleichungen (denn er ist Schnittpunkt von genau zwei der drei Geraden.)

Beweis. Wir zeigen (i). Nehmen wir an, dass $x \in P_e$ ist und schreiben wir $I := \{i \in \{1, ..., m\} : f_i(x) = c_i\}$. Es muss gelten, dass $I \neq \emptyset$, denn sonst wäre ja $f_i(x) < c_i$ für alle $i = 1, ..., m$ und damit x ein innerer Punkt, also $x \notin P_e$. Somit muss x eine Lösung des Gleichungssystems

$$f_i(z) = c_i, \quad i \in I, \tag{2.2}$$

sein. Der Punkt x ist sogar die einzige Lösung von (2.2):

Wäre dem nicht so, dann wäre der Lösungsraum mindestens eindimensional, also würde es ein $v \in \mathbb{R}^n \setminus \{0\}$ geben sodass die Gerade $G = \{x + tv : \ t \in \mathbb{R}\}$ in der Lösungsmenge enthalten ist. Falls $I = \{1, ..., m\}$ dann ist G offensichtlich in P enthalten. Ansonsten sind zumindest für hinreichend kleines $t > 0$ die Punkte $x \pm tv$ in P: Man hat $f_i(x) < c_i$ für alle $i \in \{1, ..., m\} \setminus I$, und wegen der Stetigkeit der f_i gibt es ein (hinreichend kleines) $t > 0$ sodass auch $f_i(x \pm tv) < c_i$ für alle $i \in \{1, ..., m\} \setminus I$, und zusammen mit (2.2) folgt $x \pm tv$ in P. Weil aber

$$x = \frac{1}{2}(x + tv) + \frac{1}{2}(x - tv)$$

ist, kann dann x kein Extremalpunkt von P sein.

Da nun also x die eindeutige Lösung von (2.2) sein muss, so muss der Rang des Gleichungssystems (2.2) maximal sein, also gleich n. Somit gibt es unter den f_i, $i \in I$, auf jeden Fall n linear unabhängige Linearformen $f_{i_1}, ..., f_{i_n}$, und x ist auch die eindeutige Lösung von (2.1).

Schliesslich beobachtet man noch die naive Abschätzung

$$\begin{aligned} \#\text{Extremalpunkte} &\leq \#\text{ eindeutig lösbare Gleichungssyteme vom Typ (2.1)} \\ &\leq \#\text{ Möglichkeiten, } n \text{ Linearformen aus } m \text{ auszuwählen} \\ &= \binom{m}{n}. \end{aligned}$$

Wir zeigen (ii). Seien $f_{i_1}, ..., f_{i_n}$ linear unabhängig und sei $x \in P$ die eindeutige Lösung von (2.1). Angenommen $y, z \in P$ und $0 < \lambda < 1$ sind so, dass $x = \lambda y + (1 - \lambda)z$. Dann gilt

$$c_{i_k} = f_{i_k}(x) = \lambda f_{i_k}(y) + (1 - \lambda) f_{i_k}(z),$$

und das geht nur, wenn $f_{i_k}(y) = f_{i_k}(z) = c_{i_k}$ für alle $1 \leq k \leq n$ (wenn $f_{i_k}(y)$ oder $f_{i_k}(z)$ streng kleiner als c_{i_k} sind, kann die Gleicheit nicht gelten). Das bedeutet aber, dass auch y und z das Gleichungssystem (2.1) lösen, und wegen Eindeutigkeit muss man dann $x = y = z$ haben. Das zeigt, dass x ein Extremalpunkt ist. □

Bemerkung 2.3.2. Satz 2.3.1 zeigt, wie man Extremalpunkte eines Polyeders P finden kann: Wir wählen $1 \leq i_1 < i_2 < ... < i_n \leq m$ und lösen das Gleichungssystem

$$f_{i_k}(x) = c_{i_k}, \quad 1 \leq k \leq n.$$

Ist x die eindeutige Lösung dieses Systems und gehört x zu P, so muss x ein Extremalpunkt von P sein, also $x \in P_e$. Auf diese Weise kann man *alle* Extremalpunkte von P finden.

Das folgende Korollar zeigt, dass man für lineares Optimierungsproblem a priori keinen der Extremalpunkte ausschliessen kann.

Korollar 2.3.3. *Sei $x \in P_e$. Dann gibt es eine Linearform f auf $\mathbb{R}^n$ mit*

$$f(y) < f(x) \quad \textit{für alle } y \in P \setminus \{x\}.$$

Beweis. Sei $x \in P_e$ und seien $f_{i_1}, ..., f_{i_n}$ linear unabhängig und so, dass x die eindeutige Lösung von

$$f_{i_k}(x) = c_{i_k}, \quad 1 \leq k \leq n. \tag{2.3}$$

Wir behaupten, dass dann mit $f := f_{i_1} + ... + f_{i_n}$ die Behauptung folgt. Um das zu sehen, bemerken wir, dass jedes $y \in P \setminus \{x\}$ die Ungleichungen

$$f_{i_k}(y) \leq c_{i_k}, \quad 1 \leq k \leq n,$$

erfüllt, und dass es dabei wegen der Eindeutigkeit der Lösung von (2.3) mindestens ein $k \in \{1, ..., n\}$ geben muss mit

$$f_{i_k}(y) < c_{i_k}.$$

Dann folgt aber, dass

$$f(y) = f_{i_1}(y) + ... + f_{i_n}(y) < c_{i_1} + ... + c_{i_n} = f_{i_1}(x) + ... + f_{i_n}(x) = f(x),$$

wie gewünscht. □

Bemerkung 2.3.4. Geometrisch bedeutet Korollar 2.3.3, dass es zu $x \in P_e$ einen offenen Halbraum H gibt (nämlich $H = \{f < c_{i_1} + ... + c_{i_n}\}$ in der Notation des Beweises), sodass $P \setminus \{x\} \subset H$ und $x \in \partial H$.

Man kann mit linearen Gleichungen auch die Lage von Extremalpunkten relativ zueinander beschreiben.

Definition 2.3.5. Zwei verschiedene Extremalpunkte x und y eines Polyeders P heissen *benachbart*, falls $[x, y]$ eine Seite von P ist.

Satz 2.3.6. *Sei $x \in P_e$ und seien $1 \leq i_1 < i_2 < ... < i_n \leq m$ so, dass $f_{i_1}, ..., f_{i_n}$ linear unabhängig sind und*

$$f_{i_k}(x) = c_{i_k}, \quad 1 \leq k \leq n.$$

Ist $y \in P_e \setminus \{x\}$ und gilt

$$f_{i_k}(y) = c_{i_k} \quad \text{für } n-1 \text{ der möglichen Werte von } k,$$

so sind x und y benachbart.

Wenn wir uns wieder den Fall eines Dreiecks P in der Ebene ($n = 2$) vorstellen, wird die Grundidee plausibel: Zwei verschiedene Eckpunkte sind (in Fall des Dreiecks) benachbart, es gibt also genau eine Gerade (hier in der Rolle einer Hyperebene), auf der beide liegen. Also muss genau eine der jeweils zwei Gleichungen, die diese Punkte beschreiben, von beiden Punkten gelöst werden (nämlich jene, die die Gerade beschreibt). Satz 2.3.6 schliesst nun von den Gleichungen auf die geometrische Lage.

Beweis. O.B.d.A. nehmen wir an, dass $i_k = k$ (sonst relabeln) und $f_i(y) = c_i$ für $1 \leq i \leq n-1$ und $f_n(y) < c_n$. Da y Extremalpunkt ist, gibt es nach Satz 2.3.1 n linear unabhängige Linearformen $f_{j_1}, ..., f_{j_n}$ mit

$$f_{j_k}(y) = c_{j_k}, \quad 1 \leq k \leq n.$$

Davon muss mindestens ein f_{j_k} linear unabhängig von den $f_1, ..., f_{n-1}$ sein, d.h. es gibt ein $j > n$ sodass $f_1, ..., f_{n-1}, f_j$ linear unabhängig sind und

$$f_1(y) = c_1, \quad ... \ , \quad f_{n-1}(y) = c_{n-1}, \quad f_j(y) = c_j.$$

Da die Lösung y dieses Systems eindeutig ist und $x \neq y$, so folgt

$$f_j(x) < c_j.$$

Um zu zeigen, dass $[x, y]$ eine Seite von P ist, seien $z_1, z_2 \in P$ und $0 < \lambda < 1$ und

$$z = \lambda z_1 + (1 - \lambda) z_2 \in [x, y]. \tag{2.4}$$

Wir behaupten, dass

$$z_1, z_2 \in [x, y]. \tag{2.5}$$

Da $f_i(x) = f_i(y) = c_i$, $i = 1, ..., n-1$, so folgt, dass $f_i = c_i$ auf $[x, y]$ gilt und insbesondere $f_i(z) = c_i$, $i = 1, .., n-1$. Da $f_i \leq c_i$ gilt auf ganz P, so folgt wegen (2.4), dass

$$f_i(z_1) = f_i(z_2) = c_i \quad \text{sein muss für alle } 1 \leq i \leq n-1.$$

Die Punkte z_1 und z_2 gehören also zu der Menge

$$G := \bigcap_{i=1}^{n-1} \{f_i = c_i\},$$

und da die $f_1, ..., f_{n-1}$ linear unabhängig sind, ist G eine Gerade. Sie enthält aber auch x und y, also

$$G = \{\alpha x + (1-\alpha)y : \ \alpha \in \mathbb{R}\}.$$

Um (2.5) zu sehen, genügt es nun zu zeigen, dass jeder Punkt $\widetilde{z} \in P \cap G$ eine Darstellung

$$\widetilde{z} = \alpha x + (1-\alpha)y \quad \text{mit } 0 \leq \alpha \leq 1$$

hat. In der Tat muss $\alpha \leq 1$ gelten, denn

$$c_n \geq f_n(\widetilde{z}) = \alpha f_n(x) + (1-\alpha) f_n(y) = \alpha c_n + (1-\alpha) f_n(y),$$

also

$$(1-\alpha)c_n \geq (1-\alpha) f_n(y),$$

d.h.

$$(1-\alpha)(c_n - f_n(y)) \geq 0,$$

und da $c_n - f_n(y) > 0$ ist, geht das nur mit $1-\alpha \geq 0$. Ganz analog sieht man, dass $\alpha \geq 0$ sein muss, denn

$$c_j \geq f_j(\widetilde{z}) = \alpha f_j(x) + (1-\alpha) f_j(y) = \alpha f_j(x) + (1-\alpha) c_j,$$

also am Ende

$$\alpha(c_j - f_j(x)) \geq 0,$$

wobei $c_j - f_j(x) > 0$ ist. □

Umgekehrt gilt nun folgender Schluss von der geometrischen Lage auf die Gleichungen.

Satz 2.3.7. *Seien x und y benachbarte Extremalpunkte eines Polyeders P. Dann gibt es n linear unabhängige Linearformen $f_{i_1}, ..., f_{i_n}$ mit*

$$f_{i_k}(x) = c_{i_k}, \quad 1 \leq k \leq n,$$

und

$$f_{i_k}(y) = c_{i_k} \quad \textit{für } n-1 \textit{ der möglichen Werte von } k.$$

Beweis. Seien $I_x = \{i : f_i(x) = c_i\}$ und $I_y := \{i : f_i(y) = c_i\}$ und sei $I := I_x \cap I_y$. Wir nehmen an, dass der Rang von $(f_i)_{i \in I}$ strikt kleiner ist als $n-1$ und behaupten, dass dann $[x, y]$ keine Seite sein kann.

Die Annahme impliziert, dass $\bigcap_{i \in I} \{f_i = c_i\}$ die Punkte x und y enthält und mindestens zweidimensional ist, also insbesondere eine affine Ebene E enthält, deren Elemente x und y sind.

Sei nun $z = \frac{1}{2}(x+y)$. In der affinen Ebene E können wir nun eine Gerade G durch z wählen, die weder x noch y enthält. Nun gilt für $i \neq I$, dass

$$f_i(z) = \frac{1}{2}(f_i(x) + f_i(y)) < c_i,$$

und weil die f_i stetig sind, existieren $z_1, z_2 \in G \setminus \{z\}$ mit $z = \frac{1}{2}(z_1 + z_2)$ und so, dass

$$f_i(z_1) < c_i \quad \text{und} \quad f_i(z_2) < c_i \quad \text{für alle } i \notin I.$$

Wie früher können wir aber andererseits auch schliessen, dass

$$f_i(z_1) = f_i(z_2) = c_i \quad \text{für alle } i \in I.$$

Aus diesen beiden Fakten folgt, dass $z_1, z_2 \in P$. Da $z_1, z_2 \in G \setminus \{z\}$, hat man auch $z_1, z_2 \notin [x, y]$. Weil aber $z \in [x, y]$ ist, kann $[x, y]$ dann keine Seite sein.

Also muss der Rang von $(f_i)_{i \in I}$ grösser gleich $n - 1$ sein. Es genügt nun, $n - 1$ linear unabhängige dieser f_i zu wählen und durch eine dazu linear unabhängige Linearform f_j mit $f_j(x) = c_j$ zu ergänzen. Ein solches f_j existiert, da der Rang von $(f_i)_{i \in I_x}$ gleich n ist (denn x ist ein Extremalpunkt). Der Punkt x ist dann die eindeutige Lösung des zugehörigen Gleichungssystems, und für y gilt $f_j(y) < c_j$. □

Bemerkung 2.3.8. Im Allgemeinen kann man nicht hoffen, mithilfe nur einer *festen* Familie $\{f_{i_k}\}_{1 \leq k \leq n}$ alle zu einem gegebenen Punkt $x \in P_e$ benachbarten Extremalpunkte zu finden.

Im Folgenden wollen wir den Begriff 'benachbart' für Extremalpunkte benutzen, um leichter entscheiden zu können, ob ein Extremalpunkt tatsächlich eine Maximalstelle für eine gegebene Linearform auf dem gesamten Polyeder ist.

Eine konvexe Menge $C \subset \mathbb{R}^n$ die für alle $\lambda \geq 0$ die Menge

$$\lambda C := \{\lambda x : x \in C\}$$

enthält, nennt man einen *konvexen Kegel*. Für eine gegebene konvexe Menge $K \subset \mathbb{R}^n$ ist

$$\mathbb{R}_+ K := \{\lambda x : \lambda \geq 0,\ x \in K\}$$

eine konvexer Kegel, tatsächlich der kleinste konvexe Kegel, der K enthält. (Idee: 'Unendliche Schultüte', die K fest umschliesst.)

Wir machen folgende Beobachtung. Mit 0 bezeichnen wir den Koordinatenursprung, d.h. den Nullvektor in $\mathbb{R}^n$.

Lemma 2.3.9. *Sei P ein beschränktes Polyeder mit $0 \in P_e$, aber $P \neq \{0\}$, und bezeichne P_e^0 die Menge aller zu 0 benachbarten Extremalpunkte von P. Dann hat man*

$$P \subset \mathbb{R}_+ k(P_e^0).$$

In Worten: Das Polyeder P ist eine Teilmenge des kleinsten konvexen Kegels $\mathbb{R}_+ k(P_e^0)$, welcher die konvexe Hülle $k(P_e^0)$ der Menge P_e^0 der zu 0 benachbarten Extremalpunkte enthält.

Bemerkung 2.3.10. Man hat

$$\begin{aligned} \mathbb{R}_+ k(P_e^0) &= \{\lambda x :\ \lambda \geq 0,\ x \in k(P_e^0)\} \\ &= \left\{ \sum_{i=1}^{m} \lambda_i y_i :\ m \in \mathbb{N},\ \lambda_i \geq 0,\ y_i \in P_e^0 \right\}. \end{aligned}$$

Bevor wir Lemma 2.3.9 beweisen, schauen wir uns folgende praktisch relevante Konsequenz an.

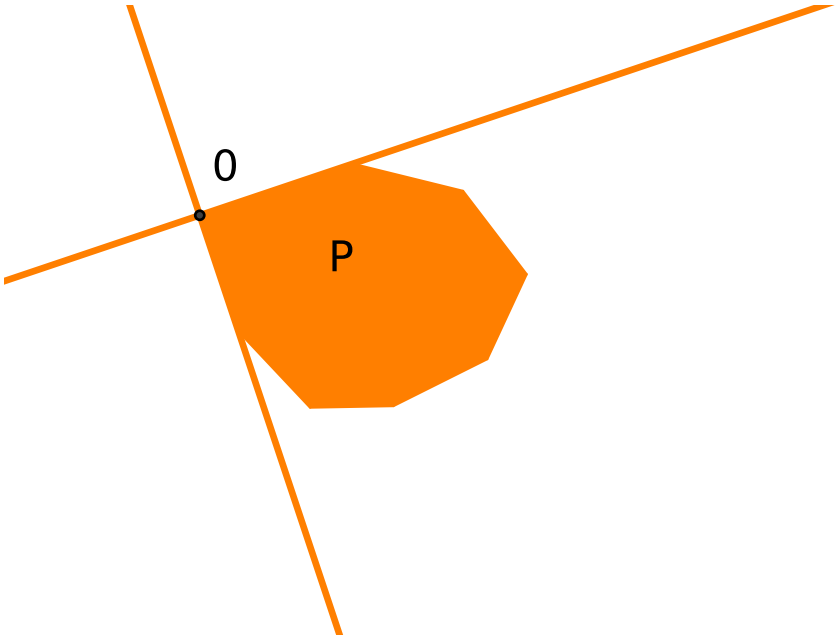


Abbildung 2.4: Polyeder als Teilmenge des Kegels

Korollar 2.3.11. *Sei P ein beschränktes Polyeder und f eine Linearform auf $\mathbb{R}^n$. Ist $x \in P_e$ und gilt $f(y) \leq f(x)$ für alle $y \in P_e$, die zu x benachbart sind, so folgt, dass*

$$f(x) = \max f(P)$$

ist.

Das bedeutet, um zu entscheiden, ob ein Extremalpunkt x eine Maximalstelle von $f|_P$ ist, müssen wir lediglich $f(x)$ mit den Werten $f(y)$ in *benachbarten* Extremalpunkten y vergleichen. In der Praxis hat man es oft mit sehr vielen Restriktionen zu tun und daher mit sehr vielen Extremalpunkten. Das Korollar kommt uns entgegen und reduziert die Komplexität der Frage, ob x Maximalstelle von $f|_P$ ist oder nicht.

Beweis. O.B.d.A. können wir annehmen, dass $x = 0$ ist (ansonsten betrachte Translation $\widetilde{P} := P - x$). Da f linear ist, hat man $f(x) = 0$. Sei nun $z \in P$. Nach Lemma 2.3.9 gibt es $y_1, ..., y_m \in P_e^0$ und $\lambda_1, ..., \lambda_m \geq 0$ mit

$$z = \sum_{i=1}^{m} \lambda_i y_i.$$

Also ergibt sich

$$f(z) = \sum_{i=1}^{m} \lambda_i f(y_i) \leq 0 = f(x),$$

denn nach Voraussetzung gilt $f(y_i) \leq f(x) = 0$, $i = 1, ..., m$. □

Wir beweisen Lemma 2.3.9.

Beweis. Wir wollen zeigen, dass $P \subset \mathbb{R}_+ k(P_e^0)$. Die Idee ist, ein Polyeder $\widetilde{P}$ zu konstruieren, welches P enthält, und für welches man 'bequem' zeigen kann, dass $\widetilde{P}_e \subset \mathbb{R}_+ P_e^0$ gilt. (Dann hat man den 'kritischen Schritt' auf den konvexen Kegel geschafft.)

Seien $f_1, ..., f_m$ Linearformen und $c_1, ..., c_m \in \mathbb{R}$ Zahlen sodass

$$P = \bigcap_{i=1}^{m} \{f_i \leq c_i\}.$$

Wegen $0 \in P$ muss $c_i \geq 0$ gelten für alle $i = 1, ..., m$ (denn $f_i(0) = 0$). Sei

$$I := \{i \in \{1, ..., m\} : \ c_i = 0\}.$$

Da $0 \in P_e$, so gibt es nach Satz 2.3.1 (i) Indizes $i_1, ..., i_n \in I$ und linear unabhängige Linearformen $f_{i_1}, ..., f_{i_n}$ sodass 0 die eindeutige Lösung des Systems

$$f_{i_k}(z) = 0, \quad k = 1, ..., n,$$

ist. Also existiert zu jedem $y \in P \setminus \{0\}$ ein $i \in I$ mit $f_i(y) < 0$. Wir definieren eine Linearform g durch

$$g := -\sum_{i \in I} f_i,$$

und offensichtlich gilt $g > 0$ auf $P \setminus \{0\}$. Weil P kompakt ist, existiert

$$\beta := \max_P g.$$

Wir definieren ein Polyeder

$$\widetilde{P} := \bigcap_{i \in I} \{f_i \leq 0\} \cap \{g \leq \beta\}.$$

Offensichtlich $P \subset \widetilde{P}$. Wir zeigen nun, dass $\widetilde{P}$ beschränkt ist:

Sei $x \in \widetilde{P}$. Dann ist

$$f_i(\lambda x) = \lambda \ f_i(x) \leq \begin{cases} 0 & \text{falls } i \in I \text{ und } \lambda > 0 \text{ und} \\ c_i & \text{falls } i \notin I \text{ und } \lambda > 0 \text{ hinreichend klein,} \end{cases}$$

und somit ist $\lambda x \in P$ für hinreichend kleine $\lambda > 0$. Das impliziert aber, dass $x \in \mathbb{R}_+ P$, und kombiniert mit Satz 2.2.21 zeigt das, dass x sich darstellen lässt in der Form

$$x = \sum_{x_i \in P_e} \lambda_i x_i = \sum_{x_i \in P_e \setminus \{0\}} \lambda_i x_i \quad \text{mit passenden } \lambda_i \geq 0.$$

Nun ist

$$\alpha := \min \big\{ g(x_i) : x_i \in P_e \setminus \{0\} \big\} > 0,$$

da $g > 0$ auf $P \setminus \{0\}$. Also

$$\sum_{i : x_i \in P_e \setminus \{0\}} \lambda_i \leq \sum_{i : x_i \in P_e \setminus \{0\}} \lambda_i \, \frac{g(x_i)}{\alpha} = \frac{1}{\alpha} \, g \left(\sum_{i : x_i \in P_e \setminus \{0\}} \lambda_i x_i \right) = \frac{1}{\alpha} \, g(x) \leq \frac{\beta}{\alpha}$$

und daher

$$\|x\| \leq \sum_{i : x_i \in P_e \setminus \{0\}} \lambda_i \, \|x_i\| \leq \max_{i : x_i \in P_e \setminus \{0\}} \|x_i\| \cdot \sum_{i : x_i \in P_e \setminus \{0\}} \lambda_i \leq \frac{\beta}{\alpha} \max_{i : x_i \in P_e \setminus \{0\}} \|x_i\| < +\infty,$$

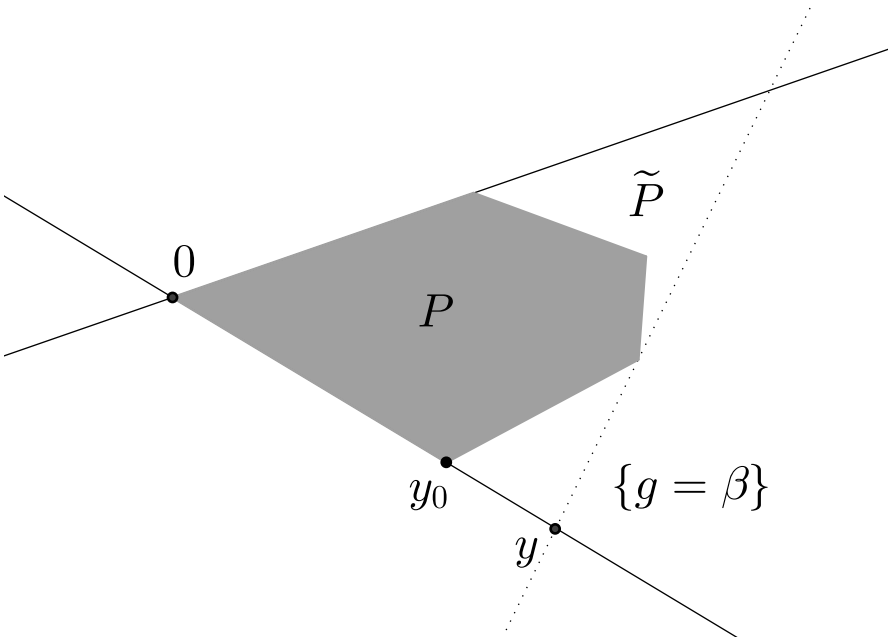


Abbildung 2.5: Polyeder P und $\widetilde{P}$

und weil weder α noch β von x abhängen, folgt damit die Beschränktheit von $\widetilde{P}$.

Sei nun $y \in \widetilde{P}_e$, $y \neq 0$. Setzen $t_0 := \sup\{t \geq 0 : ty \in P\}$ und $y_0 := t_0 y$. Dann ist $t_0 \leq 1$, da $P \subset \widetilde{P}$ und $t_0 > 0$, weil für $i = 1, ..., m$ gilt, dass

$$f_i(ty) = tf_i(y) \leq \begin{cases} 0 & \text{falls } i \in I, \text{ da } y \in \widetilde{P} \text{ und} \\ c_i & \text{falls } i \notin I \text{ und } t > 0 \text{ hinreichend klein,} \end{cases}$$

also $ty \in P$ für hinreichend kleines $t > 0$.

Weil $y \in \widetilde{P}_e$, $y \neq 0$, so ist y die eindeutige Lösung eines linearen Gleichungssystems aus n linear unabhängigen Gleichungen des Typs

$$f_i(y) = 0, \quad \text{wobei } i \in I, \text{ bzw.} \quad g(y) = \beta,$$

und weil $y \neq 0$ ist, muss eine der Gleichungen tatsächlich $g(y) = \beta$ sein. Dann ist aber nach Satz 2.3.6 der Punkt y ein in $\widetilde{P}$ zu 0 benachbarter Extremalpunkt, und $[0, y]$ ist eine Seite von $\widetilde{P}$. Wir behaupten nun, dass dann $[0, y_0]$ eine Seite von P ist:

Sei $\lambda z_1 + (1-\lambda) z_2 \in [0, y_0]$ für zwei Punkte $z_1, z_2 \in P$ und $0 < \lambda < 1$. Da $[0, y_0] \subset [0, y]$, $z_1, z_2 \in \widetilde{P}$ und $[0, y]$ Seite von P ist, hat man $z_1, z_2 \in [0, y]$. Also ist klar, dass $z_1, z_2 \in [0, y]$, mit anderen Worten, $z_1 = t_1 y$ und $z_2 = t_2 y$ für gewisse $0 \leq t_1, t_2 \leq 1$. Wegen der Definition von t_0 folgt aber $t_1, t_2 \leq t_0$, was $z_1, z_2 \in [0, y_0]$ forciert und somit zeigt, dass $[0, y_0]$ eine Seite von P ist.

Somit gilt also $y_0 \in [0, y_0]_e \subset P_e$, ferner $y_0 \neq 0$ wegen $t_0 > 0$, und $[0, y_0]$ ist Seite von P. Das heisst aber, dass $y_0 \in P_e^0$ ist.

Wir haben damit gezeigt, dass jeder Punkt $y \in \widetilde{P}_e \setminus \{0\}$ ein Punkt von $\mathbb{R}_+ P_e^0$ ist. Weil das offensichtlich auch für 0 stimmt, wissen wir nun, dass

$$\widetilde{P}_e \subset \mathbb{R}_+ P_e^0.$$

Der 'showdown' ist nun sehr bequem: Mit Satz 2.2.21 und der Monotonie der konvexen Hülle folgt nun

$$P \subset \widetilde{P} = k(\widetilde{P}_e) \subset k(\mathbb{R}_+ P_e^0) = \mathbb{R}_+ k(P_e^0).$$

□

2.4 Auflösen linearer Gleichungssysteme

Wir kommen zurück zu uns wohlbekannten linearen Gleichungssystemen und zu Strategien, sie zu lösen. Der Aspekt, den wir hier kurz diskutieren wollen, ist der prozedurale (algorithmische), und die Methode, die wir kurz anschauen, ist das klassische *Gaussverfahren* (Austauschverfahren), eines der Basics in der numerischen Mathematik.

(Mathematisch ist hier nichts neu für Sie, es geht einfach um eine clevere 'Buchhaltung', die man auch einfach als Code implementieren kann.)

Wir betrachten das lineare Gleichungssystem

$$\begin{aligned} a_{11}\,x_1 + \cdots + a_{1n}\,x_n &= c_1 \\ a_{21}\,x_1 + \cdots + a_{2n}\,x_n &= c_2 \\ &\vdots \\ a_{m1}\,x_1 + \cdots + a_{mn}\,x_n &= c_m \end{aligned}$$

mit gegebenen $a_{ij}, c_i \in \mathbb{R}$.

Für $x = (x_1, ..., x_n) \in \mathbb{R}^n$ und $1 \le i \le m$ setzen wir

$$y_i(x) := -a_{i1}\,x_1 - \cdots - a_{in}\,x_n + c_i. \tag{2.6}$$

Das definiert jeweils eine affine Abbildung $y_i : \mathbb{R}^n \to \mathbb{R}$, und im Falle $c_i = 0$ eine Linearform. Die Lösungsmenge des Gleichungssystems ist dann

$$\{x \in \mathbb{R}^n :\ y_1(x) = \cdots = y_m(x) = 0\}\,.$$

Die Grundidee für einen Lösungsalgorithmus ist nun, die $x_1, ..., x_n$ durch äquivalente Umformungen als Funktionen der $y_1, ..., y_m$ auszudrücken und dann $y_1, ..., y_m$ gleich null zu setzen. Man verwendet dabei die Darstellung der Relationen (2.6) durch folgendes Schema:

	$-x_1$	$\cdots$	$-x_n$	
y_1	a_{11}	$\cdots$	a_{1n}	c_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	c_m

(2.7)

Ein *Gauss-Eliminationsschritt* besteht in folgendem:

1. Wähle ein (k, l) mit $a_{kl} \neq 0$. Ein solches a_{kl} nennt man ein *Pivotelement*.
2. Eliminiere x_l durch Auflösen der k-ten Gleichung nach x_l:

$$y_k = -a_{kl}\,x_l - \sum_{j \neq l} a_{kj}\,x_j + c_k$$

genau dann, wenn

$$x_l = -\frac{1}{a_{kl}}\,y_l - \sum_{j \neq l} \frac{a_{kj}}{a_{kl}}\,x_j + \frac{c_k}{a_{kl}}.$$

3. Eliminiere x_l aus den anderen Gleichungen durch Einsetzen dieses Ausdrucks für x_l. Für $i \neq k$ ergibt sich

$$\begin{aligned} y_i &= -a_{il}\, x_l - \sum_{j \neq l} a_{ij}\, x_j + c_i \\ &= \frac{a_{il}}{a_{kl}}\, y_k - \sum_{j \neq k} \left(a_{ij} - \frac{a_{kj} a_{il}}{a_{kl}} \right) x_j + c_i - \frac{c_k a_{il}}{a_{kl}}. \end{aligned}$$

Dies führt zu folgendem, zu (2.7) äquivalenten Schema:

	$-x_1$	$\cdots$	$-y_k$	$\ldots$	$-x_n$	
y_1						
$\vdots$						$\vdots$
x_l	$\frac{a_{k1}}{a_{kl}} \ldots$	$\ldots$	$\frac{1}{a_{kl}}$	$\ldots$	$\frac{a_{kn}}{a_{kl}}$	$\frac{c_k}{a_{kl}}$
$\vdots$						$\vdots$
y_i	$a_{i1} - \frac{a_{k1}a_{il}}{a_{kl}} \ldots$	$\ldots$	$-\frac{a_{il}}{a_{kl}}$	$\ldots$	$a_{in} - \frac{a_{kn}a_{il}}{a_{kl}} \ldots$	$c_i - \frac{c_k a_{il}}{a_{kl}}$
$\vdots$						$\vdots$
y_m						

(2.8)

Die k-te Zeile (die *Pivotzeile*) ist die Zeile, die mit x_l beginnt. Die l-te Spalte (die *Pivotspalte*) ist jene, die oben $-y_k$ stehen hat.

Das neue Schema (2.8) wird dabei nach folgenden *Pivotregeln* für den Austausch von K-ter Zeile und l-ter Spalte gebildet:

(i) Das Pivotelement wird durch seinen Kehrwert ersetzt.

(ii) Die übrigen Elemente der Pivotzeile werden durch das Pivotelement geteilt.

(iii) Die übrigen Elemente der Pivotspalte werden durch das Pivotelement dividiert, und das Vorzeichen wird geändert.

(iv) Alle übrigen Elemente werden nach der folgenden *Rechteckregel* verändert: Ersetze in

$$\begin{array}{ccc} * & \ldots & a \\ \vdots & & \vdots \\ b & \ldots & \text{Pivot} \end{array}$$

den Eintrag $*$ durch

$$* \; - \; \frac{ab}{\text{Pivot}} \; = \; * \; + \; \frac{-a}{\text{Pivot}}\, b.$$

Man kann das nun wie folgt 'optisch vereinfachen': Der Quotient $\frac{-a}{\text{Pivot}}$ ist derjenige Eintrag, der im neuen Schema (2.8) in der l-ten Spalte (der Pivotspalte) auf der Höhe von $*$ steht. Es empfiehlt sich daher, als Zwischenschritt zunächst die neue Pivotspalte (ausser dem Kehrwert des Pivotelements) neben das alte Schema zu schreiben:

	$-x_1$	$\cdots$	$-x_n$		
y_1	a_{11}	$\cdots$	a_{1n}	c_1	$-\frac{a_{1l}}{a_{kl}}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	c_m	$-\frac{a_{ml}}{a_{kl}}$

Dann ändert sich die Rechteckregel (iv) zur Berechnung der Elemente ausserhalb von Pivotzeile und -spalte zur folgenden *Dreiecksregel*: Ersetze in

$$\begin{matrix} * & \dots & d := \frac{-a}{\text{Pivot}} \\ \vdots & & \\ b & & \end{matrix}$$

den Eintrag $*$ durch

$$* \ + \ db.$$

(Hier ist die letzte Spalte die neue, wie oben hinzugefügte Pivotspalte, und b ist das Element aus der Pivotzeile, das in derselben Spalte wie $*$ steht.)

Die neue Pivotspalte kann natürlich auch gleich an ihren Platz im neuen Schema geschrieben werden.

Beispiel 2.4.1. Das lineare Gleichungssystem

$$\begin{aligned} 2\,x_1 + 4\,x_2 + x_3 &= -4 \\ 3\,x_1 + 5\,x_2 + x_3 &= 3 \\ x_1 + x_2 + x_3 &= 0 \end{aligned}$$

führt auf das (erweiterte) Schema

	$-x_1$	$-x_2$	$-x_3$		
y_1	2	4	1	-4	-4
y_2	3	5	1	3	-5
y_3	1	1	1	0	

Als Pivot haben wir hier das Element $a_{32} = 1$ ausgewählt. Die neue Pivotspalte (abgesehen vom Kehrwert des Pivots) haben wir rechts hinzugefügt. Zum Beispiel ergibt sich

$$-\frac{a_{12}}{a_{32}} = -\frac{4}{1} = -4.$$

Wir tauschen nun die Bezeichnungen der 3. Zeile und der 2. Spalte (abgesehen vom Vorzeichen) aus und berechnen die übrigen Elemente (ausserhalb der Pivotzeile und -spalte) mit der Dreiecksregel. Zum Beispiel wird aus

$$* = a_{11} = 2$$

mit $b = a_{31} = 1$ und $d = -4$ der Eintrag

$$* + bd = 2 + 1 \cdot (-4) = -2.$$

Wir erhalten das neue Schema

	$-x_1$	$-y_3$	$-x_3$	
y_1	-2	-4	-3	-4
y_2	-2	-5	-4	3
x_2	1	1	1	0

Die Elemente der Pivotspalte werden nun nicht weiter benutzt, da für die Lösung am Ende $y_1 = y_2 = y_3 = 0$ gesetzt wird. Für den nächsten Gauss-Eliminationsschritt genügt daher das (erweiterte) Schema

	$-x_1$	$-y_3$	$-x_3$		
y_1	-2	#	-3	-4	
y_2	-2	#	-4	3	-1
x_2	1	#	1	0	0.5

Hier haben wir als neues Pivot -2 gewählt. Wir folgen nun wie zuvor den Pivotregeln und erhalten als nächstes Schema

	$-y_1$	$-y_3$	$-x_3$	
x_1	-0.5	#	1.5	2
y_2	-1	#	-1	7
x_2	0.5	#	-0.5	-2

Für den folgenden Eliminationsschritt genügt das (erweiterte) Schema

	$-y_1$	$-y_3$	$-x_3$		
x_1	#	#	1.5	2	1.5
y_2	#	#	-1	7	
x_2	#	#	-0.5	-2	-0.5

Als neues Pivot haben wir -1 gewählt. Mit den Pivotregeln ergibt sich nun das Schema

	$-y_1$	$-y_3$	$-y_2$	
x_1	#	#	1.5	$\frac{25}{2}$
x_3	#	#	-1	-7
x_2	#	#	-0.5	$-\frac{11}{2}$

Jetzt hat man also

	$-y_1$	$-y_3$	$-y_2$	
x_1	#	#	#	$\frac{25}{2}$
x_3	#	#	#	-7
x_2	#	#	#	$-\frac{11}{2}$

und damit die Lösung $x_1 = \frac{25}{2}$, $x_2 = -\frac{11}{2}$ und $x_3 = -7$ des Gleichungssystems.

Das Gauss-Eliminationsverfahren kann man auch zur Rangbestimmung für $(m \times n)$-Matrizen benutzen oder zum Invertieren regulärer $(n \times n)$-Matrizen.

Beispiel 2.4.2. Wir bestimmen den Rang der $(n \times m)$-Matrix mit $n = 4$ und $m = 5$ im Schema

	$-x_1$	$-x_2$	$-x_3$	$-x_4$
y_1	4	3	2	1
y_2	3	-1	0	2
y_3	5	7	4	0
y_4	7	2	2	3
y_5	5	-6	-2	5

(Hier gibt es keine Spalte für Zahlen c_i, denn die sind ja nicht Teil dieser Aufgabenstellung.) Als Pivot ist hier 1 ausgewählt.

Austausch von y_1 und $-x_4$ nach den Pivotregeln ergibt das neue Schema

	$-x_1$	$-x_2$	$-x_3$	$-y_1$
x_4	4	3	2	1
y_2	-5	-7	-4	-2
y_3	5	7	4	0
y_4	-5	-7	-4	-3
y_5	-15	-21	-12	-5

Wie wir Pivotzeile und -spalte bekommen, ist sofort sichtbar. Die anderen Einträge folgen wieder aus der Dreiecksregel, z.B. ergibt sich für den Eintrag a_{21} der neue Wert $3+4(-2) = -5$ und für den Eintrag a_{43} der neue Wert $2 + 2(-3) = -4$. Als neues Pivot wählen wir nun -5.

Austausch von y_2 und x_1 nach den Pivotregeln ergibt nun das neue Schema

	$-y_2$	$-x_2$	$-x_3$	$-y_1$
x_4	$\frac{4}{5}$	$\frac{-13}{5}$	$\frac{-6}{5}$	$\frac{-3}{5}$
x_1	$\frac{-1}{5}$	$\frac{7}{5}$	$\frac{4}{5}$	$\frac{2}{5}$
y_3	1	0	0	-2
y_4	-1	0	0	-1
y_5	-3	0	0	1

Nun bricht das Verfahren aber ab: An allen Positionen, an denen neue Pivotelemente stehen könnten (d.h. für welche ein Tausch der Zeilen- und Spaltenbezeichnung strategisch Sinn machen würden), stehen Nullen.

Die Linearformen y_3, y_4 und y_5 sind Linearkombinationen von y_1 und y_2: Man hat, wie sich ablesen lässt,

$$\begin{aligned} y_3 &= 2\,y_1 - y_2 \\ y_4 &= y_1 + y_2 \\ y_5 &= -y_1 + 3y_2. \end{aligned}$$

Wir sehen, dass

$$\text{Rang der Matrix } = \dim \operatorname{lin}\{y_1, ..., y_5\} = \dim \operatorname{lin}\{y_1, y_2\} \leq 2$$

gilt.

Andererseits bilden die Linearformen im header der Tabelle (hier im Beispiel also x_2, x_3, y_2 und y_2) stets eine Basis des Dualraumes, sind also insbesondere linear unabhängig, denn es sind n Elemente (hier $n = 4$), und sie spannen den Dualraum auf: auch die links stehenden x_i (hier im Beispiel x_1 und x_4) sind Linearkombinationen dieser Elemente. Also hat man

$$\dim \operatorname{lin}\{y_1, y_2\} + \dim \operatorname{lin}\{x_2, x_3\} \geq \dim \operatorname{lin}\{y_1, y_2, x_2, x_3\} = 4,$$

und weil natürlich $\dim \operatorname{lin}\{x_2, x_3\} = 2$ ist, folgt daraus $\dim \operatorname{lin}\{y_1, y_2\} \geq 2$, also

$$\text{Rang der Matrix } = 2.$$

Dieses Beispiel lässt sich unmittelbar verallgemeinern:

Satz 2.4.3. *Für eine reelle $(m \times n)$-Matrix A sind folgende Aussagen äquivalent:*

(i) Man hat $\operatorname{Rang}(A) = k$.

(ii) Das Gaussverfahren (Austauschverfahren) bricht nach k Schritten ab.

Bemerkung 2.4.4. Ähnlich kann man mit dem Gaussverfahren die zu einer gegebenen $(n \times n)$-Matrix inverse finden, falls sie existiert:

Die Matrix ist genau dann invertierbar, wenn das Verfahren erst nach n Schritten abbricht. In diesem Falle stehen dann am Ende alle x_k oben und alle y_i links. Sortiert man nun das Schema nach der natürlichen Ordnung der Indizes, so ergibt sich genau die inverse Matrix.

(Wir schauen uns das in den Übungsaufgaben an.)

2.5 Erster Simplexalgorithmus

Wir betrachten folgendes Optimierungsproblem:

Gegeben sind Linearformen $f_1, ..., f_m$ und f auf $\mathbb{R}^n$ und Zahlen $c_1, ..., c_m \in \mathbb{R}$. Gesucht ist das Maximum von f unter den Nebenbedingungen (Restriktionen)

$$x_1 \geq 0,\ x_2 \geq 0,\ \cdots,\ x_n \geq 0$$

und

$$f_1 \leq c_1,\ f_2 \leq c_2,\ \cdots,\ f_m \leq c_m.$$

Die Linearformen besitzen Darstellungen

$$f_i(x) = a_{i1}\, x_1 + \cdots + a_{in}\, x_n, \quad 1 \leq i \leq m,$$

und

$$f(x) = b_1\, x_1 + \cdots + b_n\, x_n$$

mit $a_{ij}, b_j \in \mathbb{R}$.

Für $1 \leq i \leq m$ setzen wir nun

$$y_i(x) := -f_i(x) + c_i = -\sum_{j=1}^{n} a_{ij}\, x_j + c_i,$$

man nennt die y_i auch *Schlupfvariablen*. Der zulässige Bereich (= Menge der *zulässigen Punkte*, d.h. derjenigen Punkte, die alle Nebenbedingungen erfüllen) ist das Polyeder

$$P := \bigcap_{j=1}^{n} \{x_j \geq 0\} \cap \bigcap_{i=1}^{m} \{y_i \geq 0\}.$$

In diesem Abschnitt nehmen wir stets an, dass $0 \in P$ *ist.* Äquivalent dazu ist die Forderung, dass

$$c_i \geq 0 \quad \text{für alle } 1 \leq i \leq m \text{ gilt.}$$

Ähnlich wie im vorigen Abschnitt beschreiben wir das lineare Optimierungsproblem durch ein Ausgangsschema der Gestalt

	$-x_1$	$\cdots$	$-x_n$	
y_1	a_{11}	$\cdots$	a_{1n}	c_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	c_m
f	$-b_1$	$\cdots$	$-b_n$	0

(S_0)

Definition 2.5.1. Ein Schema der Gestalt (S_0) heisst ein zum linearen Optimierungsproblem zugehöriges *Simplex-Tableau*.

Formen im Folgenden dieses Tableau durch Zeilen- und Spaltentausch in äquivalente Tableaus um der Gestalt

	$-u_1$	$\cdots$	$-u_n$	
v_1	α_{11}	$\cdots$	α_{1n}	γ_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
v_m	α_{m1}	$\cdots$	α_{mn}	γ_m
f	β_1	$\cdots$	β_n	δ

(S)

Hier sind dann $\alpha_{ij}, \beta_j, \gamma_i, \delta \in \mathbb{R}$ geeignete Koeffizienten und ähnlich wie zuvor gilt:

(i) $u_1, ..., u_n, v_1, ..., v_m$ sind die (affin linearen Funktionen) $x_1, ..., x_n, y_1, ..., y_m$ in veränderter Reihenfolge,

(ii) $v_i = -\sum_{j=1}^{n} \alpha_{ij} u_j + \gamma_j, \;\; 1 \leq i \leq m,$

(iii) $f = -\sum_{j=1}^{n} \beta_j u_j + \delta.$

Bemerkung 2.5.2. Für ein gegebenes Tableau (S) nennt man $v_1, ..., v_m$ auch die *Basisvariablen*, $u_1, ..., u_n$ die *Nichtbasisvariablen* und f die *Zielfunktion*. Das Schema (S) selbst nennt man auch eine *Basisform*. Bei gegebener Basisform werden also die Basisvariablen und die Zielfunktion durch die Nichtbasisvariablen ausgedrückt.

Wegen (i) oben gilt für jede Basisform: Der zulässige Bereich ist

$$P = \{u_1 \geq 0\} \cap \cdots \cap \{u_n \geq 0\} \cap \{v_1 \geq 0\} \cap \cdots \cap \{v_m \geq 0\}.$$

Definition 2.5.3. Eine Basisform (S) heisst *zulässig*, falls

$$\gamma_i \geq 0 \quad \text{für alle } 1 \leq i \leq m.$$

Die Voraussetzung $0 \in P$ garantiert also, dass (S_0) eine zulässige Basisform ist.

Proposition 2.5.4. *Sei (S) eine Basisform. Dann gibt es genau eine Lösung $x_S = ((x_S)_1, ..., (x_S)_n) \in \mathbb{R}^n$ des linearen Gleichungssystems*

$$u_j(x_S) = 0, \quad 1 \leq j \leq n,$$

und diese Lösung x_S ist gegeben durch

$$(x_S)_l = \begin{cases} 0 & \text{falls } (x_S)_l \in \{u_1, ..., u_n\} \text{ und} \\ \gamma_i & \text{falls } (x_S)_l = v_i. \end{cases} \tag{I}$$

Ist (S) eine zulässige Basisform, so ist x_S ein Extremalpunkt von P. Ferner ist der Wert der Zielfunktion f in x_S gegeben durch

$$f(x_S) = \delta.$$

Bemerkung 2.5.5. Die Gleichungen $u_j(x) = 0$ sind entweder von der Form

$$x_l = 0 \quad (\text{falls } u_j = x_l)$$

oder

$$f_i = c_i \quad (\text{falls } u_j = y_i = -f_i + c_i).$$

In jedem Fall definieren die Gleichungen

$$u_j(x) = 0, \quad j = 1, ..., n,$$

ein lineares Gleichungssystem im Sinne von Satz 2.3.1 mit linear unabhängigen Gleichungen. Für die eindeutige Lösung x_S gilt:

$$x_S \in P \text{ (und damit } x_S \in P_e) \Leftrightarrow (S) \text{ zulässig.}$$

Speziell für das Ausgangsschema (S_0) gilt

$$u_j = x_j, \quad \text{also } x_S = 0 \text{ und } f(x_S) = 0.$$

Man kann nun verschiedene Fälle unterscheiden.

Proposition 2.5.6. *Sei (S) ein zulässiges Schema.*

(i) Ist

$$\beta_j \geq 0 \quad \text{für alle } 1 \leq j \leq n,$$

so gilt

$$f(x_S) = \delta = \max f(P),$$

d.h. x_S ist Lösung des linearen Optimierungsproblems.

(ii) Gibt es ein $1 \leq l \leq n$, sodass $\beta_l < 0$ und

$$\alpha_{il} \leq 0 \quad \text{für alle } 1 \leq i \leq m,$$

so ist f auf P nicht nach oben beschränkt, d.h. $\sup f(P) = +\infty$.

(iii) Gibt es ein $1 \leq l \leq n$*, sodass* $\beta_l < 0$ *und*

$$\text{existiert ein } 1 \leq i \leq m \text{ mit } \alpha_{il} > 0,$$

so gilt folgendes: Bildet man für alle solche $\alpha_{il} > 0$ *die Quotienten* $\frac{\gamma_i}{\alpha_{il}}$ *und wählt man den Index* k *so, dass der Quotient minimal wird, also*

$$\frac{\gamma_k}{\alpha_{kl}} = \min\left\{\frac{\gamma_i}{\alpha_{il}} : \ i \text{ so, dass } \alpha_{il} > 0\right\}, \tag{2.9}$$

dann liefert ein Austauschschritt mit Pivotelement α_{kl} *ein neues zulässiges Schema (S'). Für den nach Proposition 2.5.4 dadurch bestimmten Extremalpunkt* $x_{S'}$ *gilt*

$$f(x_{S'}) \geq f(x_S).$$

Die Quotienten $\frac{\gamma_i}{\alpha_{il}}$ nennt man auch *charakteristische Quotienten.*

Durch Iteration erhält man folgenden *Simplexalgorithmus* für den Spezialfall, dass $x_1, ..., x_n \geq 0$ und $0 \in P$ (wie vorausgesetzt):

1. Stelle zulässiges Ausgangstableau auf
2. Falls $\beta_j \geq 0$ für alle $1 \leq j \leq n$ (Maximum erreicht), dann setze Nichtbasisvariablen $u_j = 0$. Die Werte der x_j liefern eine Maximalstelle, und δ ist der Wert von f an dieser Maximalstelle. STOPP.
3. Falls für ein $\beta_l < 0$ gilt, dass $\alpha_{il} \leq 0$ für alle $1 \leq i \leq m$, so ist das Problem unbeschränkt, $\sup f(P) = +\infty$. STOPP.
4. Wähle $\beta_l < 0$ und wähle Zeile k, für die $\alpha_{kl} > 0$ ist und der charakteristische Quotient $\frac{\gamma_k}{\alpha_{kl}}$ minimal ist wie in (2.9).
5. Führe Austauschschritt aus mit Pivotelement α_{kl}.
6. Gehe zu 2.

Bevor wir zu den Beweisen kommen, schauen wir uns das im Kontext des ersten Beispiels an.

Beispiel 2.5.7. Wir erinnern uns an Beispiel 2.1.1 (i), in dem es um die Frage ging, wieviel x_1 ha Kartoffeln und wieviel x_2 ha Getreide angebaut werden müssten, um die Zielfunktion

$$f(x_1, x_2) = 400\, x_1 + 1200\, x_2$$

unter den Restriktionen (Nebenbedingungen)

$$\begin{aligned} x_1, x_2 &\geq 0 \quad \text{(Fläche immer nichtnegativ)} \\ 100\, x_1 + 200\, x_2 &\leq 11000 \quad \text{(wegen Anbaukosten)} \\ x_1 + 4\, x_2 &\leq 160 \quad \text{(wegen Arbeitstagen)} \\ x_1 + x_2 &\leq 100 \quad \text{(wegen max. verfügbarer Fläche)} \end{aligned}$$

zu maximieren.

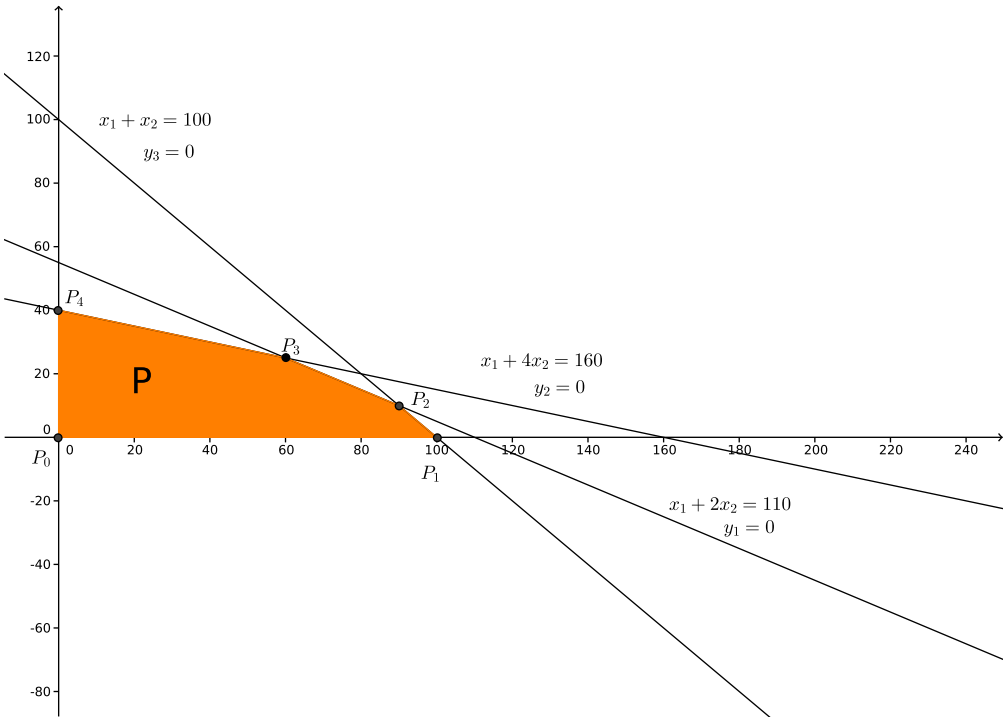


Abbildung 2.6: Simplexalgorithmus: Sukzessives Testen der Extremalpunkte.

Wir starten mit dem Ausgangsschema (der Ausgangs-Basisform)

	$-x_1$	$-x_2$	
y_1	100	200	11000
y_2	1	4	160
y_3	1	1	100
f	-400	-1200	0

Hier sind also $n = 2$ und $m = 3$. Man hat $u_1(x) = x_1$ und $u_2(x) = x_2$, nach Proposition 2.5.4 ist $x_S = (0, 0) = P_0$ die eindeutige Lösung von $u_j(x_S) = 0$, $j = 1, 2$. Weil $\gamma_1 = 11000$, $\gamma_2 = 160$ und $\gamma_3 = 100$ alle nichtnegativ sind, ist das Schema zulässig, und daher P_0 ein Extremalpunkt von P.

Nun hat man z.B. $\beta_2 = -b_2 = -1200 < 0$, also $\beta_l < 0$ für $l = 2$, somit ist nach Proposition 2.5.6 der Punkt P_0 möglicherweise noch keine Maximalstelle von $f|_P$. Da $\alpha_{22} = 4 > 0$, ist man im Fall wie in Proposition 2.5.6 (iii) beschrieben.

Der charakteristische Quotient $\frac{\gamma_2}{\alpha_{22}} = \frac{160}{4} = 40$ ist kleiner als die anderen beiden, nämlich $\frac{\gamma_1}{\alpha_{12}} = \frac{11000}{200} = 55$ und $\frac{\gamma_3}{\alpha_{32}} = \frac{100}{1} = 100$. Daher machen wir einen Austauschschritt mit Pivotelement $\alpha_{22} = 4$.

Wir tauschen also die Bezeichnungen x_2 und y_2 aus, folgen den Pivotregeln und erhalten folgendes neues Schema:

	$-x_1$	$-y_2$	
y_1	50	-50	3000
x_2	$\frac{1}{4}$	$\frac{1}{4}$	40
y_3	$\frac{3}{4}$	$-\frac{1}{4}$	60
f	-100	300	48000

Hier sind $u_1(x) = x_1$ und $u_2(x) = y_2$, also ist die Lösung von $u_j(x_S) = 0$, $j = 1, 2$, diesmal $x_S = (0, 40) = P_4$, denn $x_2 = -\frac{1}{4}\, x_1 - \frac{1}{4}\, y_2 + 40 = 40$. Wiederum ist das Schema zulässig, also P_4 extremal. Wir 'testen', ob P_4 Maximalstelle ist.

Da $\beta_1 = -100 < 0$, ist auch P_4 möglicherweise keine Maximalstelle. Es gibt für $l = 1$ Elemente α_{kl}, die grösser Null sind (sogar alle in der ersten Spalte), und die characteristischen Quotienten für $l = 1$ sind $\frac{\gamma_1}{\alpha_{11}} = \frac{3000}{50} = 60$, $\frac{\gamma_2}{\alpha_{21}} = \frac{40}{1/4} = 160$ und $\frac{\gamma_3}{\alpha_{31}} = \frac{60}{3/4} = 80$, davon ist der erste minimal. Wir machen wieder einen Austauschschritt, diesmal mit Pivot $\alpha_{11} = 50$.

Tausch der Bezeichnungen x_1 und y_1 liefert nach Pivotregeln das neue Schema

	$-y_1$	$-y_2$	
x_1	#	#	60
x_2	#	#	25
y_3	#	#	15
f	2	200	54000

Jetzt sind $u_j(x) = y_j$, $j = 1, 2$, und die eindeutige Lösung von $u_j(x_S) = 0$ kann man ablesen, nämlich $x_S = (60, 25) = P_3$. Da das Schema zulässig ist, ist P_3 extremal. Wir haben hier nun zuerst die letzte Zeile berechnet, und sehen schon, dass $\beta_1 = 2$ und $\beta_2 = 200$ beide nichtnegativ sind. Nach Proposition 2.5.6 (i) ist dann P_3 garantiert eine Maximalstelle für $f|_P$, und $f(P_3) = f(x_S) = \delta = 54000$. Die Stellen α_{kl} im neuen Schema müssen nun gar nicht mehr berechnet werden.

Der Landwirt kann also seinen Gewinn maximieren, wenn er auf $x_1 = 60$ ha Kartoffeln anbaut und auf $x_2 = 25$ ha Getreide. Er erreicht dann den Gewinn 54000.

Übrigens hätten wir im ersten Schritt auch die erste Spalte favorisieren können ($l = 1$), das hätte auf den Austausch $x_1 \leftrightarrow y_3$ geführt, was geometrisch den Schritt $P_0 \to P_1$ bedeutet hätte. Es hätten sich dann noch zwei weitere Schritte angeschlossen, nämlich $x_2 \leftrightarrow y_1$, also $P_1 \to P_2$, und $y_3 \leftrightarrow y_2$, also $P_2 \to P_3$. (Aufgabe für die Willigen: Nachchecken.)

Man läuft also, beginnend mit dem Nullpunkt, die Extremalpunkte von P ab und testet, ob sie sicher Maximalstellen sind. Kann man das garantieren für den Extremalpunkt, den man gerade betrachtet (mittels Proposition 2.5.6 (i)), so hält der Algorithmus an.

Wir beweisen zunächst Proposition 2.5.4.

Beweis. Die x_1,..., x_n sind offensichtlich n linear unabhängige Linearformen. Für jedes x_l gilt entweder

$$x_l = u_j = u_j - \underbrace{u_j(0)}_{=x_l(0)=0} \qquad \text{für ein } j \in \{1, ..., n\}$$

('x_l steht oben', d.h. x_l ist Nichtbasisvariable), oder es gilt

$$x_l = v_i = -\sum_{j=1}^{n} \alpha_{ij} u_j + \gamma_i \qquad \text{für ein } i \in \{1, ..., m\}$$

('x_l steht links', d.h. x_l ist Basisvariable), und setzt man $x = 0$ ein, so folgt

$$\gamma_i = \sum_{j=1}^{n} \alpha_{ij} u_j(0)$$

und daher

$$x_l = -\sum_{j=1}^{n} \alpha_{ij}(u_j - u_j(0)).$$

In jedem Fall ist x_l, $1 \leq l \leq n$, eine Linearkombination der Linearformen

$$u_1 - u_1(0), ..., u_n - u_n(0),$$

und das impliziert, dass die $u_1 - u_1(0)$, ..., $u_n - u_n(0)$ linear unabhängig sein müssen, denn sie spannen ja einen Raum der Dimension $\geq n$ auf.

Dann folgt aber, dass genau ein $x_S \in \mathbb{R}^n$ existiert mit

$$(u_j - u_j(0))(x_S) = -u_j(0), \quad 1 \leq j \leq n, \tag{2.10}$$

also genau ein $x_S \in \mathbb{R}^n$ mit

$$u_j(x_S) = 0, \quad 1 \leq j \leq n.$$

Für die Koordinaten von x_S gilt: Falls $x_l = u_j$ (also x_l eine Nichtbasisvariable ist), dann hat man

$$(x_S)_l = u_j(x_S) = 0.$$

Falls $x_l = v_i$ (also x_l eine Basisvariable ist), dann ist

$$(x_S)_l = v_i(x_S) = -\sum_{j=1}^{n} \alpha_{ij} u_j(x_S) + \gamma_i = \gamma_i.$$

Ist (S) zulässig, so folgt

$$u_j(x_S) = 0, \quad 1 \leq j \leq n \quad \text{nach Konstruktion}$$

und

$$v_i(x_S) = \gamma_i \geq 0, \quad 1 \leq i \leq m.$$

Somit ist $x_S \in P$, und weil x_S das lineare Gleichungssystem (2.10) eindeutig löst, ist $x_S \in P_e$ nach Satz 2.3.1. Letztlich gilt

$$f(x_S) = -\sum_{j=1}^{n} \beta_j \underbrace{u_j(x_S)}_{=0} + \delta = \delta.$$

□

Bemerkung 2.5.8. Proposition 2.5.4 sagt also, dass in jedem Simplextableau (S) die Nichtbasisvariablen ein System von n linear unabhängigen Linearformen $u_j - u_j(0)$, $1 \leq j \leq n$, definieren, die ein lineares Gleichungssystem $u_j(x) - u_j(0) = u_j(0)$, $1 \leq j \leq n$, wie in Satz 2.3.1 (i) bilden.

V07

Wir beweisen nun Proposition 2.5.6, aus welcher wir den Simplexalgorithmus abgeleitet hatten.

Beweis. Aussage (i) folgt leicht: Ist $x \in P$, so gilt $u_1(x) \geq 0$,..., $u_n(x) \geq 0$, und damit

$$f(x) = -\sum_{j=1}^{n} \underbrace{\beta_j}_{\geq 0} \underbrace{u_j(x)}_{\geq 0} + \delta \leq \delta = f(x_S).$$

Wir zeigen (ii). Da $u_1 - u_1(0)$, ..., $u_n - u_n(0)$ linear unabhängig sind, gibt es zu jedem $N \in \mathbb{N}$ genau eine Lösung $x_N \in \mathbb{R}^n$ von

$$u_j(x_N) = 0 \quad \text{für } j \in \{1, ..., n\} \setminus \{l\} \text{ und} \quad u_l(x_N) = N.$$

Wegen

$$v_i(x_N) = -\sum_{j=1}^{N} \alpha_{ij} u_j(x_N) + \gamma_i = -\underbrace{\alpha_{il}}_{\leq 0} N + \underbrace{\gamma_i}_{\geq 0} \geq 0$$

folgt $x_N \in P$ für alle N und

$$f(x_N) = -\sum_{j=1}^{n} \beta_j u_j(x_N) + \delta = -\underbrace{\beta_l}_{<0} N + \delta.$$

Wegen $\lim_{N\to\infty} f(x_N) = +\infty$ muss $\sup f(P) = +\infty$ gelten.

Wir beweisen nun (iii). Wählt man $\alpha_{kl} > 0$ als Pivot, so erhält man das neue Schema

	$-u_1 \quad \cdots$	$-v_k$	$\ldots \quad -u_n$	
v_1				
$\vdots$				$\vdots$
u_l				$\frac{\gamma_k}{\alpha_{kl}}$
$\vdots$				$\vdots$
v_i				$\gamma_i - \frac{\gamma_k \alpha_{il}}{\alpha_{kl}}$
$\vdots$				$\vdots$
v_m				
f				$\delta - \frac{\gamma_k \beta_l}{\alpha_{kl}}$

(S')

Nun soll das Schema (S') wieder zulässig sein (vorerst ein Wunsch). Da das ursprüngliche Schema zulässig war, ist $\gamma_k \geq 0$ und somit $\frac{\gamma_k}{\alpha_{kl}} \geq 0$. Für alle $i \neq k$ kann man wie folgt sehen, dass

$$\gamma_i' := \gamma_i - \frac{\gamma_k \alpha_{il}}{\alpha_{kl}} \geq 0$$

ist: Für Zeilen mit $\alpha_{il} \leq 0$ ist dies stets erfüllt, denn

$$-\frac{\gamma_k \alpha_{il}}{\alpha_{kl}} \geq 0 \quad \text{impliziert, dass} \quad \gamma_i' \geq \gamma_i \geq 0.$$

Für Zeilen mit $\alpha_{il} > 0$ muss gelten, dass

$$\gamma_i - \frac{\gamma_k \alpha_{il}}{\alpha_{kl}} \geq 0, \quad \text{und das ist äquivalent zu} \quad \frac{\gamma_k}{\alpha_{kl}} \leq \frac{\gamma_i}{\alpha_{il}},$$

d.h. die Pivotzeile (also die k-te Zeile) ist so zu wählen, dass für sie unter allen Zeilen mit $\alpha_{il} > 0$ der charakteristische Quotient $\frac{\gamma_i}{\alpha_{il}}$ minimal ist. Geht man so vor, dann sind alle $\gamma_i' \geq 0$, also ist dann (S') tatsächlich zulässig. Ausserdem gilt

$$f(x_{S'}) = \delta' := \delta - \underbrace{\frac{\gamma_k \beta_l}{\alpha_{kl}}}_{\leq 0} \geq \delta = f(x_S).$$

□

Bemerkung 2.5.9. Bei der Wahl des Pivotelementes in Schritt 4. des Simplexalgorithmus und dem anschliessenden Austauschschritt sind nun zwei Fälle (a) und (b) zu unterscheiden:

(a) In allen Zeilen mit $\alpha_{il} > 0$ gilt $\gamma_i > 0$. Dann gilt

$$f(x_{S'}) = \delta - \underbrace{\frac{\beta_l \gamma_k}{\alpha_{kl}}}_{<0} > \delta = f(x_S),$$

d.h. das neue Schema (S') beschreibt einen neuen Extremalpukt $x_{S'} \neq x_S$. Der Wert der Zielfunktion nimmt dabei echt zu. Da die Gleichungssysteme, die x_S und $x_{S'}$ beschreiben, sich nur in einer Gleichung unterscheiden (denn $n-1$ der Nichtbasisvariablen sind gleich), sind x_S und $x_{S'}$ benachbart. Also entspricht in diesem Falle ein Austauschschritt geometrisch einem Schritt zu einem benachbarten Extremalpunkt, für welchen dann der Wert der Zielfunktion echt grösser ist.

(b) Unter allen Zeilen mit $\alpha_{il} > 0$ gibt es eine mit $\gamma_i = 0$. Eine solche Zeile mit $\gamma_i = 0$ ist dann als Pivotzeile zu wählen, weil dann der charakteristische Quotient

$$\frac{\gamma_i}{\alpha_{il}} = 0$$

ist und somit minimal sein muss. Es gilt dann also $\alpha_{kl} > 0$ und $\gamma_k = 0$. Das bedeutet aber, dass die letzte Spalte beim Austausch unverändert bleibt, d.h.

$$\gamma_i' = \gamma_i, \quad 1 \leq i \leq m, \text{ und} \quad f(x_{S'}) = \delta' = \delta = f(x_S).$$

Es gilt sogar $x_{S'} = x_S$, denn weil der Austauschschritt dann der Gestalt

	$-u_l$	
v_k		0

$\rightarrow$

	$-v_k$	
u_l		0

ist, führen beide Schemata bei Nullsetzen der Nichtbasisvariablen $u_1, ..., u_n$ (im ursprünglichen Schema) bzw. $u_1, ..., u_{k-1}, v_k, u_{k+1}, ..., u_n$ (im neuen Schema) auf die Gleichungen

$$u_j(x_S) = u_j(x_{S'}) = 0, \quad 1 \leq j \leq n \quad \text{und} \quad v_k(x_S) = v_k(x_{S'}) = 0.$$

Dieses lineare Gleichungssystem ist eindeutig lösbar (denn man kann n linear unabhängige Gleichungen auswählen) mit Lösung $x_{S'} = x_S$.

Definition 2.5.10. Ein Austausch mit Pivotelement α_{kl}, für welches $\gamma_k = 0$ ist, heisst *stationärer Austausch.*

Bemerkung 2.5.11.

(i) Ein stationärer Austausch beschreibt also den Übergang zwischen zwei Gleichungssystemen, die im Sinne von Satz 2.3.1 denselben Extremalpunkt beschreiben.

(ii) Es können dabei zwei Situationen auftreten:

(b.1) Der Punkt x_S besitzt einen benachbarten Extremalpunkt $\widetilde{x}$ mit

$$f(x_S) < f(\widetilde{x}).$$

Dann gibt es nach Satz 2.3.7 eine endliche Folge stationärer Austauschschritte, nach welchen wieder der Fall (a) eintritt. (Denn es gibt nur endlich viele Möglichkeiten für Gleichungssysteme mit eindeutiger Lösung x_S, und sobald man einen Austausch macht, der zu einer Gleichung führt, die x_S selbst nicht mehr erfüllt, ist dann nach Satz 2.3.7 die eindeutige Lösung des neuen Systems ein zu x_S benachbarter Extremalpunkt.)

(b.2) Der Punkt x_S besitzt keinen benachbarten Extremalpunkt $\widetilde{x}$ mit

$$f(x_S) < f(\widetilde{x}).$$

In diesem Falle ist x_S bereits ein Extremalpunkt, an welchem die Zielfunktion f maximal wird, oder f ist unbeschränkt auf P.

(iii) Eine Folge stationärer Austauschschritte kann einen Zyklus (loop) bilden. (Dann würde sich der Algorithmus aufhängen.) Um das zu vermeiden, gibt es geeignete Auswahlregeln, z.B. *Bland's Antizyklenregel*:

Ordne die Variablen in einer Reihenfolge, z.B.

$$z_1 = x_1,\ \dots\ ,\ z_n = x_n,\ z_{n+1} = y_1,\ \dots\ ,\ z_{n+m} = y_m.$$

Wenn bei einem Austausch mehrere Spalten oder Zeilen wählbar sind, so wähle stets die Variable z_i mit dem kleinsten Index i. (Der Index der basisverlassenden Variablen soll möglichst klein gewählt sein.)

Wir ergänzen Schritt 4. im Simplexalgorithmus durch folgende *Auswahlregel*:

(i) Versuche stationäre Austauschschritte zu vermeiden: Gibt es mehrere $\beta_l < 0$ (mehrere mögliche Pivotspalten), so wähle wenn möglich eine solche, für welche der Fall (a) eintritt. Falls nicht möglich, wende Blands Antizyklenregel an.

(ii) *(Regel von Dantzig)* Sind mehrere Pivotspalten für einen nicht stationären Austausch möglich, so wähle die Spalte mit dem betragsmässig grössten β_l (Faustregel, Grundidee ist maximale Kostenreduktion; aber nicht unbedingt optimal).

Mit dieser Auswahlregel ist im Simplexalgorithmus jeder Schritt bestimmt (anders als zuvor, und eigentlich kann man deshalb erst jetzt von einem Algorithmus sprechen). Zusätzlich ist ein Aufhängen ausgeschlossen:

Korollar 2.5.12. *Mit der zusätzlichen Auswahlregel terminiert der Simplex-Algorithmus nach endlich vielen Austauschschritten.*

Beweis. Der Algorithmus verharrt höchstens endlich viele (stationäre) Schritte in einem Extremalpunkt, in solchen Schritten (Fall (b)) bleibt der Wert der Zielfunktion gleich. In jedem nicht stationären Schritt (Fall (a)) erfolgt ein Wechsel zu einem benachbarten Extremalpunkt, währenddessen der Wert der Zielfunktion echt grösser wird. Jeder Extremalpunkt wird also höchstens einmal durchlaufen. Da es nur endlich viele Extremalpunkte gibt, tritt nach endlich vielen Schritten eine der Abbruchbedingungen 2. (Maximum von $f|_P$ erreicht) oder 3. (Problem stellt sich als unbeschränkt heraus) ein. □

Bemerkung 2.5.13. Selbst wenn der zulässige Bereich ein unbeschränktes Polyeder P ist, gilt: Ist die Zielfunktion f auf P beschränkt, so nimmt $f|_P$ sein Maximum in einem Extremalpunkt von P an.

Bemerkung 2.5.14. Man kann die Effekte der Auswahlregel auch noch stärker formalisiert beschreiben, allerdings ist das eher notationsintensiv. Wir beschränken uns daher auf diese eher knappe und konzeptionelle Diskussion.

2.6 Simplexverfahren in allgemeineren Situationen

Bisher hatten wir stets unter der Bedingung gearbeitet, dass alle Variablen nichtnegativ sein sollen, und dass 0 ein Element des zulässigen Bereiches ist. Wir schauen uns nun an, wie man diese Voraussetzungen fallen lassen kann.

2.6.1 Variablen nicht nach unten beschränkt

Wir möchten $f(x) = b_1\, x_1 + ... + b_n\, x_n$ maximieren unter den Nebenbedingungen

$$f_i(x) = \sum_{j=1}^{n} a_{ij} x_j \leq c_i, \quad 1 \leq i \leq m,$$

wobei wie zuvor $c_i \geq 0$ gelten soll für alle i. (Also soll 0 wieder ein Punkt des zulässigen Bereiches sein.) Wir verlangen aber nun *nicht* mehr, dass $x_j \geq 0$ gelten soll für alle $1 \leq j \leq n$.

Für jedes j, für das $x_j \geq 0$ keine vorgegebene Restriktion mehr ist, kann man nun wie folgt vorgehen: Man ersetzt x_j einfach durch die Differenz zweier neuer Variablen x'_j und x''_j, also

$$x_j = x'_j - x''_j,$$

und stellt die neuen Nebenbedingungen

$$x'_j \geq 0 \quad \text{und} \quad x''_j \geq 0.$$

(Geometrisch erhöht man also künstlich die Dimension.)

Beispiel 2.6.1. Nehmen wir an, $f(x) = x_1 - x_2$, $x = (x_1, x_2) \in \mathbb{R}^2$, soll maximiert werden unter den Nebenbedingungen

$$\begin{aligned} 3\, x_1 - x_2 &\leq 6 \\ x_1 - 2\, x_2 &\leq 4 \\ x_2 &\leq 1 \\ x_1 &\geq 0. \end{aligned}$$

Wir setzen $x_2 := x'_2 - x''_2$. Das führt auf das modifizierte Problem,

$$f(x) = x_1 - x'_2 + x''_2, \quad x = (x_1, x'_2, x''_2) \in \mathbb{R}^3,$$

zu maximieren unter

$$\begin{aligned} 3\, x_1 - x'_2 + x''_2 &\leq 6 \\ x_1 - 2\, x'_2 + 2\, x''_2 &\leq 4 \\ x'_2 - x''_2 &\leq 1 \end{aligned}$$

sowie $x_1 \geq 0$, $x_2' \geq 0$ und $x_2'' \geq 0$. Das führt auf das Ausgangsschema

	$-x_1$	$-x_2'$	$-x_2''$	
y_1	3	-1	1	6
y_2	1	-2	2	4
y_3	0	1	-1	1
f	-1	1	-1	0

Man kann nun wie im letzten Abschnitt mit dem Simplexverfahren lösen, und am Ende erhält man x_2 aus $x_2 = x_2' - x_2''$:

Der erste Austauschschritt mit Pivot 2 ergibt das Schema

	$-x_1$	$-x_2'$	$-y_2$	
y_1	$\frac{5}{2}$	0	$-\frac{1}{2}$	4
x_2''	$\frac{1}{2}$	-1	$\frac{1}{2}$	2
y_3	$\frac{1}{2}$	0	$\frac{1}{2}$	3
f	$-\frac{1}{2}$	0	$\frac{1}{2}$	2

Der zweite, mit Pivot $\frac{5}{2}$, ergibt

	$-y_1$	$-x_2'$	$-y_2$	
x_1	$\frac{2}{5}$	0	$-\frac{1}{5}$	$\frac{8}{5}$
x_2''	$-\frac{1}{5}$	-1	$\frac{6}{10}$	$\frac{6}{5}$
y_3	$-\frac{1}{5}$	0	$\frac{6}{10}$	$\frac{11}{5}$
f	$\frac{1}{5}$	0	$\frac{4}{10}$	$\frac{14}{5}$

Nun ist man aber schon in der Situation von Proposition 2.5.6 (i), denn die Einträge β_j in der letzten Zeile sind alle nichtnegativ. Die eindeutige Lösung des Gleichungssytems, welches aus diesem Schema durch Nullsetzen der Nichtbasisvariablen entsteht, ist

$$(x_1,\ x_2',\ x_2'') = (\frac{8}{5},\ 0,\ \frac{6}{5}),$$

und f hat in diesem Punkt den Wert $\frac{14}{5}$. Rückübersetzen ergibt

$$x_2 = x_2' - x_2'' = -\frac{6}{5},$$

also nimmt die Zielfunktion $f|_P$ ihr Maximum in

$$(x_1,\ x_2) = (\frac{8}{5},\ -\frac{6}{5})$$

an,

$$\max f(P) = \frac{14}{5} = f(\frac{8}{5},\ -\frac{6}{5}).$$

2.6.2 Ausgangsschema nicht zulässig

Nehmen an, dass es ein $i \in \{1, ..., m\}$ gibt mit $c_i < 0$. Dann ist also $0 \notin P$. Das Simplexverfahren in der besprochenen Form benötigt aber ein zulässiges Ausgangsschema. Wir ermitteln daher ein solches.

Nehmen wir an, die gegebenen Restriktionen lauten

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

und $x_1, ..., x_n \geq 0$. Der zulässige Bereich ist dann das Polyeder

$$P = \left\{ x = (x_1, ..., x_n) \in \mathbb{R}^n_+ : \sum_{j=1}^{n} a_{ij} x_j \leq c_i, \quad i = 1, ..., m \right\}.$$

Wir betrachten nun ein Hilfsproblem mit $n+1$ Variablen $x_1, ..., x_n, t$ und Zielfunktion

$$\widetilde{f}(x, t) = -t$$

unter den Nebenbedingungen

$$\sum_{j=1}^{n} a_{ij} x_j - t \leq c_i, \quad 1 \leq i \leq m,$$

und $x_1, ..., x_n, t \geq 0$. Der zugehörige zulässige Bereich ist das Polyeder

$$\widetilde{P} = \left\{ (x, t) \in \mathbb{R}^{n+1}_+ : \sum_{j=1}^{n} a_{ij} x_j - t \leq c_i, \quad i = 1, ..., m \right\}.$$

Dann gilt:

(i) Für $c := -\min\{c_i : \ i = 1, ..., m\}$ ist $(0, ..., 0, c) \in \widetilde{P}$, d.h. $\widetilde{P} \neq \emptyset$ und, da $\widetilde{f}$ nach oben beschränkt ist durch 0, ist das Hilfsproblem lösbar.

(ii) Man hat $x \in P$ genau dann, wenn $(x, 0) \in \widetilde{P}$.

(iii) Man hat $P \neq \emptyset$ genau dann, wenn $\max \widetilde{f}(\widetilde{P}) = 0$, denn:

Wenn $x_0 \in P$, dann $(x_0, 0) \in \widetilde{P}$, also $\max \widetilde{f}(\widetilde{P}) \geq \widetilde{f}(x_0, 0) = 0$ und somit (wegen $\widetilde{f} \leq 0$) $\max \widetilde{f}(\widetilde{P}) = 0$.

Umgekehrt gilt folgt aus $\max \widetilde{f}(\widetilde{P}) = 0$, dass es $(x_0, t_0) \in \widetilde{P}$ geben muss mit $\widetilde{f}(x_0, t_0) = 0$; da aber $\widetilde{f}(x_0, t_0) = -t_0$ ist, muss dann $t_0 = 0$ sein, also $(x_0, t_0) = (x_0, 0) \in \widetilde{P}$, also wegen (ii) $x_0 \in P$.

In Tableauschreibweise liefert das den folgenden Algorithmus: Das Ausgangsschema ist von der Form

	$-x_1$	$\cdots$	$-x_n$	
y_1	a_{11}	$\cdots$	a_{1n}	c_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	c_m
f	$-b_1$	$\cdots$	$-b_n$	0

und für mindestens ein i ist $c_i < 0$. Für das Hilfsproblem haben wir das Schema

	$-x_1$	$\cdots$	$-x_n$	$-t$	
y_1	a_{11}	$\cdots$	a_{1n}	-1	c_1
$\vdots$	$\vdots$			$\vdots$	$\vdots$
y_k				-1	
$\vdots$	$\vdots$			$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	-1	c_m
$\widetilde{f}$	0	$\cdots$	0	1	0

und das ist ebenfalls nicht zulässig. Wählen nun $k \in \{1, ..., m\}$ so, dass

$$c_k = \min\{c_i : \ i = 1, ..., m\}$$

und tauschen y_k und t, das ergibt ein Schema der Gestalt

	$-x_1$	$\cdots$	$-x_n$	$-y_k$	
y_1	$\#$	$\cdots$	$\#$	-1	γ_1
$\vdots$	$\vdots$			$\vdots$	$\vdots$
t				-1	
$\vdots$	$\vdots$			$\vdots$	$\vdots$
y_m	$\#$	$\cdots$	$\#$	-1	γ_m
$\widetilde{f}$	$\#$	$\cdots$	$\#$	1	$\#$

mit $\gamma_k = -c_k > 0$, und für $j \neq k$ gilt

$$\gamma_j = c_j + c_k \cdot (-1) = c_j - c_k \geq 0.$$

Also ist dieses Schema zulässig.

Wir suchen nun nach dem Maximum von $\widetilde{f}(x, t) = -t$ auf $\widetilde{P}$. Falls $\max \widetilde{f}(\widetilde{P}) < 0$ folgt $P = \emptyset$, und man kann also auch kein Maximum für das ursprüngliche Problem finden. Falls $\max \widetilde{f}(\widetilde{P}) = 0$, so folgt $P \neq \emptyset$ (wie oben diskutiert). In diesem Fall wird das Maximum von $\widetilde{f}|_{\widetilde{P}}$ in einem Punkt $(x_0, 0) \in \mathbb{R}^{n+1}_+$ angenommen, und dann muss auch $x_0 \in P$ gelten.

Nun können wir erreichen, dass t eine Nichbasisvariable ist, denn: Ist t Basisvariable, so muss in

	$-u_1$	$\cdots$	$-u_{n+1}$	
	$\#$	$\cdots$	$\#$	γ_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
t	α_{k1}		α_{km}	γ_k
$\vdots$	$\vdots$		$\vdots$	$\vdots$
	$\#$	$\cdots$	$\#$	γ_m
$\widetilde{f}$	$\#$	$\cdots$	$\#$	0

$\gamma_k = 0$ gelten, weil die Maximalstelle (also die Lösung des Gleichungssystems) $(x, t) = (x_0, 0)$ ist. Nun muss es auch mindestens ein $j \in \{1, ..., n + 1\}$ geben mit $\alpha_{kj} \neq 0$, denn

sonst wäre t konstant null als Funktion der $u_1, ..., u_{n+1}$, und das wäre ein Widerspruch dazu, dass t auf $\widetilde{P}$ unbeschränkt ist. Wir wählen nun ein solches $\alpha_{kj} \neq 0$ als Pivotelement. Ein Austauschschritt mit diesem Pivot ergibt ein zulässiges Schema mit t als Nichtbasisvariable.

Wegen $\gamma_k = 0$ ist dieser Austauschschritt ist stationär, er ändert also nicht den betrachteten Extremalpunkt von $\widetilde{P}$. Deshalb kann man nun $t = 0$ setzen, d.h. die t-Spalte streichen. Was man erhält, ist ein zulässiges Schema, das die Restriktionen von P beschreibt.

Was wir jetzt noch machen müssen, ist, die letzte Zeile durch die korrekte Zeile für die ursprüngliche Zielfunktion f zu ersetzen, aber dargestellt als Funktion der aktuellen Nichtbasisvariablen. Dazu kann man einfach jedes x_k, dass eine Basisvariable ist, durch die Nichtbasisvariablen ausdrücken,

$$x_k = \sum_{j=1}^{n} \alpha_{ij} u_j + \gamma_i, \qquad \text{falls } x_k \text{ in der } i\text{-ten Zeile steht.}$$

V08 Nun haben wir ein *zulässiges* Ausgangsschema.

Bemerkung 2.6.2. Man löst also zuerst das Hilfsproblem und nutzt die Lösung, um ein zulässiges Ausgangsschema für das ursprüngliche Problem zu konstruieren. Danach löst man das ursprüngliche Problem wie gewohnt.

Beispiel 2.6.3. Wir wollen $f(x) = 2\,x_1 - x_2 + 2\,x_3$ maximieren unter den Restriktionen $x_1, x_2, x_3 \geq 0$ und

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 6 \\ -x_1 + x_2 &\leq -1 \\ -x_2 + x_3 &\leq -1. \end{aligned}$$

Das liefert ein nicht zulässiges Ausgangsschema. Wir maximieren deshalb $\widetilde{f}(x,t) = -t$ unter den Restriktionen $x_1, x_2, x_3, t \geq 0$ und

$$\begin{aligned} x_1 + x_2 + x_3 - t &\leq 6 \\ -x_1 + x_2 - t &\leq -1 \\ -x_2 + x_3 - t &\leq -1. \end{aligned}$$

Das zugehörige Schema ist

	$-x_1$	$-x_2$	$-x_3$	$-t$	
y_1	1	1	1	-1	6
y_2	-1	1	0	-1	-1
y_3	0	-1	1	-1	-1
$\widetilde{f}$	0	0	0	1	0

Der minimale Wert in der letzten Spalte ist -1, er wird in der 2. und der 3. Zeile angenommen. Wir tauschen t mit y_2, (Pivot ist -1) und erhalten

	$-x_1$	$-x_2$	$-x_3$	$-y_2$	
y_1	2	0	1	-1	7
t	1	-1	0	-1	1
y_3	1	-2	1	-1	0
$\widetilde{f}$	-1	1	0	1	-1

Dieses Schema ist nun zulässig. Die charakteristischen Quotienten für $l = 1$ sind $\frac{7}{2}$, 1 und 0 (stationärer Austausch). Mit Pivot 1 ergibt sich nun

	$-y_3$	$-x_2$	$-x_3$	$-y_2$	
y_1	-2	4	-1	1	7
t	-1	1	-1	0	1
x_1	1	-2	1	-1	0
$\widetilde{f}$	1	-1	1	0	-1

und die charakteristischen Quotienten zu $l = 2$ sind $\frac{7}{4}$ und 1, letzterer minimal. Mit Pivot 1 erhalten wir

	$-y_3$	$-t$	$-x_3$	$-y_2$	
y_1	2	-4	3	1	3
x_2	-1	1	-1	0	1
x_1	-1	2	-1	-1	2
$\widetilde{f}$	0	1	0	0	0

Dieses Schema ist optimal im dem Sinne, dass $\widetilde{f}$ ihren maximalen Wert 0 annimmt, also ist insbesondere $P \neq \emptyset$.

Wir dürfen die t-Spalte 'vergessen' und $t = 0$ setzen. Stellen nun f als Funktion von y_3, x_3 und y_2 dar: Hatten $f(x) = 2\,x_1 - x_2 + 2\,x_3$. Aus 2. und 3. Zeile folgt, dass (wegen $t = 0$)

$$\begin{aligned} x_2 &= y_3 + x_3 + 1 \\ x_1 &= y_3 + x_3 + y_2 + 2, \end{aligned}$$

also

$$f(x) = y_3 + 3\,x_3 + 2\,y_2 + 3.$$

Wir haben nun folgendes zulässiges Schema für das ursprüngliche Problem gefunden:

	$-y_3$	$-x_3$	$-y_2$	
y_1	2	3	1	3
x_2	-1	-1	0	1
x_1	-1	-1	-1	2
f	-1	-3	-2	3

Folgen nun wie früher dem Simplexalgorithmus. Für $l = 3$ hat man nur den charakteristischen Quotienten 3, wählen also Pivot 1 und erhalten

	$-y_3$	$-x_3$	$-y_1$	
y_2	2	3	1	3
x_2	-1	-1	0	1
x_1	1	2	1	5
f	3	3	2	9

Da alle β_l nichtnegativ sind, ist das Schema optimal, und Nullsetzen der Nichtbasisvariablen ergibt die Maximalstelle $(5, 1, 0)$ für $f|_P$, und der Wert von f an dieser Stelle ist 9.

2.7 Alternativsätze

Wir schauen und ein hilfreiches Werkzeug an, das man benutzen kann, um die Gültigkeit linearer Ungleichungen zu charakterisieren.

2.7.1 Notation und Wiederholung

Wir fassen Vektoren als Spaltenvektoren auf,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \in \mathbb{R}^n.$$

Wir schreiben

$$x \geq 0 \quad \text{falls} \quad x_1 \geq 0,\ x_2 \geq 0,\ ...,\ x_n \geq 0$$

(wie in der Übung), und $x \geq y$ oder $y \leq x$, falls $x - y \geq 0$. Für eine Matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \qquad \text{bezeichnet} \qquad A^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}$$

ihre transponierte, insbesondere

$$(x_1, ..., x_n)^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Man hat $(A \cdot B)^T = B^T \cdot A^T$. Mit

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

bezeichnen wir das Euklidische Skalarprodukt auf $\mathbb{R}^n$. Wir bemerken, dass $x^T \cdot y = \langle x, y \rangle$. Ist $L \subset \mathbb{R}^n$ ein Untervektorraum von $\mathbb{R}^n$, dann heisst

$$L^\perp := \{y \in \mathbb{R}^n : \langle x, y \rangle = 0 \quad \text{für alle } x \in L\}$$

das *orthogonale Komplement von L*. Man hat

$$L \oplus L^\perp = \mathbb{R}^n \quad \text{(orthogonale Summe) und} \quad (L^\perp)^\perp = L.$$

2.7.2 Formulierung der Sätze und Beweise

Die folgende Beobachtung nennt man das *Lemma von Farkas*.

Satz 2.7.1. *Sei $L \subset \mathbb{R}^n$ ein Untervektorraum. Dann gilt genau eine der beiden Alternativen:*

(i) Es gibt ein $x \in L$ mit $x \geq 0$ und $x_1 > 0$.

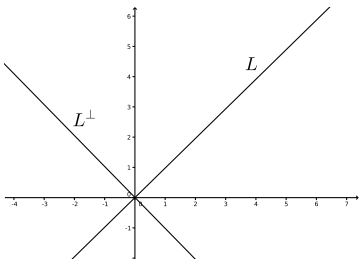


Abbildung 2.7: Fall (i)

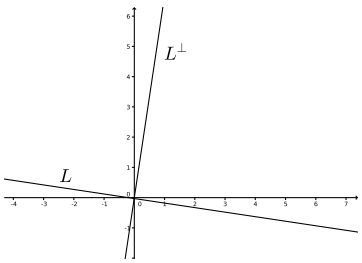


Abbildung 2.8: Fall (ii)

(ii) Es gibt ein $y \in L^\perp$ mit $y \geq 0$ und $y_1 > 0$.

Zum Beweis nutzen wir folgende Aussage.

Lemma 2.7.2. *Sei $n \geq 2$, $L \subset \mathbb{R}^n$ ein Untervektorraum und $L^\perp$ sein orthogonales Komplement. Seien*

$$\widetilde{L} := \left\{\widetilde{x} \in \mathbb{R}^{n-1} : \ (\widetilde{x}, 0) \in L\right\}$$

('Schnitt von L mit der Hyperebene $H := \{x_n = 0\}$') und

$$\hat{L^\perp} := \left\{\hat{y} \in \mathbb{R}^{n-1} : \ \text{es gibt ein } y_n \in \mathbb{R} \text{ mit } (\hat{y}, y_n) \in L^\perp\right\}$$

('Bild der Projektion von $L^\perp$ auf H'). Dann sind $\widetilde{L}$ und $\hat{L^\perp}$ zueinander orthogonal komplementäre Unterräume des $\mathbb{R}^{n-1}$,

$$(\widetilde{L})^\perp = \hat{L^\perp}.$$

Beweis. Wir zeigen zunächst, dass $\hat{L^\perp} \subset (\widetilde{L})^\perp$: Sei $\hat{y} \in \hat{L^\perp}$. Dann gibt es also ein $y_n \in \mathbb{R}$ sodass $(\hat{y}, y_n) \in L^\perp$. Für alle $\widetilde{x} = (x_1, ..., x_{n-1}) \in \widetilde{L}$ hat man dann

$$\langle \widetilde{x}, \hat{y} \rangle = \sum_{i=1}^{n-1} x_i \cdot y_i = \sum_{i=1}^{n-1} x_i \cdot y_i + 0 \cdot y_n = \big\langle \underbrace{(\widetilde{x}, 0)}_{\in L}, \underbrace{(\hat{y}, y_n)}_{\in L^\perp} \big\rangle = 0.$$

Also ist $\hat{y} \in (\widetilde{L})^\perp$. Daraus folgt, dass $\hat{L^\perp}$ und $\widetilde{L}$ orthogonal zueinander sind (d.h. dass ihre Elemente paarweise orthogonal sind).

Nun genügt es zu zeigen, dass

$$\dim \widetilde{L} + \dim \hat{L^\perp} = \dim H = n - 1.$$

(Wir nutzen also ein Dimensionsargument, um den Beweis zu vervollständigen.) Man hat

$$L \subset H \quad \text{genau dann gilt, wenn} \quad e_n := (0, ..., 0, 1)^T \in L^\perp.$$

Sei nun $P : L^\perp \to H$ die orthogonale Projektion auf H. Dann ist $y \in \ker P$ genau dann, wenn y von der Form $y = (0, ..., 0, y_n)^T$ ist, also genau dann, wenn $y \in L^\perp \cap \operatorname{lin}\{e_n\}$; insbesondere

$$\dim \ker P = \begin{cases} 1 & \text{wenn } e_n \in L^\perp \\ 0 & \text{wenn } e_n \notin L^\perp. \end{cases}$$

Mit der Dimensionsformel für lineare Abbildungen erhält man

$$\dim \hat{L^\perp} = \dim P(L^\perp) = \dim L^\perp - \dim \ker P.$$

Die Dimensionsformel für Summen von Unterräumen liefert

$$\begin{aligned}\dim \widetilde{L} &= \dim L \cap H \\ &= \dim L + \dim H - \dim(L+H) \\ &= \dim L + n - 1 - \begin{cases}(n-1) & \text{falls } L \subset H \\ n & \text{falls } L \not\subset H\end{cases} \\ &= \dim L - 1 + \begin{cases}1 & \text{falls } L \subset H \\ 0 & \text{falls } L \not\subset H\end{cases} \\ &= \dim L - 1 + \dim \ker P\end{aligned}$$

(denn, wie oben bemerkt, gilt $e_n \in L^\perp$ genau dann, wenn $L \subset H$). Summation ergibt

$$\begin{aligned}\dim \widetilde{L} + \dim \hat{L^\perp} &= \dim L - 1 + \dim \ker P + \dim L^\perp - \dim \ker P \\ &= \dim L + \dim L^\perp - 1 \\ &= n - 1.\end{aligned}$$

□

Wir beweisen Satz 2.7.1.

Beweis. Die Aussagen (i) und (ii) schliessen einander aus, denn wären beide gültig, so folgte

$$0 = x^T y = \sum_{i=1}^{n} \underbrace{x_i \cdot y_i}_{\geq 0} \geq x_1 \cdot y_1 > 0,$$

was offensichtlich Quatsch ist. Es genügt daher zu zeigen, dass einer der beiden Fälle (i) oder (ii) eintritt. Das kann man mit Induktion bezüglich n sehen:

Ist $n = 1$, so gilt $L = \mathbb{R}$, $L^\perp = \{0\}$ oder $L = \{0\}$ und $L^\perp = \mathbb{R}$. In jedem Fall ist (i) oder (ii) richtig.

Wir nehmen an, dass für $n-1$ (i) oder (ii) gilt und zeigen selbiges für n. Sei $L \subset \mathbb{R}^n$ ein Unterraum. Betrachte in $\mathbb{R}^{n-1}$ die Unterräume

$$\widetilde{L} := \left\{\widetilde{x} \in \mathbb{R}^{n-1} : \ (\widetilde{x}, 0) \in L\right\},$$

$$\hat{L} := \left\{\hat{x} \in \mathbb{R}^{n-1} : \ \text{es gibt } x_n \in \mathbb{R} \text{ mit } (\hat{x}, x_n) \in L\right\},$$

$$\widetilde{L^\perp} := \left\{\widetilde{y} \in \mathbb{R}^{n-1} : \ (\widetilde{y}, 0) \in L^\perp\right\}$$

und

$$\hat{L^\perp} = \left\{\hat{y} \in \mathbb{R}^{n-1} : \ \text{es gibt } y_n \in \mathbb{R} \text{ mit } (\hat{y}, y_n) \in L^\perp\right\}.$$

Nach Lemma 2.7.2 sind $\widetilde{L}$ und $\hat{L^\perp}$ sowie $\widetilde{L^\perp}$ und $(\hat{L^\perp})^\perp = \hat{L}$ zueinander orthogonal komplementäre Unterräume des $\mathbb{R}^{n-1}$.

Nach Induktionsvoraussetzung für das Paar $\widetilde{L}$ und $\hat{L^\perp}$ gilt: Es gibt einen Vektor der Gestalt

(1a) $(+, \oplus, ..., \oplus, 0) \in L$ oder

(1b) $(+, \oplus, ..., \oplus, *) \in L^\perp$,

hiernei bezeichnet + eine strikt positive Koordinate, $\oplus$ eine nichtnegative Koordinate und $*$ eine beliebige Koordinate.

Nach Induktionsvoraussetzung für das Paar $\widetilde{L^\perp}$ und $\hat{L}$ folgt: Es gibt einen Vektor der Gestalt

(2a) $(+, \oplus, ..., \oplus, *) \in L$ oder

(2b) $(+, \oplus, ..., \oplus, 0) \in L^\perp$.

Im Fall (1a) ist Alternative (i) erfüllt und im Fall (2b) Alternative (ii). Um den Beweis fortzusetzen dürfen wir also annehmen, dass (1b) und (2a) erfüllt sind, d.h. es gibt

$$x = (\hat{x}, x_n) \in L \text{ und } y = (\hat{y}, y_n) \in L^\perp \text{ mit } x_n, y_n \in \mathbb{R},\ \hat{x}, \hat{y} \geq 0,\ x_1, y_1 > 0.$$

Dann ist aber

$$0 = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \underbrace{\sum_{i=2}^{n-1} x_i y_i}_{\geq 0} + x_n y_n \geq x_1 y_1 + x_n y_n,$$

also

$$-x_n y_n \geq x_1 y_1 > 0.$$

Daraus folgt, dass $x_n > 0$, wonach (i) gilt, oder $y_n > 0$, wonach (ii) gilt. □

V09

Die folgende Beobachtung macht Satz 2.7.1 'praktisch leicht nutzbar'.

Bemerkung 2.7.3. Sei A eine reelle $(m \times n)$-Matrix. Dann gilt ja

$$\langle y, Ax \rangle = y^T(Ax) = (y^T A)x = (A^T y)^T x = \left\langle A^T y, x \right\rangle.$$

Daher gilt für Kern

$$\ker(A) = \{x \in \mathbb{R}^n :\ Ax = 0\}$$

und Bild

$$\text{Bild}(A^T) = \left\{A^T y :\ y \in \mathbb{R}^m\right\}$$

der zugehörigen linearen Abbildungen (welche wir hier pragmatisch ebenfalls mit A bezeichnen), dass $\ker(A)$ und $\text{Bild}(A^T)$ orthogonale Komplemente im $\mathbb{R}^n$ sind, denn für beliebiges $x \in \mathbb{R}^n$ gilt:

$$\begin{aligned} x \in \ker(A) \Leftrightarrow Ax = 0 &\Leftrightarrow \langle y, Ax \rangle = 0 \text{ für alle } y \in \mathbb{R}^m \\ &\Leftrightarrow \left\langle A^T y, x \right\rangle = 0 \text{ für alle } y \in \mathbb{R}^m \ \Leftrightarrow x \in (\text{Bild}(A^T))^\perp. \end{aligned}$$

Bemerkung 2.7.4. Bemerkung 2.7.3 besagt: A ist nicht injektiv, genau dann wenn A^T nicht surjektiv ist. Also ist A^T surjektiv die Alternative zu A nicht injektiv.

Durch geeignete Wahl von A lassen sich nun verschiedene Alternativsätze beweisen. Wir betrachten zunächst den *Satz von Minkowski-Farkas für lineare Gleichungen.*

Satz 2.7.5. *Sei A eine reelle $(m \times n)$-Matrix und $b \in \mathbb{R}^m$. Dann sind äquivalent:*

(i) Die Gleichung $Ax = b$ besitzt eine Lösung $x \geq 0$.

(ii) Für alle $u \in \mathbb{R}^m$ mit $u^T A \leq 0$ gilt $u^T b \leq 0$.

Beweis. Wir bemerken, dass die Verneinung der Aussage (ii) wie folgt lautet: Es gibt $u \in \mathbb{R}^m$ mit $u^T A \leq 0$ und $u^T b > 0$. Wir müssen also zeigen, dass entweder (i) oder nicht (ii) gilt. Nun ist es aber nicht möglich, dass gleichzeitig (i) und nicht (ii) gelten, denn ergäbe den Widerspruch

$$u^T b = u^T(Ax) = \underbrace{(u^T A)}_{\leq 0} \underbrace{x}_{\geq 0} \leq 0$$

zu $u^T b > 0$. Also genügt es nun zu zeigen, dass (i) oder nicht (ii) gilt.

Wir betrachten die reelle $(m \times (n+1))$-Matrix $M := (-b, A)$. Nach Satz 2.7.1 und Bemerkung 2.7.3 mit $L = \ker(M)$ und $L^\perp = \text{Bild}(M^T)$ gilt nun entweder

(1) es gibt $(\underbrace{x_0}_{\in\mathbb{R}}, \underbrace{x}_{\in\mathbb{R}^n})^T \in \ker(M)$ mit $x_0 > 0$ und $x \geq 0$ oder

(2) es gibt $\widetilde{u} \in \mathbb{R}^m$ mit $\underbrace{-b^T\widetilde{u}}_{=-\widetilde{u}^T b} > 0$ und $\underbrace{A^T\widetilde{u}}_{=(\widetilde{u}^T A)^T} \geq 0$, denn

$$M^T\widetilde{u} = \begin{pmatrix} -b^T\widetilde{u} \\ A^T\widetilde{u} \end{pmatrix}.$$

In Fall (1) gilt (i) mit $\widetilde{x} := \frac{x}{x_0} \geq 0$, denn

$$0 = \frac{1}{x_0} M(x_0, x)^T = \frac{1}{x_0}(-x_0 b + Ax) = -b + A\widetilde{x},$$

also $A\widetilde{x} = b$. Im Fall (2) gilt nicht (ii) mit $u := -\widetilde{u}$, denn

$$u^T b > 0 \quad \text{und} \quad u^T A \leq 0.$$

□

Der *Satz von Minkowski-Farkas* gilt auch *für lineare Ungleichungen.*

Satz 2.7.6. *Sei A eine reelle $(m \times n)$-Matrix und $b \in \mathbb{R}^m$. Dann sind äquivalent:*

(i) Die Ungleichung $Ax \geq b$ besitzt eine Lösung $x \geq 0$.

(ii) Für alle $u \in \mathbb{R}^m$ mit $u \geq 0$ und $u^T A \leq 0$ gilt $u^T b \leq 0$.

Beweis. Nun lautet nicht (ii) wie folgt: Es existiert ein $u \in \mathbb{R}^m$ mit $u \geq 0$, $u^T A \leq 0$, so dass $u^T b > 0$. Wie zuvor schliessen sich (i) und nicht (ii) aus, denn für u wie eben formuliert und x wie in (i) ergäbe sich der Widerspruch

$$\underbrace{u^T}_{\geq 0} b \leq \underbrace{u^T A}_{\leq 0} \underbrace{x}_{\geq 0} \leq 0.$$

Mit der reellen $(m \times (m+n+1))$-Matrix $M = (b, -A, E_m)$, wobei E_m die $(m \times m)$-Einheitsmatrix bezeichnet, folgt nach Satz 2.7.1 und Bemerkung 2.7.3, dass entweder gilt

(1) es gibt $(\underbrace{x_0}_{\in\mathbb{R}}, \underbrace{x}_{\in\mathbb{R}^n}, \underbrace{z}_{\in\mathbb{R}^m})^T \in \ker(M)$ mit $x_0 > 0$, $x \geq 0$ und $z \geq 0$ oder

(2) es gibt $u \in \mathbb{R}^m$ mit $b^T u = u^T b > 0$, $A^T u \leq 0$ und $u \geq 0$, also gilt nicht (ii); bemerke, dass

$$M^T u = \begin{pmatrix} b^T u \\ -A^T u \\ u \end{pmatrix}.$$

In Fall (1) setzen wir wieder $\widetilde{x} := \frac{x}{x_0} \geq 0$ und beobachten, dass

$$0 = \frac{1}{x_0} M(x_0, x, z)^T = \frac{1}{x_0}(x_0 b - Ax + z) = b - A\widetilde{x} + \frac{z}{x_0},$$

also

$$A\widetilde{x} = b + \underbrace{\frac{z}{x_0}}_{\geq 0} \geq b,$$

und somit (i) gilt. □

2.8 Dualitätssatz

Wir entwickeln eine Theorie 'dualer' Probleme und studieren sie dann aus der Perspektive des Simplexalgorithmus.

2.8.1 Formulierung und Beweis des Dualitätssatzes

Gegeben sei ein lineares Optimierungsproblem wie zuvor, in dem eine lineare Zielfunktion

$$f(x) = b_1 x_1 + ... + b_n x_n$$

zu maximieren ist unter den Restriktionen $x_1 \geq 0$, ..., $x_n \geq 0$ und

$$f_i(x) = a_{i1} x_1 + ... + a_{in} x_n \leq c_i, \quad 1 \leq i \leq m,$$

in Kurzschreibweise:

$$\text{Maximiere } b^T x \text{ unter den Bedingungen } x \geq 0 \text{ und } Ax \leq c, \tag{2.11}$$

wobei

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad c = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

gegeben sind und $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$. Im Kontext von Dualitätsdiskussionen nennen wir (2.11) auch das *primale Problem*.

Definition 2.8.1. Das zu (2.11) *duale Problem* lautet:

$$\text{Minimiere } c^T y \text{ unter den Bedingungen } y \geq 0 \text{ und } A^T y \geq b, \quad y \in \mathbb{R}^m.$$

Bemerkung 2.8.2.

(i) Durch Übergang zu $-c$, $-A$, $-b$ kann man auch das duale Problem wieder wie gewohnt schreiben, nämlich:

$$\text{Maximiere } (-c)^T y \text{ unter den Bedingungen } y \geq 0 \text{ und } (-A)^T y \leq -b.$$

(ii) Das zum dualen Problem duale Problem ist wieder das primale Problem.

Im Folgenden schreiben wir

$$P := \{x \in \mathbb{R}^n : \ x \geq 0, \ Ax \leq c\}$$

und

$$\widetilde{P} := \left\{y \in \mathbb{R}^m : \ y \geq 0, \ A^T y \geq b\right\}$$

für die zulässigen Bereiche des primalen und des dualen Problems. Wir beobachten folgenden *schwachen Dualitätssatz*:

Lemma 2.8.3. *Für alle $x \in P$ und $y \in \widetilde{P}$ gilt*

$$b^T x \leq c^T y. \tag{2.12}$$

Beweis. Man hat, da $x \geq 0$ und $y \geq 0$ sind,

$$b^T x \leq (A^T y)^T x = y^T (Ax) \leq y^T c = c^T y.$$

□

Bemerkung 2.8.4.

(i) Ist also $\widetilde{P} \neq \emptyset$, so ist die Zielfunktion $b^T x$ des primalen Problems nach oben beschränkt. Analog: Ist $P \neq \emptyset$, so ist die Zielfunktion $c^T y$ des dualen Problems nach unten beschränkt.

(ii) Gibt es $x_0 \in P$ und $y_0 \in \widetilde{P}$ mit

$$b^T x_0 \geq c^T y_0,$$

so folgt mit (2.12), dass

$$b^T x_0 \geq c^T y_0 \geq b^T x, \quad x \in P,$$

und

$$c^T y_0 \leq b^T x_0 \leq c^T y, \quad y \in \widetilde{P},$$

dass heisst, x_0 ist Lösung des primalen Problems und y_0 is Lösung des dualen, und wegen (2.12) gilt

$$b^T x_0 = c^T y_0.$$

Die optimalen Lösungen des primalen und des dualen Problems führen also auf denselben Wert (der jeweiligen Zielfunktion).

Wir haben also folgende Aussage gezeigt:

Satz 2.8.5. *Gibt es $x_0 \in P$ und $y_0 \in \widetilde{P}$ mit $b^T x_0 \geq c^T y_0$, so folgt*

$$b^T x_0 = \max_{x \in P} b^T x = \min_{y \in \widetilde{P}} c^T y = c^T y_0.$$

Wir beweisen nun folgenden *Dualitätssatz* über die Äquivalenz der Lösbarkeiten.

Satz 2.8.6. *Folgende Aussagen sind äquivalent:*

(i) Das primale Problem besitzt eine Lösung $x_0 \in P$,

$$b^T x_0 = \max_{x \in P} b^T x.$$

(ii) Das duale Problem besitzt eine Lösung $y_0 \in \widetilde{P}$,

$$c^T y_0 = \min_{y \in \widetilde{P}} c^T y.$$

(iii) Es gilt $P \neq \emptyset$ und $\widetilde{P} \neq \emptyset$.

Falls eine (und damit alle) dieser Aussagen gelten, so folgt, dass

$$b^T x_0 = c^T y_0. \tag{2.13}$$

Beweis. Wir zeigen (i) $\Rightarrow$ (iii). Gilt (i), so ist $P \neq \emptyset$. Setzen

$$S := \{u \geq 0 : \ Au \leq 0\}.$$

Sind $x \in P$ und $u \in S$, dann ist $x + u \in P$, denn $x + u \geq 0$ und

$$A(x+u) = \underbrace{Ax}_{\leq c} + \underbrace{Au}_{\leq 0} \leq c.$$

Man kann nun schliessen, dass

$$b^T u \leq 0 \quad \text{für alle } u \in S.$$

Wäre nämlich $b^T u > 0$ für ein $u \in S$, so wäre $nu \in S$ für beliebiges $n \in \mathbb{N}$ und somit $x_0 + nu \in P$, das aber würde bedeuten, dass

$$b^T(x_0 + nu) = b^T x_0 + n b^T u$$

unendlich gross würde für $n \to \infty$, das Problem also unbeschränkt wäre und somit nicht lösbar. Somit gilt also:

$$\text{Für alle } u \in \mathbb{R}^n \text{ mit } u \geq 0 \text{ und } u^T A^T \leq 0 \text{ ist } u^T b \leq 0.$$

Nach Satz 2.7.6 (Minkowski-Farkas) ist das aber äquivalent dazu, dass ein $y \geq 0$ existiert mit $A^T y \geq b$. Somit muss also $\widetilde{P} \neq \emptyset$ sein.

Wir zeigen (iii) $\Rightarrow$ (i). Gilt $\widetilde{P} \neq \emptyset$, so folgt nach Bemerkung 2.8.4 (i), dass die Zielfunktion $b^T x$ nach oben beschränkt ist auf P. Dann muss sie aber ihr Maximum auf $P \neq \emptyset$ annehmen (wie früher bereits argumentiert), und somit folgt (i).

Ganz analog sieht man, dass auch die Äquivalenz (ii) $\Leftrightarrow$ (iii) gilt.

Es bleibt zu zeigen, dass (2.13) folgt. Nach Satz 2.8.5 genügt es zu zeigen, dass es $x_0 \in \mathbb{R}^n$ und $y_0 \in \mathbb{R}^m$ gibt mit

$$Ax_0 \leq c, \quad x_0 \geq 0 \quad (\text{also } x_0 \in P),$$

$$A^T y_0 \geq b, \quad y_0 \geq 0 \quad (\text{also } y_0 \in \widetilde{P}),$$

und

$$b^T x_0 \geq c^T y_0.$$

Gesucht ist also eine Lösung $(x_0, y_0)^T$ der linearen Ungleichung

$$\underbrace{\begin{pmatrix} -A & 0 \\ 0 & A^T \\ b^T & -c^T \end{pmatrix}}_{=:D} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \geq \begin{pmatrix} -c \\ b \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} x \\ y \end{pmatrix} \geq 0,$$

hier ist $(x, y)^T \in \mathbb{R}^{n+m}$ und D ist eine $((m+n+1) \times (m+n))$-Matrix. Nach Satz 2.7.6 (Minkowski-Farkas) existiert genau dann eine solche Lösung, wenn für alle $u \in \mathbb{R}^{m+n+1}$ mit $u \geq 0$ und $D^T u \leq 0$ die Ungleichung

$$\begin{pmatrix} -c \\ b \\ 0 \end{pmatrix}^T \cdot u \leq 0$$

gilt. Genau das weisen wir jetzt nach. Sei

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \geq 0 \quad \text{und} \quad D^T u = \begin{pmatrix} -A^T & 0 & b \\ 0 & A & -c \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \leq 0,$$

hierbei ist $u_1 \in \mathbb{R}^m$, $u_2 \in \mathbb{R}^n$ und $u_3 \in \mathbb{R}$. Dann folgt

$$-A^T u_1 + u_3 b \leq 0 \quad \text{und} \quad Au_2 - u_3 c \leq 0, \tag{2.14}$$

und nach Multiplikation von links mit $u_2^T \geq 0$ bzw. $u_1^T \geq 0$ folgt

$$-u_2^T A^T u_1 + u_3 u_2^T b \leq 0 \quad \text{und} \quad u_1^T A u_2 - u_3 u_1^T c \leq 0$$

und somit

$$u_3 u_2^T b \leq \underbrace{u_2^T A^T u_1}_{\in \mathbb{R}} = (u_2^T A^T u_1)^T = u_1^T A u_2 \leq u_3 u_1^T c.$$

Falls nun $u_3 > 0$ ist, dann folgt $u_2^T b \leq u_1^T c$, also

$$\begin{pmatrix} -c \\ b \\ 0 \end{pmatrix}^T \cdot u = -c^T u_1 + b^T u_2 \leq 0,$$

wie gewünscht.

Ist hingegen $u_3 = 0$, so folgt wegen (2.14), dass

$$-A^T u_1 \leq 0 \quad \text{und} \quad Au_2 \leq 0,$$

gleichzeitig hat man nach Voraussetzung $u_1 \geq 0$ und $u_2 \geq 0$. Da gemäss (iii) $P \neq 0$ ist, gibt es $x \geq 0$ mit $Ax \leq c$, also $-Ax \geq -c$, und mit Satz 2.7.6 folgt

$$u_1^T \cdot (-c) \leq 0.$$

Da (iii) auch $\widetilde{P} \neq 0$ garantiert, gibt es $y \geq 0$ mit $A^T y \geq b$, und Satz 2.7.6 liefert

$$u_2^T \cdot b \leq 0.$$

Addition ergibt nun

$$\begin{pmatrix} -c \\ b \\ 0 \end{pmatrix}^T \cdot u = -c^T u_1 + b^T u_2 \leq 0.$$

□

Bemerkung 2.8.7. Nach Satz 2.8.6 können also vier Fälle auftreten:

(1) Man hat $P = \emptyset$ und $\widetilde{P} = \emptyset$.

(2) Man hat $P \neq \emptyset$ und $\widetilde{P} = \emptyset$. Dann ist das primäre Problem unbeschränkt.

(3) Man hat $P = \emptyset$ und $\widetilde{P} \neq \emptyset$. Dann ist das duale Problem unbeschränkt.

(4) Man hat $P \neq \emptyset$ und $\widetilde{P} \neq \emptyset$. Dann sind beide Probleme lösbar (wie im Satz).

Insbesondere gilt also:

Ist $P \neq \emptyset$, so ist das primale Problem genau dann lösbar, wenn $\widetilde{P} \neq \emptyset$.

und analog

Ist $\widetilde{P} \neq \emptyset$, so ist das duale Problem genau dann lösbar, wenn $P \neq \emptyset$.

2.8.2 Übertragung auf den Simplexalgorithmus

Gegeben ist das primale Problem, $b^T x$ zu maximieren unter den Bedingungen $x \geq 0$ und $Ax \leq c$. Das ist (wie wir wissen) beschrieben durch das Ausgangstableau

	$-x_1$	$\cdots$	$-x_n$	
y_1	a_{11}	$\cdots$	a_{1n}	c_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_m	a_{m1}	$\cdots$	a_{mn}	c_m
f	$-b_1$	$\cdots$	$-b_n$	0

(S_0)

Das duale Problem, $c^T y$ zu minimieren unter den Bedingungen $y \geq 0$ und $A^T y \geq b$ kann man durch eine modifizierte Interpretation des Ausgangstableaus beschreiben,

x_1	$\cdots$	x_n	g	
a_{11}	$\cdots$	a_{1n}	c_1	y_1
$\vdots$		$\vdots$	$\vdots$	$\vdots$
a_{m1}	$\cdots$	a_{mn}	c_m	y_m
$-b_1$	$\cdots$	$-b_n$	0	

$(\widetilde{S}_0)$

Die *duale Interpretation* ist wie folgt:

Wir fassen $y_1, ..., y_m$ als Linearformen auf $\mathbb{R}^m$ auf, $y_i : \mathbb{R}^m \to \mathbb{R}$, $y_i(y) = y_i$, (genauso wie vormals die x_i) und definieren affin lineare Funktionen

$$x_j(y) := \sum_{i=1}^{m} a_{ij} y_i - b_j, \quad 1 \leq j \leq n,$$

und

$$g(y) := \sum_{i=1}^{m} c_i y_i.$$

Für den zulässigen Bereich gilt dann

$$\widetilde{P} = \underbrace{\{x_1 \geq 0, ..., x_n \geq 0\}}_{=\{A^T y \geq b\}} \cap \{y_1 \geq 0, ..., y_m \geq 0\}.$$

Führt man nun einen Austausschritt (oder mehrere Austausschritte) nach Pivotregeln durch, so erhält man ein neues Tableau

$$\begin{array}{|ccc|c|c|} \hline u_1 & \cdots & u_n & g & \\ \hline \alpha_{11} & \cdots & \alpha_{1n} & \gamma_1 & v_1 \\ \vdots & & \vdots & \vdots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} & \gamma_m & v_m \\ \hline \beta_1 & \cdots & \beta_n & \delta & \\ \hline \end{array} \qquad (\widetilde{S})$$

Proposition 2.8.8. *Die Gleichungen*

$$u_j = \sum_{i=1}^{m} \alpha_{ij} v_i + \beta_j, \quad 1 \leq j \leq n,$$

und

$$g = \sum_{i=1}^{m} \gamma_i v_i + \delta$$

des Schemas $(\widetilde{S})$ sind äquivalent zu den Gleichungen des Schemas $(\widetilde{S}_0)$. Insbesondere gilt

$$\widetilde{P} = \{u_1 \geq 0, ..., u_n \geq 0\} \cap \{v_1 \geq 0, ..., v_m \geq 0\}.$$

Beweis. Wir müssen zeigen, dass nach einem Austauschschritt (gemäss Pivotregeln) auch für das neue Schema die duale Interpretation gültig bleibt. Nehmen wir also an, u_l und v_k werden getauscht. Dann wird v_k im neuen Schema durch die Variablen v_i, $i \neq k$, und

$$u_l = \sum_{i=1}^{m} \alpha_{il} v_i + \beta_l$$

(Darstellung von u_l im alten Schema) ausgedrückt. Man hat

$$v_k = \frac{1}{\alpha_{kl}} u_l - \sum_{i \neq k} \frac{\alpha_{il}}{\alpha_{kl}} v_i - \frac{\beta_l}{\alpha_{kl}},$$

wobei die Koeffizienten $\frac{1}{\alpha_{kl}}$, $-\frac{\alpha_{il}}{\alpha_{kl}}$ und $-\frac{\beta_l}{\alpha_{kl}}$ die Elemente der neuen Pivotspalte sind. Für $j \neq l$ gilt

$$\begin{aligned} u_j &= \sum_{i=1}^{m} \alpha_{ij} v_i + \beta_j \\ &= \alpha_{kj} v_k + \sum_{i \neq k} \alpha_{ij} v_i + \beta_j \\ &= \frac{\alpha_{kj}}{\alpha_{kl}} u_l + \sum_{i \neq k} \left(\alpha_{ij} - \frac{\alpha_{il}\alpha_{kj}}{\alpha_{kl}} \right) v_i + \left(\beta_j - \frac{\beta_l \alpha_{kj}}{\alpha_{kl}} \right), \end{aligned}$$

hierbei sind die Koeffizienten

$$\frac{\alpha_{kj}}{\alpha_{kl}}, \quad \left(\alpha_{ij} - \frac{\alpha_{il}\alpha_{kj}}{\alpha_{kl}} \right), \quad \left(\beta_j - \frac{\beta_l \alpha_{kj}}{\alpha_{kl}} \right)$$

nun die Einträge der neuen j-ten Spalte. Analog ergibt sich

$$\begin{aligned} g &= \sum_{i=1}^{m} \gamma_i v_i + \delta \\ &= \gamma_k v_k + \sum_{i \neq k} \gamma_i v_i + \delta \\ &= \frac{\gamma_k}{\alpha_{kl}} u_l + \sum_{i \neq k} \left(\gamma_i - \frac{\alpha_{il}}{\alpha_{kl}} \gamma_k \right) v_i + \left(\delta - \frac{\beta_l \gamma_k}{\alpha_{kl}} \right), \end{aligned}$$

wobei die Koeffizienten nun die Einträge der neuen letzten Spalte sind. □

Korollar 2.8.9. *Startet man mit einem zulässigen Ausgangsschema für das primale Problem und findet man mittels Simplexalgorithmus das dafür optimale Schema (S), dann gilt:*

(i) Man hat $\gamma_1 \geq 0$, ..., $\gamma_m \geq 0$, denn (S) ist ein zulässiges Schema.

(ii) Man hat $\beta_1 \geq 0$, ..., $\beta_n \geq 0$, denn das Schema ist optimal.

(iii) Auch für (S) gilt die duale Interpretation.

Nach Aussage (iii) in diesem Korollar kann man nun feststellen:

(a) Der Fakt, dass $\beta_1 \geq 0$, ..., $\beta_n \geq 0$ gilt impliziert, dass, wenn man die Variablen v_1, ..., v_m alle gleich Null setzt, man

$$u_1 = \beta_1 \geq 0, \quad ..., \quad u_m = \beta_m \geq 0$$

erhält. Somit beschreibt das Schema (S) einen Punkt $y_0 \in \widetilde{P}$. Wir können das auch dadurch ausdrücken, dass wir sagen, *(S) ist zulässig für das duale Problem.*

(b) Der Fakt dass $\gamma_1 \geq 0$, ..., $\gamma_m \geq 0$ gilt bedeutet, dass *(S) optimal ist für das duale Problem*, d.h. die Funktion $g|_{\widetilde{P}}$ nimmt ihr Minimum δ in y_0 an, denn:

$$g = \underbrace{\gamma_1}_{\geq 0} v_1 + ... + \underbrace{\gamma_m}_{\geq 0} v_m + \delta$$

wird minimal für $v_1 = ... = v_m = 0$.

Wir fassen das in folgendem Satz zusammen. Er besagt, dass wir mit dem Simplexalgorithmus tatsächlich zwei Probleme gleichzeitig lösen.

Satz 2.8.10. *Findet man mit dem Simplexalgorithmus ein optimales zulässiges Tableau, so bestimmt dies simultan Lösungen des primalen und des dualen Problems. Für das duale Problem gilt: Der minimale Wert der Zielfunktion ist δ, und dieser wird angenommen im Punkt $y_0 \in \mathbb{R}^m$ mit*

$$(y_0)_k = \begin{cases} 0 & \text{falls } y_k = v_i \text{ für ein } i \\ \beta_j & \text{falls } y_k = u_j \text{ für ein } j. \end{cases}$$

Beispiel 2.8.11. Betrachten wir das Problem, $y_1 + y_2$, $y = (y_1, y_2) \in \mathbb{R}^2$, zu minimieren unter den Bedingungen $y_1 \geq 0$, $y_2 \geq 0$ und

$$\begin{aligned} 4\,y_1 + y_2 &\geq 2 \\ 3\,y_1 + 2\,y_2 &\geq 4 \\ y_1 + 8\,y_2 &\geq 10 \end{aligned}$$

oder, in Normalform: Wir suchen das Maximum von $-y_1 - y_2$ unter den Bedingungen $y_1 \geq 0$, $y_2 \geq 0$ und

$$\begin{aligned} -4\,y_1 - y_2 &\leq -2 \\ -3\,y_1 - 2\,y_2 &\leq -4 \\ -y_1 - 8\,y_2 &\leq -10. \end{aligned}$$

Das zugehörige Ausgangsschema

	$-y_1$	$-y_2$	
	-4	-1	-2
	-3	-2	-4
	-1	-8	-10
f	1	1	0

ist nicht zulässig. Aber: Dieses Problem ist das duale Problem dazu, die Funktion $2\,x_1 + 4\,x_2 + 10\,x_3$, $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ zu minimieren unter den Bedingungen $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ und

$$\begin{aligned} 4\,x_1 + 3\,x_2 + x_3 &\leq 1 \\ x_1 + 2\,x_2 + 8\,x_3 &\leq 1. \end{aligned}$$

Dieses liefert das zulässige Ausgangsschema

	$-x_1$	$-x_2$	$-x_3$	
y_1	4	3	1	1
y_2	1	2	8	1
	-2	-4	-10	0

und Austausch mit Pivot 8 liefert

	$-x_1$	$-x_2$	$-y_2$	
y_1	$\frac{31}{8}$	$\frac{22}{8}$	$-\frac{1}{8}$	$\frac{7}{8}$
x_3	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
	$-\frac{6}{8}$	$-\frac{12}{8}$	$\frac{10}{8}$	$\frac{10}{8}$

Ein weiterer Austausch mit Pivot $\frac{22}{8}$ ergibt das optimale Schema

	$-x_1$	$-y_1$	$-y_2$	
x_2	#	#	#	$\frac{7}{22}$
x_3	#	#	#	$\frac{1}{22}$
	$\frac{30}{22}$	$\frac{12}{22}$	$\frac{26}{22}$	$\frac{38}{22}$

Hier ist der maximale Wert also $\frac{19}{11} = \frac{38}{22}$, und er ergibt sich in $x_0 = (0, \frac{7}{22}, \frac{1}{22})$. Die duale Interpretation

x_1	y_1	y_2		
#	#	#	$\frac{7}{22}$	x_2
#	#	#	$\frac{1}{22}$	x_3
$\frac{30}{22}$	$\frac{12}{22}$	$\frac{26}{22}$	$\frac{38}{22}$	

dazu liefert nun die Lösung des ursprünglichen Problems, nämlich den minimalen Wert $\frac{19}{11}$ für $y_1 + y_2$ auf dem zulässigen Bereich im Punkt $y_0 = (\frac{12}{22}, \frac{26}{22})$.

2.9 Interpretation des dualen Problems

Die besprochene Dualität besitzt eine ökonomische bzw. strategische Interpretation.

Beispiel 2.9.1. Wir erinnern uns an Beispiel 2.1.1. Für den Landwirt gibt es zwei mögliche *Aktivitäten*, nämlich den Anbau von Kartoffeln auf x_1 ha und den Anbau von Getreide auf x_2 ha. Offensichtlich sind also $x_1 \geq 0$ und $x_2 \geq 0$. Er kann über drei beschränkte *Ressourcen* verfügen, nämlich

$$\begin{aligned} \text{Kapital:} \quad & 100\,x_1 + 200\,x_2 \leq 11000 \\ \text{Arbeitszeit:} \quad & x_1 + 4\,x_2 \leq 160 \\ \text{Land:} \quad & x_1 + x_2 \leq 100. \end{aligned}$$

Der Gewinn ergibt sich als

$$f(x_1, x_2) = 400\,x_1 + 1200\,x_2.$$

Wir hatten gesehen, dass sich der maximale Gewinn 54000 im Punkt $(60, 25)$ ergibt.

Man kann sich nun fragen: Inwiefern ist es sinnvoll, die Beschränkungen für die Ressourcen zu lockern, d.h. die Schranken 11000, 160 und 100 zu vergrössern? Welcher Preis wäre fr den zukauf von Ressourcen gerechtfertigt?

An der Maximalstelle gilt in unserem Beispiel: die Restriktionen für Kapital und Arbeitszeit sind *straff*, d.h. mit Gleichheit erfüllt,

$$\begin{aligned} 100 \cdot 60 + 200 \cdot 25 &= 11000 \\ 60 + 4 \cdot 25 &= 160. \end{aligned}$$

Die Restriktion für Land ist nicht straff, man hat

$$60 + 25 < 100.$$

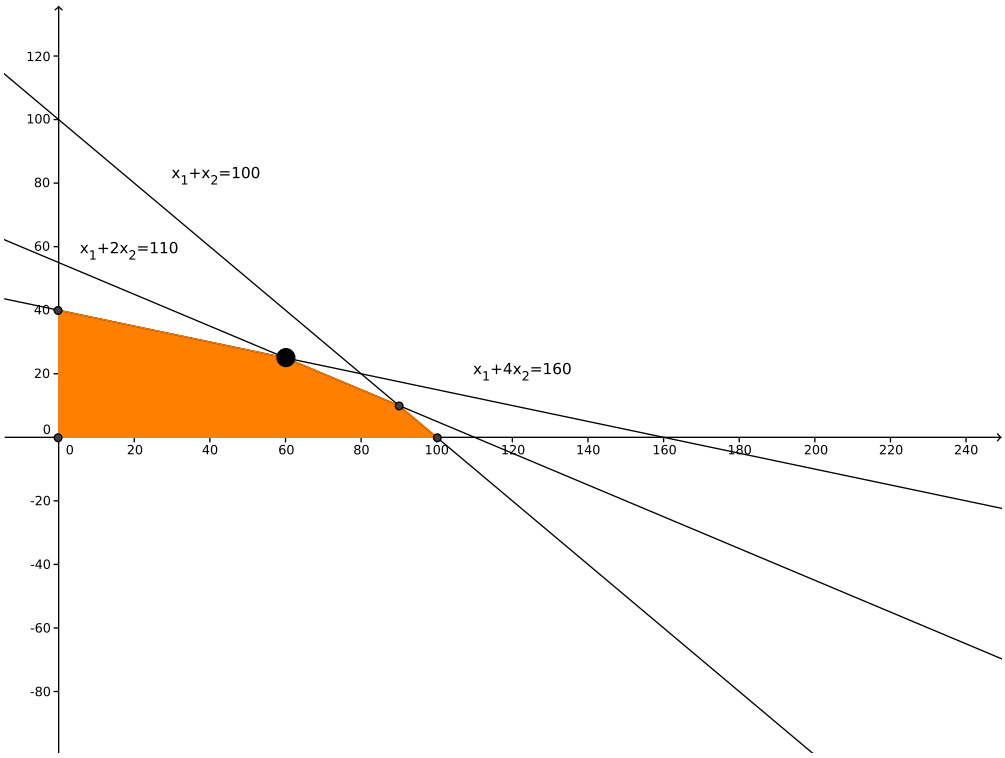


Abbildung 2.9: Beschränkte Ressourcen Kapital, Arbeitszeit und Land.

Es würde also nichts bringen, Land dazuzukaufen.

Welcher Preis je Einheit wäre für zusätzliches Kapital und zusätzliche Arbeitszeit gerechtfertigt ? Wir suchen eine Preisvektor

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

hier stehen y_1, y_2 und y_3 jeweils für die Preise für eine Einheit Kapital, Arbeitszeit und Land.

Die Kosten am Markt für den Zukauf zur Produktion von Kartoffeln pro ha sind

$$100\, y_1 + y_2 + y_3,$$

die Kosten für die Produktion von Getreide pro ha sind

$$200\, y_1 + 4\, y_2 + y_3.$$

Falls diese Kosten den Gewinn, den der Landwirt je ha erwirtschaften kann, übersteigen, wird er nicht kaufen. Das also ist der Fall wenn

$$\begin{aligned} 100\, y_1 + y_2 + y_3 &\geq 400 \\ 200\, y_1 + 4\, y_2 + y_3 &\geq 1200. \end{aligned}$$

gilt. Dies sind die Restriktionen des dualen Problems. Der Landwirt wird unter diesen Umständen aber eventeuell seine Ressourcen verkaufen. Käufer*innen werden interessiert sein, den Gesamtpreis der Ressourcen

$$11000\, y_1 + 160\, y_2 + 100\, y_3$$

zu minimieren. Dieser definiert die Zielfunktion des dualen Problems. Der Preisvektor y ist also die Lösung des dualen Problems sein.

Das Ausgangsschema des primalen Problems war

	$-x_1$	$-x_2$	
y_1	100	200	11000
y_2	1	4	160
y_3	1	1	100
	-400	-1200	0

gewesen, und das daraus folgende optimale Schema

	$(-)y_1$	$(-)y_2$	
x_1	#	#	60
x_2	#	#	25
y_3	#	#	15
	2	200	54000

Daraus kann man nun ablesen, dass sich das Minimum für das duale Problem in $y = (y_1, y_2, y_3)^T$ mit

$$y_1 = 2, \quad y_2 = 200, \quad y_3 = 0$$

ergibt. Dass $y_3 = 0$ ist korrespondiert zu der nicht straffen Restriktion für Land: braches Land steht zur Verfügung, Käufer*innen sind nicht bereit, Geld für Land auszugeben.

Der Preisvektor y beschreibt keinen realen Preis, sondern die Preise, die zu bezahlen sinnvoll wäre, um die jeweiligen Ressourcen zu vergrössern. Man nennt solche fiktiven Preise auch *Schattenpreise*.

Bemerkung 2.9.2. Ist eine Restriktion für das primale Problem nicht straff an der Maximalstelle, so gilt für die zugehörige Variable y_i des dualen Problems an der Minimalstelle $y_i = 0$ (Satz vom schwachen komplementären Schlupf).

Allgemein kann man folgende *Aktivitätsanalyse* formulieren:

Gegeben sind n *Aktivitäten*, durch welche sich eine Ware (z.B. Geld) erzeugen lässt, und dazu stehen m *Ressourcen* zur Verfügung. Die Koeffizienten des Problems kann man wie folgt interpretieren:

c_i ist der Vorrat, der von der i-ten Ressource zu Verfügung steht.

a_{ij} ist die Menge der i-ten Ressource, welche bei der j-ten Aktivität verbraucht wird.

b_j ist die Menge der Ware, die mit der j-ten Aktivität produziert wird.

Hierbei sind $1 \leq i \leq m$ und $1 \leq j \leq n$.

Gesucht sind die *Aktivitätsniveaus* $x_j \geq 0$, $1 \leq j \leq n$, mit

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

d.h. derart, dass nicht mehr verbraucht wird, als von der jeweiligen Ressource zur Vefügung steht, und für welche die Menge der produzierten Ware maximal wird,

$$\sum_{j=1}^{n} b_j x_j \stackrel{!}{=} \max .$$

Beim dazu dualen Problem sucht man *Schattenpreise* y_i pro Einheit der i-ten Ressource (und damit also einen Preisvektor y), sodass der Gesamtpreis aller Materialien minimiert wird,

$$\sum_{i=1}^{m} c_i y_i \stackrel{!}{=} \min,$$

und die Nebenbedingungen

$$\sum_{i=1}^{m} a_{ij} y_i \geq b_j, \quad 1 \leq j \leq n,$$

erfüllt sind; letztere besagen, dass die Gesamtkosten der j-ten Aktivität mindestens b_j sein sollen.

Primales und duales Problem können Rollen tauschen, d.h. statt *'Gewinnmaximierung ohne Erschöpfung der Ressourcen'* dann *'Aufwandsminimierung unter Einhaltung minimaler Standards'*.

Beispiel 2.9.3. Die Lebensmittel A und B enthalten die Nährstoffe I, II, III und IV gemäss fogender Tabelle (in Nährstoffeinheiten pro Mengeneinheit des Lebensmittels):

	I	II	III	IV
A	2.5	2.0	1.5	1.0
B	0.5	1.0	1.5	3.0

Ausserdem wissen wir, dass Lebensmittel A den Preis 20 (pro Mengeneinheit) hat und Lebensmittel B den Preis 10. Man sucht eine preisgünstigste Kombination der beiden Lebensmittel, welche aber insgesamt mindestens 140 Einheiten von I, 300 Einheiten von II, 270 Einheiten von III und 300 Einheiten von IV enthält.

Bezeichnen x_1 und x_2 die Einheiten von A bzw. B in der Kombination, so führt das auf das lineare Optimierungsproblem, $20x_1+10x_2$ zu minimieren unter den Nebenbedingungen $x_1 \geq 0$, $x_2 \geq 0$ und

$$\begin{aligned} 2.5\, x_1 + 0.5\, x_2 &\geq 140 \\ 2.0\, x_1 + 1.0\, x_2 &\geq 300 \\ 1.5\, x_1 + 1.5\, x_2 &\geq 270 \\ 1.0\, x_1 + 3.0\, x_2 &\geq 300. \end{aligned}$$

Das dazu duale Problem kann man nun wie folgt interpretieren:

Man sucht die Preise, zu welchen es sich lohnt, Nährstoffe I, II, III und IV zuzukaufen, oder, anders gesagt, die Preise y_I, y_{II}, y_{III} und y_{IV}, welche ein Pharma-Hersteller für Präparate mit jeweils einer Einheit I, II, III bzw. IV verlangen kann.

Die Präparate dürfen als Nährstoffquellen insgesamt nicht teurer sein als A und B (sonst kaufen wir sie nicht), das ergibt die Restriktionen

$$2.5\, y_I + 2.0\, y_{II} + 1.5\, y_{III} + 1.0\, y_{IV} \leq 20$$
$$0.5\, y_I + 1.0\, y_{II} + 1.5\, y_{III} + 3.0\, y_{IV} \leq 10.$$

Der Verkaufserlös

$$140\, y_I + 300\, y_{II} + 270\, y_{III} + 300\, y_{IV}$$

soll dabei maximiert werden.

Nun kann man den Simplexalgorithmus ansetzen. Das erste Problem führt nach Transformation auf Normalform auf ein nicht zulässiges Ausgangsschema. Wir betrachten deshalb lieber das Ausgangsschema für das zweite Problem:

	$-y_I$	$-y_{II}$	$-y_{III}$	$-y_{IV}$	
x_1	2.5	2.0	1.5	1.0	20
x_2	0.5	1.0	1.5	3.0	10
	-140	-300	-270	-300	0

Ein Austauschschritt mit Pivot 1.0 liefert das Schema

	$-y_I$	$-x_2$	$-y_{III}$	$-y_{IV}$	
x_1	1.5	-2.0	-1.5	-5	0
y_{II}	0.5	1.0	1.5	3.0	10
	10	300	180	600	3000

Die duale Interpretation dieses Schemas erlaubt es nun, die Lösung für das erste Problem abzulesen, nämlich

$$x_1 = 0 \quad \text{und} \quad x_2 = 300.$$

Als Schattenpreise für die Präparate ergeben sich

$$y_I = y_{II} = y_{IV} = 0 \quad \text{und} \quad y_{III} = 10.$$

Wir geben eine *weitere Interpretation der Schattenpreise*. Ist ein primales Problem gegeben, in welchem $b^T x$ maximiert werden soll unter den Bedingungen $x \geq 0$ und $Ax \leq c$, dann kann man fragen:

Wie verändert sich der Wert der Zielfunktion an der Maximalstelle, wenn man die Schranken c für die Ressourcen durch $c + \delta$ ersetzt, $\delta \in \mathbb{R}^m$ klein ?

Angenommen, das duale Problem besitzt die *eindeutige* Lösung $y_0 \in \widetilde{P}_e$, für die dann gilt, dass

$$c^T y_0 < c^T y, \quad y \in \widetilde{P}_e \setminus \{y_0\}.$$

Durch Variation von c verändert sich der zulässige Bereich

$$\widetilde{P} = \{y \geq 0,\ A^T y \geq b\}$$

des dualen Problems nicht. Also gilt für δ hinreichend klein, dass

$$(c+\delta)^T y_0 < (c+\delta)^T y, \quad y \in \widetilde{P}_e \setminus \{y_0\}.$$

Das bedeutet aber, dass y_0 auch noch optimal ist nach dem Übergang von c zu $c+\delta$.

Mit dem Dualitätssatz sehen wir aber, dass die Änderung des optimalen Wertes der primalen Zielfunktion gleich der Änderung des optimalen Wertes der dualen Zielfunktion ist, und das ist gerade

$$(c+\delta)^T y_0 - c^T y_0 = \delta^T y_0 = \sum_{i=1}^{m} \delta_i (y_0)_i.$$

V11 Der Schattenpreis $(y_0)_i$ gibt also an, *welchen Beitrag eine Vergrösserung der i-ten Ressource zum Gesamterlös ergibt.*

2.10 Beschreibung allgemeiner Polyeder

Wir schauen uns die Struktur von Polyedern etwas genauer an, insbesondere mit dem Ziel, sie für unbeschränkte Polyeder noch besser zu verstehen.

2.10.1 Notation und Wiederholung

Zur Erinnerung:

Eine Linearkombination $\sum_{i=1}^m \lambda_i a_i$ von Elementen $a_1, ..., a_m \in \mathbb{R}^n$ heisst *Affinkombination*, falls $\sum_{i=1}^m \lambda_i = 1$.

Die Menge aller Affinkombinationen von Elementen einer Menge $S \subset \mathbb{R}^n$ heisst die *affine Hülle* von S, wir schreiben dafür aff S.

Eine Teilmenge $M \subset \mathbb{R}^n$ heisst *affiner Unterraum* von $\mathbb{R}^n$, falls sie ihre eigene affine Hülle ist, $M = \text{aff}\, M$.

Es gilt:

(i) M ist genau dann ein affiner Unterraum von $\mathbb{R}^n$, wenn $M = a + U$ gilt für ein $a \in \mathbb{R}^n$ und einen linearen Unterraum U von $\mathbb{R}^n$. Der lineare Unterraum U ist dabei eindeutig bestimmt als $U = M - M$. Die *Dimension* von M ist definiert als

$$\dim M := \dim U.$$

(ii) Der Schnitt affiner Unterräume ist wieder ein affiner Unterraum.

(iii) Die affine Hülle einer Menge $S \subset \mathbb{R}^n$ ist der kleinste affine Unterraum von $\mathbb{R}^n$, der S enthält.

2.10.2 Darstellungssatz für Polyeder

Wir beobachten nun folgendes.

Lemma 2.10.1. *Für jede Seite S einer konvexen Menge $A \subset \mathbb{R}^n$ gilt*

$$S = A \cap \operatorname{aff} S.$$

(Geht man von der trivialen Identität $S = A \cap S$ aus, macht also der Übergang von S zu aff S auf der rechten Seite die Gleichheit nicht kaputt.)

Beweis. Offensichtlich ist $S \subset A \cap \operatorname{aff} S$. Um auch die umgekehrte Inklusion zu zeigen, sei nun $x \in A \cap \operatorname{aff} S$. Dann ist x von der Form

$$x = \sum_{i=1}^{m} \mu_i y_i \quad \text{mit } \mu_i \in \mathbb{R} \text{ sodass } \sum_{i=1}^{m} \mu_i = 1 \text{ und } y_i \in S.$$

Falls $\mu_i \geq 0$ gilt für alle i, so ist $x \in k(S)$, und da S konvex ist, folgt $x \in S = k(S)$. Falls es ein i gibt mit $\mu_i < 0$, dann ist

$$\alpha := -\sum_{i:\mu_i<0} \mu_i > 0,$$

und die Punkte

$$z_1 := \frac{1}{1+\alpha} \sum_{i:y_i>0} \mu_i y_i \quad \text{und} \quad z_2 := -\frac{1}{\alpha} \sum_{i:\mu_i<0} \mu_i y_i$$

sind Elemente der konvexen Hülle von $y_1, ..., y_m$, also

$$z_1, z_2 \in k(\{y_1, ..., y_m\}) \subset S.$$

Nun gilt aber

$$z_1 = \frac{1}{1+\alpha} \left(x - \sum_{i:\mu_i<0} \mu_i y_i \right) = \frac{1}{1+\alpha}\, x + \frac{\alpha}{1+\alpha}\, z_2 \quad \in \quad]x, z_2].$$

Weil $z_1 \in S$ ist und andererseits $x, z_2 \in A$, so folgt daraus $x \in S$, denn S ist eine Seite von A. □

Sei nun

$$P := \bigcap_{i=1}^{m} \{f_i \leq c_i\}$$

ein gegebenes Polyeder.

Definition 2.10.2. Für jedes i nennen wir

$$H_i := \{f_i = c_i\}$$

eine *Restriktionshyperebene (RHE)* von P.

Man kann nun eine Beschreibung der affinen Hülle einer (nichtleeren) Seite von P als Schnitt 'guter' RHE erhalten.

Satz 2.10.3. *Sei $S \neq \emptyset$ eine Seite von P und $I := \{i \in \{1,...,m\} : \ S \subset H_i\}$. Dann ist*

$$\operatorname{aff} S = \bigcap_{i \in I} H_i.$$

Beweis. Sei $M := \bigcap_{i\in I} H_i$. Da $S \subset M$ ist, folgt $\operatorname{aff} S \subset \operatorname{aff} M = M$.

Sei daher umgekehrt $x \in M$. Wir schreiben $J := \{1,...,m\} \setminus I$ und zeigen nun, dass es ein $y \in S$ gibt mit $f_j(y) < c_j$ für alle $j \in J$.

(Mit diesem y basteln wir uns dann eine Darstellung von x, die den gewünschten Schluss erlaubt.)

Für jedes $j \in J$ gibt es ein $y_j \in S$ mit $f_j(y_j) < c_j$. Wir setzen

$$y := \frac{1}{k} \sum_{j \in J} y_j \quad \in S, \quad \text{wobei } k = \#J \text{ ist.}$$

Dann gilt in der Tat für jedes $j_0 \in J$, dass

$$f_{j_0}(y) = \frac{1}{k}\Big(\underbrace{f_{j_0}(y_{j_0})}_{<c_{j_0}} + \sum_{j \in J\setminus\{j_0\}} \underbrace{f_{j_0}(y_j)}_{\leq c_{j_0}} \Big) < \frac{1}{k}\Big(c_{j_0} + \sum_{j\in J\setminus\{j_0\}} c_{j_0} \Big) = c_{j_0}.$$

Wir betrachten nun die Punkte

$$z_1 := y + \varepsilon(y - x) \quad \text{und} \quad z_2 := y - \varepsilon(y - x) \tag{2.15}$$

mit $0 < \varepsilon < 1$ so klein, dass

$$f_j(z_1), f_j(z_2) < c_j \text{ für alle } j \in J. \tag{2.16}$$

Da

$$f_i(x) = c_i \quad \text{für alle } i \in I \text{ wegen } x \in M$$

und

$$f_i(y) = c_i \quad \text{für alle } i \in I \text{ wegen } y \in S \subset H_i \text{ (nach Definition von } I),$$

hat man auch

$$f_i(z_1) = f_i(z_2) = c_i \text{ für alle } i \in I. \tag{2.17}$$

Zusammen erlauben (2.16) und (2.17) zu schliessen, dass $z_1, z_2 \in P$ gilt. Somit hat man (wegen (2.15))

$$y = \frac{1}{2}z_1 + \frac{1}{2}z_2 \in S,$$

und weil S eine Seite ist, folgt $z_1, z_2 \in S$. Das zeigt aber, dass

$$x = y + \frac{1}{2\varepsilon}(z_2 - z_1) \quad \in \quad \operatorname{aff}\{y, z_1, z_2\} \subset \operatorname{aff} S.$$

Hier haben wir wieder (2.15) benutzt, und $1 + \frac{1}{2\varepsilon} - \frac{1}{2\varepsilon} = 1$. □

Mit Lemma 2.10.1 und Satz 2.10.3 folgt nun eine elegante Beschreibung einer (nichtleeren) Seite von P als Schnitt von P selbst und 'guten' RHE.

Korollar 2.10.4. *Für jede Seite $S \neq \emptyset$ von P gilt*

$$S = P \cap \bigcap_{i : S \subset H_i} H_i,$$

wobei die H_i die RHE von P bezeichnen.

Bemerkung 2.10.5.

(i) Speziell für $S = P$ folgt aus Satz 2.10.3, dass

$$\operatorname{aff} P = \bigcap_{i:P\subset H_i} H_i.$$

Die RHE, in denen P gänzlich enthalten ist (d.h. $P \subset H_i$), heissen *singuläre* RHE.

(ii) Wir definieren die Dimension von P als

$$\dim P := \dim \operatorname{aff} P.$$

Wegen (i) kann $\dim P$ kleiner sein als $n = \dim \mathbb{R}^n$. Falls $\dim P < n$ hätte P, als Teilmenge von $\mathbb{R}^n$ gesehen, dann allerdings nie innere Punkte. Für topologische Fragen ist es daher günstiger, eine Einschränkung auf den Unterraum $\operatorname{aff} P$ vorzunehmen. Wir definieren das *relative Innere von* P als

$$\operatorname{int}_r P := \{x \in P : \ B_\delta(x) \cap \operatorname{aff} P \subset P \text{ für ein } \delta > 0\}$$

und den *relativen Rand von* P als

$$\partial_r P := P \setminus \operatorname{int}_r P.$$

Da $\operatorname{aff} P \subset \mathbb{R}^n$ abgeschlossen ist, stimmen Abschluss und relativer Abschluss einer Menge in $\operatorname{aff} P$ überein. Insbesondere ist also

$$\partial_r P = \overline{P} \setminus \operatorname{int}_r P = P \setminus \operatorname{int}_r P.$$

(iii) Für $P = \bigcap_{i=1}^m \{f_i \le c_i\}$ ist

$$\operatorname{int}_r P = P \cap \bigcap_{i:P\not\subset H_i} \{f_i < c_i\}.$$

Hierbei wird der Schnitt also über alle nichtsingulären RHE genommen.

(iv) Jeder Punkt $x \in \partial_r P$ ist Element einer *echten Seite* S von P, d.h. einer Seite $S \neq P$ von P, denn:

Aus $x \in \partial_r P$ folgt, dass $x \in P$, aber auch $x \notin \operatorname{int}_r P$. Wegen (iii) gibt es also eine nichtsinguläre RHE mit $\{f_i = c_i\}$ mit $f_i(x) = c_i$. Dann ist aber $S := P \cap \{f_i = c_i\}$ eine Seite von P und $x \in S$. Da $\{f_i = c_i\}$ nichtsingulär ist, muss S ein echte Seite sein.

(v) Umgekehrt gilt für jede echte Seite S von P, dass

$$S \subset \partial_r P.$$

Korollar 2.10.6. *Ist S eine echte Seite von P, so gilt*

$$\dim S < \dim P.$$

Beweis. Nach Satz 2.10.3 ist

$$\bigcap_{S \subset H_i} H_i = \operatorname{aff} S \subset \operatorname{aff} P = \bigcap_{P \subset H_i} H_i,$$

und nach Korollar 2.10.4 gilt

$$P \cap \bigcap_{S \subset H_i} H_i = S \subsetneq P = \bigcap_{P \subset H_i} H_i.$$

Daraus folgt, dass $\operatorname{aff} S \subsetneq \operatorname{aff} P$ ist, und das impliziert $\dim S < \dim P$. □

V12

Wir können nun überlegen, wie man ein Polyeder aus seinen Seiten reproduzieren kann. Da die Definition des Begriffs 'Polyeder' eher weit ist, funktioniert das nicht immer. Es gibt Polyeder, die man nicht im Sinne der folgenden Diskussion 'aus kleineren Teilen zusammensetzen kann', die also nicht reduzierbar sind. Andererseits eignen sich solche Teile vielleicht besonders gut als 'Einzelbausteine'. Die Situation ist ideell also ähnlich wie bei Primzahlen und Primfaktorzerlegung.

Definition 2.10.7. Ein nichtleeres Polyeder heisst *primitiv*, wenn es nicht die konvexe Hülle seiner echten Seiten ist.

Beispiel 2.10.8.

(i) Affine Unterräume (z.B. Punkte, Geraden, Ebenen) sind primitive Polyeder.

(ii) Affine Halbräume, d.h. nichtleere Schnitte eines affinen Unterraums M mit einem Halbraum $\{f \leq c\}$ mit $M \not\subset \{f \leq c\}$ (z.B. Halbgeraden, Halbebenen), sind primitive Polyeder.

Wir hatten nämlich gesehen, dass echte Seiten im relativen Rand enthalten sind. Dieser ist im Fall (i) leer und im Fall (ii) in $\{f = c\}$ enthalten.

Bemerkung 2.10.9. Jeder affine Halbraum ist die Summe einer Halbgeraden und eines Untervektorraumes, d.h. von der Gestalt

$$x_0 + \mathbb{R}_+ \cdot v + U,$$

hier ist x_0 ein fixierter Punkt, $\mathbb{R}_+ \cdot v$ eine Halbgerade (definiert durch einen gegebenen Vektor v), und U ist ein linearer Unterraum. (Übung)

Satz 2.10.10. *Jedes Polyeder $P \neq \emptyset$ ist die konvexe Hülle seiner primitiven Seiten.*

Beweis. Ist P selbst primitiv, so ist nichts zu zeigen. Wir dürfen also annehmen, dass P nicht primitiv ist und somit die konvexe Hülle seiner echten Seiten S. Weil $\dim S < \dim P$, so folgt mit Induktion nach $\dim P$: Jede echte Seite von S ist die konvexe Hülle der primitiven Seiten von S. Seiten von S sind aber auch Seiten von P. Daraus folgt die Behauptung. (Ist $\dim S = 0$, so ist S ein Punkt, also primitiv.) □

Man kann nun sogar zeigen, dass mehr als die o.g. Beispiele für primitive Polyeder gar nicht vorkommen:

Satz 2.10.11. *Ein primitives Polyeder P ist entweder ein affiner Unterraum oder ein affiner Halbraum.*

Beweis. Falls $\partial_r P = \emptyset$ dann ist $P = \text{int}_r P$. Da $\text{int}_r P$ offen ist in aff P und P abgeschlossen in aff P, folgt daraus, dass diese Menge $\emptyset$ sein muss oder aff P (denn das sind die einzigen Teilmengen von aff P, die gleichzeitig offen und abgeschlossen sind). Da nach Definition P nichtleer ist, muss also

$$P = \text{int}_r P = \text{aff } P$$

gelten, wie gewünscht.

Falls $\partial_r P \neq \emptyset$, so gibt es einen Punkt $x \in \partial_r P$, insbesondere also $x \notin \text{int}_r P$. Nach Bemerkung 2.10.5 (iii) gibt es dann eine nichtsinguläre RHE H mit $x \in H$.

Wir beweisen nun zunächst folgende *Behauptung*: Es gibt genau eine nichtsinguläre RHE, die ganz $\partial_r P$ umfasst.

Das kann man wie folgt sehen: Angenommen, es gibt nichtsinguläre RHE $H_1 = \{f_1 = c_1\}$ und $H_2 = \{f_2 = c_2\}$ und $v_1, v_2 \in \partial_r P$ mit $v_1 \in H_1 \setminus H_2$ und $v_2 \in H_2 \setminus H_1$. Dann folgt

$$f_1(v_1) = c_1 \quad \text{und} \quad f_2(v_1) < c_2$$

sowie

$$f_1(v_2) < c_1 \quad \text{und} \quad f_2(v_2) = c_2.$$

Somit also $f_1(v_1 - v_2) > 0$ und $f_2(v_1 - v_2) < 0$. Sei nun $u \in P$ beliebig. Dann gibt es $\alpha_1, \alpha_2 > 0$ sodass

$$f_1(\underbrace{u + \alpha_1(v_1 - v_2)}_{=:y_1}) > c_1 \quad \text{und} \quad f_2(\underbrace{u + \alpha_2(v_2 - v_1)}_{=:y_2}) > c_2.$$

Das bedeutet, wir haben $u \in P$ und $y_1, y_2 \notin P$, und $u \in [y_1, y_2]$. Dann müssen aber

$$\widetilde{y}_1 \in [u, y_1] \quad \text{und} \quad \widetilde{y}_2 \in [u, y_2]$$

existieren mit $\widetilde{y}_1, \widetilde{y}_2 \in \partial_r P$. Nach Bemerkung 2.10.5 (iv) liegen $\widetilde{y}_1$, $\widetilde{y}_2$ also in echten Seiten von P. Andererseits hat man aber $u \in [\widetilde{y}_1, \widetilde{y}_2]$, somit ist u also Konvexkombi von Punkten echter Seiten von P. Da aber $u \in P$ beliebig war, folgt, dass P nicht primitiv sein kann, Widerspruch. Also ist obige Behauptung wahr.

Sei nun also $H =: \{f = c\}$ die nichtsinguläre RHE mit $\partial_r P \subset H$. Wir schreiben $\widetilde{H} := \{f \leq c\}$ und $M := \text{aff } P$. Offensichtlich ist $P \subset \widetilde{H}$.

Wir stellen nun eine weitere *Behauptung* auf, nämlich, dass $P = M \cap \widetilde{H}$ ist. Aus dieser Behauptung folgt offensichtlich der Satz.

Um sie zu zeigen, bemerken wir zunächst, dass $P \subset M \cap \widetilde{H}$ gilt. Wir müssen die umgekehrte Inklusion zeigen. Angenommen, es gibt einen Punkt $y \in (M \cap \widetilde{H}) \setminus P$. Da H nichtsingulär ist, muss es ein $u \in P \setminus H$ geben. Also hat man $u \in P$, $y \in M \setminus P$, und damit existiert zwangsläufig ein $x \in [u, y] \subset M$ mit $x \in \partial_r P$, somit $x \in M \cap H$. Nun gilt

$$x \neq u \text{ (da } u \in P \setminus H) \text{ und } x \neq y \text{ (da } x \in \partial_r P \subset P \text{ und } y \notin P).$$

Nach Satz 2.2.13 ist $M \cap H$ eine Seite von $M \cap \widetilde{H}$, denn

$$M \cap H = \underbrace{M \cap \widetilde{H}}_{\subset \{f \leq c\}} \cap \{f = c\}.$$

Also muss wegen $x \in M \cap H$ und $u, y \in M \cap \widetilde{H}$ dann gelten, dass $u \in M \cap H$ ist. Das aber widerspricht $u \notin H$.

Somit muss $M \cap \widetilde{H} \subset P$ gelten und damit die letzte Behauptung. □

Aus den Sätzen 2.10.10 und 2.10.11 erhalten wir nun einen ersten *Darstellungssatz für Polyeder*:

Korollar 2.10.12. *Jedes Polyeder ist die konvexe Hülle endlich vieler affiner Unterräume und affiner Halbräume.*

Wir kommen nochmal zurück zum Begriff des konvexen Kegels und formalisieren ihn nun nochmal richtig.

Definition 2.10.13.

(i) Eine Linearkombination $\sum_{i=1}^{m} \lambda_i a_i$ von Vektoren $a_i \in \mathbb{R}^n$ heisst *konisch*, falls $\lambda_i \geq 0$ für alle i.

(ii) Eine konvexe Menge $K \subset \mathbb{R}^n$ ist ein *konvexer Kegel*, falls für jedes $x \in K$ und $\lambda \geq 0$ auch $\lambda x \in K$ gilt.

(iii) Für $S \subset \mathbb{R}^n$ heisst

$$\text{cone}(S) := \Big\{ \sum_{i=1}^{m} \lambda_i a_i : \ m \in \mathbb{N},\ a_i \in S,\ \lambda_i \geq 0,\ i = 1, ..., m \Big\}$$

der von S erzeugte konvexe Kegel.

Bemerkung 2.10.14.

(i) $\text{cone}(S)$ ist der kleinste konvexe Kegel, der S enthält.

(ii) Man hat $0 \in K$ für jeden konvexen Kegel K.

(iii) Für eine reelle $(m \times n)$-Matrix ist $\{Ax \leq 0\}$ ein *polyedrischer Kegel*, d.h. ein konvexer Kegel, der auch ein Polyeder ist.

Lemma 2.10.15. *Jeder affine Unterraum und jeder affine Halbraum ist von der Form*

$$u + \text{cone}(Y)$$

mit $u \in \mathbb{R}^n$ und $Y = \{y_1, ..., y_m\} \subset \mathbb{R}^n$ endlich.

Beweis. Sei $M = x_0 + U$ ein affiner Unterraum, wobei $x_0 \in M$ und U ein Untervektorraum ist. Ist $\{b_1, ..., b_k\}$ eine Basis von U, so ist

$$U = \text{cone}(\{b_1, ..., b_k, -b_1, ..., -b_k\}).$$

Daraus folgt die Behauptung mit $u = x_0$ und $Y = \{b_1, ..., b_k, -b_1, ..., -b_k\}$.

Ist $\widetilde{H}$ ein affiner Halbraum, dann ist er nach Bemerkung 2.10.9 von der Gestalt

$$\widetilde{H} = x_0 + \mathbb{R}_+ \cdot v + U$$

mit $x_0, v \in \mathbb{R}^n$ und einem Untervektorraum U. Ist nun $\{b_1, ..., b_k\}$ eine Basis von U, dann folgt

$$\widetilde{H} = x_0 + \text{cone}(\{v, b_1, ..., b_k, -b_1, ..., -b_k\})$$

und damit die Behauptung. □

V13

Wir erhalten folgenden *Darstellungssatz für Polyeder*:

Satz 2.10.16. *Für jedes Polyeder P gibt es endliche Mengen $X \subset \mathbb{R}^n$ und $Y \subset \mathbb{R}^n$, sodass*

$$P = k(X) + \text{cone}(Y).$$

Beweis. Sei o.E. $P \neq \emptyset$ (sonst wähle $X = Y = \emptyset$). Nach Korollar 2.10.12 gibt es affine Unter- oder Halbräume $A_1, ..., A_p$, deren konvexe Hülle P ist. Für jedes $i = 1, ..., p$ gilt nach Lemma 2.10.15

$$A_i = u_i + \text{cone}(Y_i)$$

mit geeigneten $u_i \in P$ und $Y_i = \{y_1^{(i)}, ..., y_{m(i)}^{(i)}\} \subset \mathbb{R}^n$.

Sei nun $x \in P$ beliebig. Dann gilt

$$x = \sum_{i=1}^{p} \lambda_i a_i \quad \text{ mit } a_i \in A_i,\ \lambda_i \geq 0,\ \sum_{i=1}^{p} \lambda_i = 1.$$

Weiter findet man für jedes $i = 1, ..., p$ Koeffizienten $\mu_j^{(i)} \geq 0$ sodass

$$a_i = u_i + \sum_{j=1}^{m(i)} \mu_j^{(i)} y_j^{(i)}.$$

Also folgt

$$x = \sum_{i=1}^{p} \lambda_i u_i + \sum_{i=1}^{p} \sum_{j=1}^{m(i)} \lambda_i \mu_j^{(i)} y_j^{(i)} \quad \in \quad k(\{u_1, ..., u_p\}) + \text{cone}(Y_1 \cup ... \cup Y_p).$$

Das bedeutet, wir haben $P \subset k(X) + \text{cone}(Y)$ mit $X = \{u_1, ..., u_p\}$ und $Y = Y_1 \cup ... \cup Y_p$.

Wir zeigen nun, dass auch $k(X) + \text{cone}(Y) \subset P$ gilt. Sei $x \in k(X) + \text{cone}(Y)$, also

$$x = u + \sum_j \lambda_j y_j$$

mit $u \in k(u_1, ..., u_p)$, $y_j \in Y_1 \cup ... \cup Y_p$, $\lambda_j \geq 0$. Setze $\lambda := \sum_j \lambda_j$.

Falls $\lambda = 0$, so ist $x = u \in k(\{u_1, ..., u_p\}) \subset P$, denn $u_1, ..., u_p \in P$. Ist $\lambda > 0$, so betrachten wir $\mu_j := \frac{\lambda_j}{\lambda}$. Dann gilt $\mu_j \geq 0$ für alle j und $\sum_j \mu_j = 1$, und ausserdem ist

$$x = \sum_j \mu_j (u + \lambda y_j).$$

Falls wir nun zeigen können, dass $u + \lambda y_j \in P$ ist für alle j, dann impliziert das $x \in P$. Seien i_0 und j_0 so, dass $y_j = y_{j_0}^{(i_0)}$ ist. Dann hat man $A_{i_0} = u_{i_0} + \text{cone}(Y_{i_0})$ und daher $u_{i_0} + \lambda y_j \in P$ für alle $\lambda \geq 0$. Sei für $0 < \alpha < 1$ nun

$$x_\alpha := u_{i_0} + \lambda y_j + (1 - \alpha)(u - u_{i_0}),$$

für $\alpha \to 0$ konvergiert das gegen $u + \lambda y_j$. Da P abgeschlossen ist, hat man also $u + \lambda y_j \in P$, sobald man zeigen kann, dass $x\alpha \in P$ ist für alle hinreichend kleinen α. Nun ist aber

$$x_\alpha = \alpha(u_{i_0} + \frac{\lambda}{\alpha} y_j) + (1 - \alpha)u \quad \in \quad P,$$

denn es ist Konvexkombination von $u_{i_0} + \frac{\lambda}{\alpha} y_j \in P$ und $u \in P$. Also folgt insgesamt, dass $x \in P$. □

Korollar 2.10.17. *Ist S ein polyedrischer Kegel, so gibt es eine endliche Menge $Y = \{y_1, ..., y_m\} \subset \mathbb{R}^n$ mit*

$$S = \text{cone}(Y).$$

V14 **Übung 2.10.18.** *Man beweise dieses Korollar.*

2.10.3 Lösbarkeit von Optimierungsproblemen

Satz 2.10.16 erlaubt uns nun, eine frühere Behauptung zur Lösbarkeit linearer Optimierungsprobleme zu beweisen:

Satz 2.10.19. *Sei $P \neq \emptyset$ ein Polyeder und f eine Linearform, die auf P nach oben beschränkt ist. Dann nimmt $f|_P$ auf P ihr Maximum an.*

Beweis. Sei $f(x) = b^T x$. Nach Satz 2.10.16 besitzt $x \in P$ die Darstellung

$$x = \sum_{i=1}^{p} \lambda_i u_i + \sum_{j=1}^{l} \mu_j v_j$$

mit $\lambda_i, \mu_j \geq 0$ und $\sum_{i=1}^{p} \lambda_i = 1$. Daraus folgt

$$f(x) = \sum_{i=1}^{p} \lambda_i f(u_i) + \sum_{j=1}^{l} \mu_j f(v_j).$$

Da f nach oben beschränkt ist und μ_j beliebig gross wählbar, folgt daraus $f(v_j) \leq 0$ für alle $j = 1, ..., l$, denn: Gäbe es ein j_0 mit $f(v_{j_0}) > 0$, so könnte man $\lambda_j = 0$ wählen für $j \neq j_0$, und man müsste trotzdem

$$+\infty > \sup_P f \geq \sum_{i=1}^{p} \lambda_i f(u_i) + \lambda_{j_0} f(v_{j_0})$$

haben für beliebig grosse $\lambda_{j_0} > 0$, was unmöglich ist. Daher folgt

$$f(x) \leq \sum_{i=1}^{p} \lambda_i f(u_i) \leq \max_{i=1,...,p} f(u_i) =: f(u_{i_0})$$

für ein geeignetes $i_0 \in \{1, ..., p\}$. Also nimmt $f|_P$ ihr Maximum in u_{i_0} an. □

2.10.4 Darstellungssatz von Weyl

Satz 2.10.16 kann man sogar umkehren in der folgenden Form, man nennt diese Aussage den *Darstellungssatz von Weyl*.

Wir zeigen zunächst einen Spezialfall:

Satz 2.10.20. *Sei*

$$P := \text{cone}(Y)$$

für $Y = \{y_1, ..., y_k\} \in \mathbb{R}^n$ endlich. Dann gibt es eine reelle $(m \times n)$-Matrix A mit

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

d.h. P ist ein polyedrischer Kegel.

Satz 2.10.21. *Seien X und Y endliche Teilmengen des $\mathbb{R}^n$. Dann ist*

$$k(X) + \text{cone}(Y)$$

ein Polyeder.

Beweis. Sei $P^0 := \{z \in \mathbb{R}^n : \ y^T z \leq 0 \text{ für alle } y \in P\}$. Dann gilt $z \in P^0$ genau dann, wenn

$$\sum_{i=1}^{k} \lambda_i \, y_i^T \, z \leq 0 \quad \text{ für alle } \lambda_i \geq 0,$$

d.h. genau dann, wenn $y_i^T z \leq 0$ für $i = 1, ..., k$, und das kann man auch schreiben als

$$Bz \leq 0,$$

wobei B die reelle $(k \times n)$-Matrix

$$B = \begin{pmatrix} y_1^T \\ \vdots \\ y_k^T \end{pmatrix}$$

ist. Also gilt

$$P^0 = \{z \in \mathbb{R}^n : \ Bz \leq 0\},$$

und das ist ein polyedrischer Kegel. Nach Korollar 2.10.17 gibt es also eine Menge $Z = \{z_1, ..., z_m\} \subset \mathbb{R}^n$ mit

$$P^0 = \text{cone}(Z).$$

Sei nun $P^{00} := (P^0)^0 = \{x \in \mathbb{R}^n : \ x^T z \leq 0 \text{ für alle } z \in P^0\}$. Dann folgt ganz analog wie eben, dass

$$P^{00} = \{x \in \mathbb{R}^n : \ Ax \leq 0\}$$

mit der reellen $(m \times n)$-Matrix

$$A = \begin{pmatrix} z_1^T \\ \vdots \\ z_m^T \end{pmatrix}.$$

Wir zeigen nun $P^{00} = P$, und daraus folgt dann die Behauptung des Satzes.

Nach Voraussetzung ist $b \in P$ genau dann, wenn $b \in \text{cone}(\{y_1, ..., y_k\})$, und das ist genau dann der Fall, wenn es ein $x \geq 0$ gibt mit $B^T x = b$. Nach Satz 2.7.5 (Minkowski-Farkas für lineare Gleichungen) ist das äquivalent dazu, dass für alle $u \in \mathbb{R}^n$ mit $Bu \leq 0$ die Ungleichung $b^T u \leq 0$ gilt. Das wiederum ist genau dann der Fall, wenn für alle $u \in P^0$ gilt, dass $b^T u \leq 0$, also genau dann, wenn $b \in P^{00}$.

□

Der allgemeine Satz 2.10.21 folgt nun so:

Beweis. Seien $X = \{x_1, ..., x_k\}$ und $Y = \{y_1, ..., y_l\}$ endliche Teilmengen des $\mathbb{R}^n$ wie im Satz. Wir betrachten den Kegel

$$C := \text{cone}\left(\left\{\begin{pmatrix} x \\ 1 \end{pmatrix} : x \in X\right\} \cup \left\{\begin{pmatrix} y \\ 0 \end{pmatrix} : y \in Y\right\}\right)$$

in $\mathbb{R}^{n+1}$. Sei $H := \{z_{n+1} = 1\} \subset \mathbb{R}^{n+1}$. Dann gilt

$$\begin{pmatrix} z \\ z_{n+1} \end{pmatrix} \in C$$

genau dann, wenn

$$\begin{pmatrix} z \\ z_{n+1} \end{pmatrix} = \sum_{i=1}^{k} \underbrace{\lambda_i}_{\geq 0} \begin{pmatrix} x_i \\ 1 \end{pmatrix} + \sum_{j=1}^{l} \underbrace{\mu_j}_{\geq 0} \begin{pmatrix} y_j \\ 0 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{k} \lambda_i x_i + \sum_{j=1}^{l} \mu_j y_j \\ \sum_{i=1}^{k} \lambda_i \end{pmatrix}.$$

Daher ist also

$$\begin{pmatrix} z \\ z_{n+1} \end{pmatrix} \in C \cap H$$

genau dann, wenn

$$\sum_{i=1}^{k} \lambda_i = 1 \quad \text{und} \quad z \in k(X) + \text{cone}(Y) = P.$$

Nach Satz 2.10.20 ist C also die Lösungsmenge eines Ungleichungssystems der Form

$$a_{i1}\, x_1 + ... + a_{in}\, x_n + a_{in+1}\, x_{n+1} \leq 0, \quad i = 1, ..., m.$$

Das bedeutet aber insbesondere, dass P die Lösungsmenge ist von

$$a_{i1}\, x_1 + ... + a_{in}\, x_n \leq -a_{in+1} =: c_i, \quad i = 1, ..., m.$$

□

Das folgende Charakterisierung beschränkter Polyeder ist eine einfache Konsequenz des Umstandes, dass cone(Y) unbeschränkt ist für alle $Y \neq \emptyset$.

Korollar 2.10.22. *Die folgenden Aussagen sind äquivalent:*

(i) P ist ein beschränktes Polyeder.

(ii) P ist die konvexe Hülle endlich vieler Punkte.

Die Darstellung eines Polyeders wie in Satz 2.10.16 und Satz 2.10.21 ist eindeutig in folgendem Sinne:

Lemma 2.10.23. *Sei P ein Polyeder. In der Zerlegung*

$$P = K + S$$

mit $K = k(\{x_1, ..., x_k\})$ und $S = \text{cone}(\{y_1, ..., y_l\})$ ist S eindeutig bestimmt. Für jedes $x_0 \in P$ gilt

$$S = \{y \in \mathbb{R}^n : \ x_0 + \lambda y \in P \quad \textit{für alle } \lambda \geq 0\}.$$

Ist $P = \bigcap_{i=1}^{m} \{f_i \leq c_i\}$, so gilt $S = \bigcap_{i=1}^{m} \{f_i \leq 0\}$.

Beweis. Sei

$$S_x := \{y \in \mathbb{R}^n : x + \lambda y \in P \quad \text{für alle } \lambda \geq 0\}.$$

Wie am Ende des Beweises von Satz 2.10.16 kann man sehen, dass für jedes $x \in P$ die Gleichheit $S_x = S_{x_0}$ gilt. (Übungsaufgabe.)

Wir behaupten nun, dass $S = S_{x_0}$ sein muss. Für $y \in S$ gilt nach Voraussetzung, dass $x_0 + \lambda y \in P$ ist für alle $\lambda \geq 0$. Somit ist $S \subset S_{x_0}$ klar.

Um die umgekehrte Inklusion zu zeigen, sei nun $y_0 \in S_{x_0}$. Angenommen, $y_0 \notin S$. Dann hat

$$Ax = y_0, \quad x \geq 0$$

keine Lösung x; hier ist A die reelle $(n \times l)$-Matrix $A := (y_1, ..., y_l)$. Nach Satz 2.7.5 (Minkowski-Farkas) gibt es dann $u \in \mathbb{R}^n$ mit $u^T y_0 > 0$ und $u^T A \leq 0$, und die letzte Bedingung ist äquivalent zu $u^T y_j \leq 0$, $j = 1, ..., l$. Folglich gilt für alle $x \in P$, also für alle

$$x = \sum_{i=1}^{k} \lambda_i x_i + \sum_{j=1}^{l} \mu_j y_j \quad \text{mit} \quad \lambda_i, \mu_j \geq 0, \quad \sum_{i=1}^{k} \lambda_i = 1,$$

dass

$$u^T x = \sum_{i=1}^{k} \lambda_i u^T x_i + \sum_{j=1}^{l} \mu_j \underbrace{u^T y_j}_{\leq 0} \leq \sum_{i=1}^{k} \lambda_i u^T x_i \leq \max_{1 \leq i \leq k} u^T x_i < +\infty.$$

Wegen $y_0 \in S_{x_0} = S_x$ ist aber $x + \lambda y_0 \in P$ für alle $\lambda \geq 0$, und wegen $u^T y_0 > 0$ würde damit folgen, dass

$$u^T(x + \lambda y_0) \to +\infty \quad \text{für } \lambda \to +\infty,$$

offensichtlich ein Widerspruch. Also muss gelten, dass $y_0 \in S$, und somit $S_{x_0} \subset S$, was die obige Behauptung zeigt.

Sei schliesslich $P = \bigcap_{i=1}^{m} \{f_i \leq c_i\}$. Man hat $y \in S$ genau dann, wenn $x_0 + \lambda y \in P$ für alle $\lambda \geq 0$, und dass ist genau dann der Fall, wenn

$$\underbrace{f_i(x_0)}_{\leq c_i} + \lambda f_i(y) = f_i(x_0 + \lambda y) \leq c_i \quad \text{für alle } i \text{ und alle } \lambda \geq 0.$$

Das ist aber äquivalent zu $f_i(y) \leq 0$, $i = 1, ..., m$. □

Wir schauen als nächstes, unter welchen Umständen ein allgemeines Polyeder P überhaupt Extremalpunkte ('Ecken') hat.

Lemma 2.10.24. *Sei S ein polyedrischer Kegel. Dann ist*

$$L := S \cap (-S)$$

der grösste in S enthaltene Untervektorraum.

Beweis. Tatsächlich ist L ein Untervektorraum: Sind $x, y \in L = S \cap (-S)$, so folgt $x + y \in S \cap (-S) = L$ (Ist S Kegel, so ist auch $-S$ ein Kegel). Für $\lambda \geq 0$ and $x \in L$ folgt sofort, dass $\lambda x \in S \cap (-S) = L$. Für $\lambda < 0$ und $x \in L$ hat man $\lambda x \in (-S)$, da $x \in S$, und $\lambda x \in S$, da $x \in (-S)$. Also $\lambda x \in L$.

Der Untervektorraum ist maximal: Ist $W \subset S$ ein weiterer Untervektorraum, so gilt $W \subset S \cap (-S) = L$. □

Definition 2.10.25. Sei $P = K + S$ ein Polyeder, wobei K ein beschränktes Polyeder ist und S der gemäss Lemma 2.10.23 eindeutig bestimmte polyedrische Kegel.

(i) Der Vektorraum $L = S \cap (-S)$ heisst der *Linienraum* von P.

(ii) Das Polyeder P heisst *spitz*, falls $P_e \neq \emptyset$.

Satz 2.10.26. *Die folgenden Aussagen sind äquivalent:*

(i) Das Polyeder P ist spitz.

(ii) Der Linienraum von P ist trivial, $L = \{0\}$.

(iii) Ist $P = \{x \in \mathbb{R}^n : \ Ax \leq c\}$, so ist der Rang von A gleich n.

Bemerkung 2.10.27. Die zusätzlichen Nebenbedingungen $x \geq 0$ in (iii) für ein lineares Optimierungsproblem stellen also sicher, dass der zulässige Bereich spitz ist.

Wir beweisen Satz 2.10.26.

Beweis. Wir beweisen zunächst die Äquivalenz von (i) und (ii). Sei $L \neq \{0\}$ und $0 \neq w \in L$. Dann sind $w, -w \in S$. Für beliebiges $x_0 \in P$ folgt: $x_0 + w \in P$ und $x_0 - W \in P$, also

$$x_0 = \frac{1}{2}(x_0 + w) + \frac{1}{2}(x_0 - w) \notin P_e.$$

Das geht aber nur, wenn $P_e = \emptyset$. Sei nun $L = \{0\}$. Nach Satz 2.10.10 ist P die konvexe Hülle seiner primitiven Seiten $\widetilde{S}$. Da nach Satz 2.10.11 eine primitive Seite nur ein affiner Unterraum oder affiner Halbraum sein kann, andererseits aber $L = \{0\}$ ist, können nur Punkte oder Halbgeraden als primitive Seiten $\widetilde{S}$ von P vorkommen. (Denn: ist W ein Untervektorraum und $x_0 + W \subset \widetilde{S} \subset P$, so folgt nach Lemma 2.10.23 insbesondere $W \subset S$, also $W \subset L$, somit $W = \{0\}$.) Da eine Halbgerade einen Extremalpunkt hat und das Polyeder nichtleer ist, hat P nach Satz 2.2.15 auch Extremalpunkte, also $P_e \neq \emptyset$.

Um die Äquivalenz von (ii) und (iii) zu sehen, bemerken wir, dass

$$L = S \cap (-S) = \{x \in \mathbb{R}^n : \ Ax \leq 0\} \cap \{x \in \mathbb{R}^n : \ -Ax \leq 0\}$$
$$= \{x \in \mathbb{R}^n : \ Ax = 0\} = \ker A.$$

Also ist $L = \{0\}$ genau dann, wenn $\ker A = \{0\}$, also genau dann, wenn der Rang von A gleich n ist. □

V15

2.10.5 Anwendungen auf lineare Optimierungsprobleme

Wir schauen uns Anwendungen auf lineare Optimierungsprobleme an. Angenommen, f soll maximiert werden unter den Bedingungen

$$f_i(x) \leq c_i, \quad i = 1, ..., m.$$

Dann ist der zulässige Bereich $P = \bigcap_{i=1}^m \{f_i \leq c_i\}$, und er nimmt die eindeutige Zerlegung $P = K + S$ an (wie in Lemma 2.10.23). Der Linienraum ist $L = S \cap (-S)$.

Satz 2.10.28.

(i) Wir haben $\sup_P f < +\infty$ genau dann, wenn $f|_S \leq 0$. In diesem Falle nimmt die Funktion $f|_P$ ihr Maximum an.

(ii) Ist $f|_L \neq 0$, so ist $\sup_P f = +\infty$.

(iii) Ist P spitz und $\sup_P f < +\infty$, so mimmt $f|_P$ ihr Maximum in einem Extremalpunkt von P an.

Beweis. Um (i) zu sehen, sei $x = k+s$ mit $k \in K$ und $s \in S$. Dann gilt $f(x) = f(k)+f(s)$. Da K kompakt ist, gilt $\sup_K |f| < +\infty$, also ist $\sup_P f < +\infty$ genau dann, wenn $\sup_S f < +\infty$. Dies ist äquivalent zu $f|_S \leq 0$, denn: Falls $f(s) > 0$ für ein $s \in S$, so kann man $f(\lambda s) = \lambda f(s)$ beliebig gross machen, indem man $\lambda \geq 0$ gross wählt, und das erzwingt $\sup_S f = +\infty$.

Um (ii) zu sehen nehmen wir an, dass $f(x) \neq 0$ für ein $x \in L$. In diesem Falle ist auch $-x \in L$. Also folgt, dass o.E. $f(x) > 0$ für ein $x \in L \subset S$. Dann folgt aber $\sup_P f = +\infty$ aus (i).

Wir zeigen (iii). Ist $\alpha := \sup_P f < +\infty$, dann nimmt f dieses Maximum α an nach Satz 2.10.19, und nach Satz 2.2.13 ist

$$M := P \cap \{f = \alpha\}$$

eine Seite von P und selbst ein Polyeder. Dann ist aber der Linienraum von M in jenem von P enthalten, und weil P spitz ist, muss auch M spitz sein. Also gilt $M_e \neq \emptyset$, und somit gibt es ein $x \in M_e$ sodass $f(x) = \alpha$. Wegen $M_e \subset P_e$ folgt nun (iii). □

Bemerkung 2.10.29. Wenn P nicht spitz ist, dann kann man zeigen, dass $P \cap L^\perp$ spitz sein muss. Ein nach oben beschränkte Zielfunktion f nimmt dann ihr Maximum auf P in $(P \cap L^\perp)_e$ an und ist konstant längs der Richtungen von L.

Kapitel 3

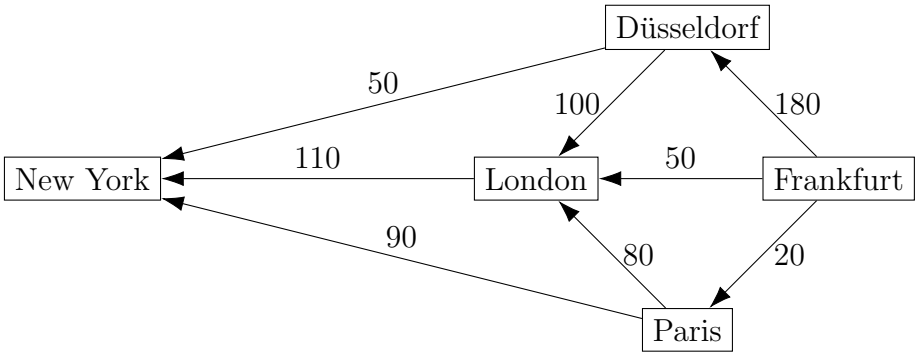
Netzplantechnik und Netzwerkoptimierung

Wir betrachten Optimierungsprobleme auf Graphen und Netzwerken.

3.1 Flussprobleme

Wir beginnen mit einem Beispiel.

Beispiel 3.1.1. Eine Airline muss zusätzlich zur bereits bestehenden Auslastung möglichst viele Mitglieder einer Reisegruppe von Frankfurt nach New York fliegen, Direktflüge sind bereits ausgebucht, aber Verbindungen mit Umsteigen sind möglich. Die vorhandenen Kapazitäten an Sitzplätzen sind wie folgt:



Die Fragen, die sich nun stellen, sind:

(a) Wieviele Passagiere können mit diesen Kapazitäten insgesamt befördert werden ?

(b) Wie ist die Beförderung zu organisieren ?

Ganz ähnliche Probleme kann man z.B. auch studieren für:

- andere Verkehrsmittel (Strassen, Bahnen, Schiffsverbindungen)
- Netze (Strom-, Wasser-, Telekommunikationsnetze)
- betriebliche Ablaufplanung (workflow, Produktionsplan)

Wir benötigen einige Grundbegriffe der Graphentheorie.

Definition 3.1.2. Ein *gerichteter Graph* ist ein Paar $G = (X, \Gamma)$, bestehend aus einer Menge X und einer Teilmenge

$$\Gamma \subset (X \times X) \setminus \{(x, x) : \ x \in X\},$$

sodass $(y, x) \notin \Gamma$ falls $(x, y) \in \Gamma$. Die Elemente $x \in X$ heissen *Knoten* oder *Punkte* des Graphen, die Elemente $\gamma \in \Gamma$ heissen *(gerichtete) Kanten*.

Ein gerichteter Graph ist also zunächst nur ein rein kombinatorisches Objekt. Für die angesprochenen Probleme verlangt man nun noch eine spezielle Struktur des Graphen und weist den Kanten noch eine quantitative Bewertung zu.

Definition 3.1.3. Ein *Netzwerk (endliches Transportnetzwerk)* ist ein Quadrupel (G, a, b, c), wobei

(i) $G = (X, \Gamma)$ ein gerichteter Graph ist mit einer endlichen Menge X,

(ii) $a, b \in X$ mit $(b, a) \in \Gamma$; a heisst *Quelle*, b *Senke*, (b, a) *Kontrollkante*,

(iii) $c : \Gamma \to \mathbb{N} \cup \{+\infty\}$ ist eine Funktion mit $c(\gamma) \in \mathbb{N}$ für alle $\gamma \in \Gamma \setminus \{(b, a)\}$ und $c((b, a)) = +\infty$; $c(\gamma)$ heisst die *Kapazität* der Kante γ.

Bemerkung 3.1.4. Diese Definition ist etwas restriktiver als andere, die man in der Literatur zur Graphentheorie findet, sie ist unserem Zweck angepasst. Die Terminologien zu Graphen unterscheiden sich in verschiedenen Texten zum Teil erheblich, man muss jeweils genau schauen, wie die Begriffe dort definiert sind.

Um die Notation zu vereinfachen, schreiben wir auch als Abkürzung auch

$$c(x, y) := c((x, y))$$

für den Wert einer Funktion c auf Γ an der Stelle $(x, y) \in \Gamma$; wir nutzen diese Abkürzung mit dieser Interpretation auch stillschweigend im Folgenden.

Definition 3.1.5.

(i) Ein *Fluss* in einem Netzwerk $((X, \Gamma), a, b, c)$ ist eine Abbildung

$$\varphi : \Gamma \to \mathbb{N},$$

welche die *Kirchhoffsche Regel* erfüllt,

$$\sum_{\gamma \in x_\rightarrow} \varphi(\gamma) = \sum_{\gamma \in x_\leftarrow} \varphi(\gamma) \quad \text{für alle } x \in X.$$

hier ist

$$x_\rightarrow := \{(x, y) \in (X \times X) : \ (x, y) \in \Gamma\}$$

die Menge aller von x ausgehenden Kanten und

$$x_\leftarrow := \{(y, x) \in (X \times X) : \ (y, x) \in \Gamma\}$$

die Menge aller zu x hinführenden Kanten.

(ii) Ein Fluss φ heisst *zulässig*, falls

$$\varphi(\gamma) \leq c(\gamma) \quad \text{für alle } \gamma \in \Gamma.$$

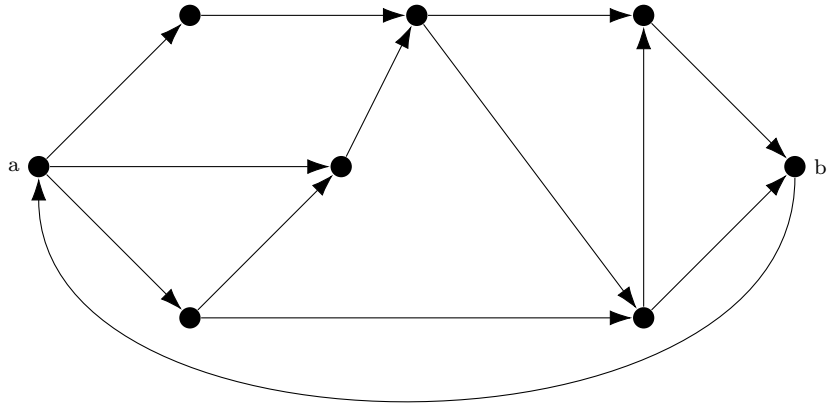
(iii) Ist φ ein Fluss, so heisst $\varphi(b,a)$ der *Wert* von φ. Ein zulässiger Fluss mit maximalem Wert heisst *Maximalfluss*.

Bemerkung 3.1.6.

(i) Ein Fluss beschreibt einen Transport von Masse (Material, Personen, Daten, etc.) von der Quelle a zur Senke b, bei dem an keiner Stelle Masse verloren geht. Letzteres ist codiert in der Kirchhoffsche Regel, die sagt, dass an jedem Knoten die eingehende Gesamtmasse gleich der ausgehenden sein muss.

(ii) Ein zulässiger Fluss überschreitet also auf keiner Kante die Kapazität. Die Kapazität kann man als das Durchleitungsvermögen der Kante (den 'Rohrquerschnitt') interpretieren.

(iii) Die Kontrollkante ist quasi formal (kommt nicht aus dem praktischen Problem) und wird zusätzlich eingeführt, damit die Kirchhoffsche Regel auch in der Quelle a und der Senke b gelten.



(iv) Der Gesamtfluss drch das Netzwerk von der Quelle a zu der Senke b fliesst über die Kontrollkante zurück, ist also gleich dem Wert $\varphi(b,a)$ des Flusses φ. Weil dieser Rückfluss durch die Kontrollkante nicht beschränkt sein soll, setzt man $c(b,a) = +\infty$.

Für eine gegebene Menge $A \subset X$ von Knoten definieren wir nun ganz allgemein die Menge

$$A_{\rightarrow} := \{(x,y) \in \Gamma : \ x \in A, y \notin A\}$$

der aus A herausführenden Kanten und die Menge

$$A_{\leftarrow} := \{(x,y) \in \Gamma : \ x \notin A, y \in A\}$$

der nach A hineinführenden Kanten.

Man kann nun eine kleine Verallgemeinerung der Kirchhoffschen Regel bekommen.

Lemma 3.1.7. *Sei φ ein Fluss und $A \subset X$. Dann hat man*

$$\sum_{\gamma \in A_{\rightarrow}} \varphi(\gamma) = \sum_{\gamma \in A_{\leftarrow}} \varphi(\gamma).$$

Beweis. Summation über alle $x \in A$ und Anwendung der Kirchhoffschen Regel für jedes solche x ergibt

$$\sum_{x\in A}\sum_{\gamma\in x_\rightarrow}\varphi(\gamma)=\sum_{x\in A}\sum_{\gamma\in x_\leftarrow}\varphi(\gamma).$$

Die linke Seite ist gleich

$$\sum_{x\in A}\sum_{\substack{y\in X\\(x,y)\in\Gamma}}\varphi(x,y)=\sum_{x\in A}\sum_{\substack{y\in A\\(x,y)\in\Gamma}}\varphi(x,y)+\underbrace{\sum_{x\in A}\sum_{\substack{y\notin A\\(x,y)\in\Gamma}}\varphi(x,y)}_{=\sum_{\gamma\in A_\rightarrow}\varphi(\gamma)},$$

die rechte Seite

$$\sum_{x\in A}\sum_{\substack{y\in X\\(y,x)\in\Gamma}}\varphi(y,x)=\sum_{x\in A}\sum_{\substack{y\in A\\(y,x)\in\Gamma}}\varphi(y,x)+\underbrace{\sum_{x\in A}\sum_{\substack{y\notin A\\(y,x)\in\Gamma}}\varphi(y,x)}_{=\sum_{\gamma\in A_\leftarrow}\varphi(\gamma)}.$$

Weil auch die ersten Summanden auf den rechten Seiten dieser beiden Identitäten gleich sind, folgt damit die Behauptung. □

Korollar 3.1.8. *Sei φ ein zulässiger Fluss und sei $A \subset X$ so, dass $a \in A$, aber $b \notin A$. Dann ist*

$$\varphi(b,a)\le\sum_{\gamma\in A_\rightarrow}c(\gamma).$$

Beweis. Da $(b,a) \in A_\leftarrow$ hat man

$$\varphi(b,a)\le\sum_{\gamma\in A_\leftarrow}\varphi(\gamma)=\sum_{\gamma\in A_\rightarrow}\varphi(\gamma)\le\sum_{\gamma\in A_\rightarrow}c(\gamma).$$

□

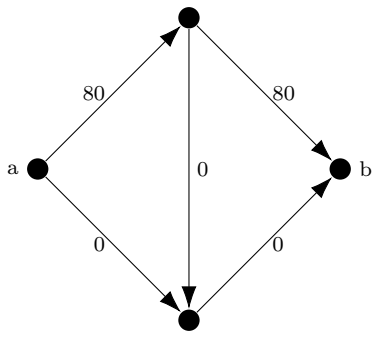
V16

3.2 Der Algorithmus von Ford-Fulkerson

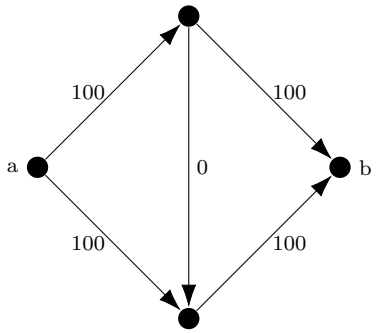
Wir betrachten ein rekursives Verfahren zur Vergrösserung des Wertes von zulässigen Flüssen. Im Folgenden sei $((X,\Gamma),a,b,c)$ ein Netzwerk.

3.2.1 Vorbetrachtungen

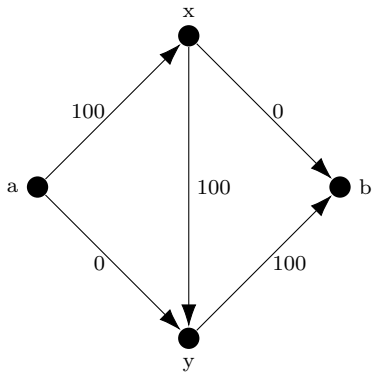
Angenommen, ein Netzwerk der folgenden Form ist gegeben, dessen Kanten alle die Kapazität 100 haben, und auf diesem Netzwerk ein Fluss mit Werten



Einen zulässigen Fluss mit maximalem Wert



erhält man durch *Vergrössern* des Flusses durch geeignete Kanten. Manchmal ist aber auch ein *Verkleinern* des Flusses durch einzelne Kanten nötig: Ist z.B. ein zulässiger Ausgangsfluss der Form



gegeben, so muss der Fluss durch (x, y) auf Null reduziert werden. Im Allgemeinen benötigt man beide Mechanismen, Vergrösserung entlang einer Kante und Verkleinerung entlang einer Kante, um zu einem zulässigen Fluss mit maximalem Wert zu kommen.

3.2.2 Algorithmus von Ford-Fulkerson

Wir formulieren zunächst den Algorithmus und sehen uns dann ein einfaches Beispiel an. Danach machen wir einige theoretische Beobachtungen und betrachten noch ein weiteres (in gewissem Sinne 'weniger einfaches') Beispiel.

(1) Wähle den zulässigen Ausgangsfluss $\varphi \equiv 0$.

(2) Zu einem gegebenen Fluss φ markiere Punkte des Graphen durch Anwenden der folgenden Regeln, solange möglich:

Regel 0: Markiere Quelle a.

Regel 1: Ist $\gamma = (x, y) \in \Gamma$, x markiert, y nicht markiert und $\varphi(\gamma) < c(\gamma)$, so markiere y.

Regel 2: Ist $\gamma = (x, y) \in \Gamma$, $\gamma \neq (b, a)$, x nicht markiert, y markiert und $\varphi(\gamma) > 0$, so markiere x.

(3) Wenn b nicht markiert ist, dann Stopp (haben optimalen Fluss gefunden). Wenn b markiert, dann gibt es nach Konstruktion einen Weg von a nach b längs gewisser

Kanten aus $\Gamma \setminus \{(b,a)\}$ (ohne Beachtung der Richtung) durch markierte Punkte, d.h. durch Punkte

$$a = a_1,\ a_2,\ \ldots,\ a_n = b,$$

sodass für $2 \leq i \leq n$ gilt: Entweder

(α) $(a_{i-1}, a_i) \in \Gamma$ und $\varphi(a_{i-1}, a_i) < c(a_{i-1}, a_i)$ oder
(β) $(a_i, a_{i-1}) \in \Gamma$ und $\varphi(a_i, a_{i-1}) > 0$.

Setze

$$\delta_\alpha := \min\{c(a_{i-1}, a_i) - \varphi(a_{i-1}, a_i) :\ (a_{i-1}, a_i) \in \Gamma\},$$
$$\delta_\beta := \min\{\varphi(a_i, a_{i-1}) :\ (a_i, a_{i-1}) \in \Gamma\},$$

wobei das $\min \emptyset := +\infty$ gesetzt wird, und

$$\delta := \min(\delta_\alpha, \delta_\beta) > 0.$$

Definiere einen neuen Fluss $\widetilde{\varphi}$ durch

$$\widetilde{\varphi}(a_{i-1}, a_i) := \varphi(a_{i-1}, a_i) + \delta \quad \text{falls } (a_{i-1}, a_i) \in \Gamma$$

und

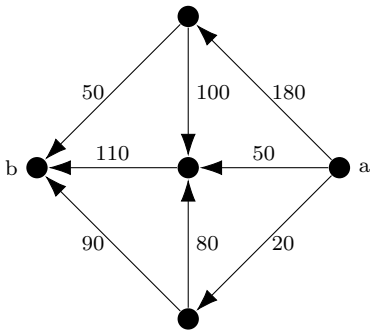
$$\widetilde{\varphi}(a_i, a_{i-1}) := \varphi(a_i, a_{i-1}) - \delta \quad \text{falls } (a_i, a_{i-1}) \in \Gamma$$

für $1 \leq i \leq n$, wobei $a_0 := b$, und

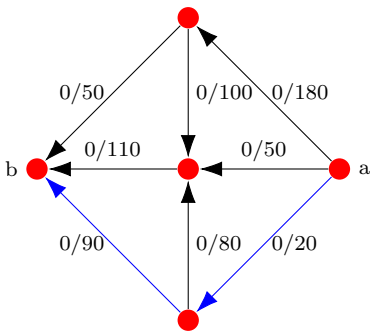
$$\widetilde{\varphi}(\gamma) := \varphi(\gamma) \quad \text{für alle übrigen } \gamma \in \Gamma.$$

(4) Gehe zu (2).

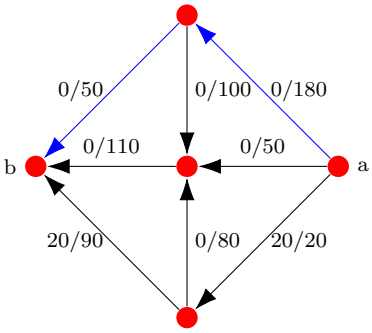
Beispiel 3.2.1. Wir erinnern uns an Beispiel 3.1.1. In abstrakter Form sieht das zugehörige Netzwerk so aus:



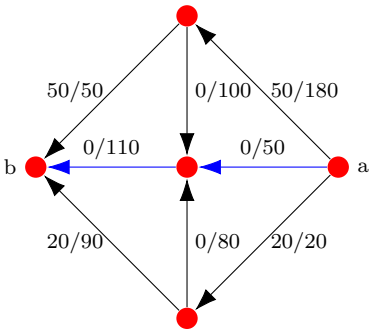
Die Punkte a und b stehen für Frankfurt und New York. Für den Ausgangsfluss $\varphi_0 \equiv 0$ kann man hier alle Knoten (rot) markieren, und wir können einen (blauen) Weg von a nach b wählen:



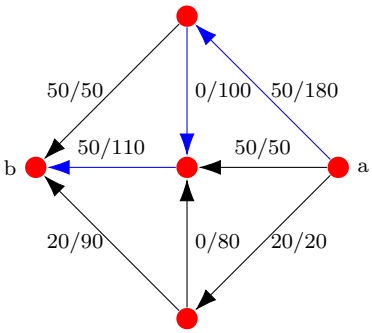
Man hat $\delta = \min(20, 90) = 20$ und erhält damit einen neuen Fluss φ_1 mit



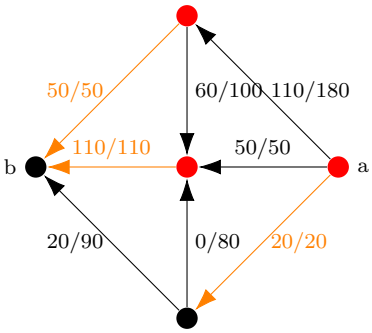
Hier konnten wir wieder alle Knoten (rot) markieren und einen (blauen) Weg wählen. Diesmal ergibt sich $\delta = \min(180, 50) = 50$. Der neue Fluss φ_2 ergibt sich nun als



Wieder können alle Knoten markiert werden, und wir finden einen (blauen) Weg von a nach b. Diesmal folgt $\delta = \min(50, 110) = 50$, und wir bekommen einen neuen Fluss φ_3 mit



Wieder können alle Knoten markiert werden, und wir können immer noch einen (blauen) Weg von a nach b finden. Nun ist $\delta = \min(130, 100, 60) = 60$. Für den neuen Fluss φ_4 können ausser a nur noch zwei weitere Punkte markiert werden:



Der Fluss φ_4 ist also ein Maximalfluss. Sein Wert ist

$$\varphi(b,a) = \sum_{\gamma \in a_\rightarrow} \varphi(\gamma) = \sum_{\gamma \in b_\leftarrow} \varphi(\gamma) = 110 + 50 + 20 = 180.$$

In diesem Beispiel hat sich immer nur der Fall (α), ergeben, wir haben nirgends einen Fluss entlang einer Kante verkleinert. Dieses Beispiel ist also besonders 'einfach' in dem Sinne, dass wir nur einen der beiden Mechanismen benutzt haben. (Wir betrachten in Kürze ein Beispiel, in dem beide Mechanismen benutzt werden.)

Wir kommen zur Formulierung des Algorithmus zurück und benutzen dieselbe Notation wie dort. Wir weisen nun auch formal nach, dass mit jedem Update vom ursprünglichen Fluss φ zum neuen Fluss $\widetilde{\varphi}$ Flusseigenschaft und Zulässigkeit erhalten bleiben und der Wert sich erhöht. Sei dazu

$$a =: a_1,\ a_2,\ ...,\ a_n := b$$

ein Weg von a nach b durch markierte Punkte wie in Schritt 3 des Algorithmus, und sei $a_0 := b$, wie ebenfalls dort vereinbart.

Proposition 3.2.2. *Für $\widetilde{\varphi}$ gilt:*

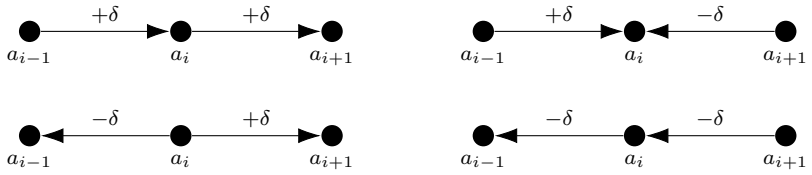
(i) $\widetilde{\varphi}$ ist ein Fluss (erfüllt Kirchhoffsche Regel).

(ii) $\widetilde{\varphi}$ ist zulässig.

(iii) $\widetilde{\varphi}$ hat einen grösseren Wert als φ, d.h.

$$\widetilde{\varphi}(b,a) > \varphi(b,a).$$

Beweis. Wir zeigen (i): Für $x \in X \setminus \{a_1, ..., a_n\}$ ist $\widetilde{\varphi} = \varphi$ auf $x_\leftarrow$ and $x_\rightarrow$, in solchen Knoten bleibt also die Kirchhoff-Bedingung stets erfüllt. Für $x = a_i$, $1 \leq i \leq n$, ergibt sich stets einer der folgenden Fälle:



In allen Fällen bleibt bei dem Update die Kirchhoff-Bedingung in $x = a_i$ erfüllt, im ersten Fall hat man

$$\sum_{\gamma \in x_\rightarrow} \widetilde{\varphi}(\gamma) = \sum_{\gamma \in x_\rightarrow} \varphi(\gamma) + \delta = \sum_{\gamma \in x_\leftarrow} \varphi(\gamma) + \delta = \sum_{\gamma \in x_\leftarrow} \widetilde{\varphi}(\gamma),$$

die anderen Fälle funktionieren analog. Aussage (ii) ist klar nach der Definition von δ. Aussage (iii) folgt wegen

$$\widetilde{\varphi}(b,a) = \widetilde{\varphi}(a_0,a_1) = \varphi(a_0,a_1) + \delta = \varphi(b,a) + \delta > \varphi(b,a).$$

□

Satz 3.2.3. *Der Ford-Fulkerson-Algorithmus bricht nach endlich vielen Schritten ab und liefert dabei einen zulässigen Fluss mit maximalem Wert.*

Beweis. Da der Wert des Flusses stets ganzzahlig ist, wird er wegen Proposition 3.2.2 um mindestens 1 grösser. Nach Korollar 3.1.8 gilt

$$\varphi(b,a) \leq \sum_{\gamma \in a_{\rightarrow}} c(\gamma) < +\infty,$$

also ist der maximal mögliche Wert beschränkt, und somit muss der Algorithmus zwangsläufig nach endlich vielen Schritten abbrechen. Für den am Ende erhaltenen Fluss φ wird in Schritt 2 der Punkt b nicht mehr markiert. Sei A die Menge aller markierten Punkte. Dann ist $a \in A$, $b \notin A$, ferner $\varphi(\gamma) = c(\gamma)$ für alle $\gamma \in A_{\rightarrow}$ und $\varphi(\gamma) = 0$ für alle $\gamma \in A_{\leftarrow} \setminus \{(b,a)\}$. Also hat man wegen Lemma 3.1.7

$$\varphi(b,a) = \sum_{\gamma \in A_{\leftarrow}} \varphi(\gamma) = \sum_{\gamma \in A_{\rightarrow}} \varphi(\gamma) = \sum_{\gamma \in A_{\rightarrow}} c(\gamma).$$

Nach Korollar 3.1.8 ist aber $\sum_{\gamma \in A_{\rightarrow}} c(\gamma)$ eine obere Schranke für den Wert eines Flusses, somit besitzt φ den maximalen Wert. □

Definition 3.2.4. Sei $((X,\Gamma),a,b,c)$ ein Netzwerk und $\Gamma' := \Gamma \setminus \{(b,a)\}$. Wir nennen eine endliche Folge von Knoten

$$a = x_0,\ x_1,\ ...,\ x_m = b$$

einen *Weg* von a nach b, falls für jedes $i = 1, ..., m$ entweder $(x_{i-1}, x_i) \in \Gamma'$ oder $(x_i, x_{i-1}) \in \Gamma'$ gilt.

Definition 3.2.5. Sei φ ein zulässiger Fluss auf einen Netzwerk $((X,\Gamma),a,b,c)$ und $\Gamma' := \Gamma \setminus \{(b,a)\}$. Ein Weg

$$a = x_0,\ x_1,\ ...,\ x_m = b$$

heisst *vergrössernder Weg* bezüglich φ, falls für jedes $i = 1, .., m$ gilt: Entweder

(α) $(x_{i-1}, x_i) \in \Gamma'$ und $\varphi(x_{i-1}, x_i) < c(x_{i-1}, x_i)$ oder

(β) $(x_i, x_{i-1}) \in \Gamma'$ und $\varphi(x_i, x_{i-1}) > 0$.

Im Fall (α) nennt man (x_{i-1}, x_i) eine *Vorwärtskante*, im Fall (β) nennt man (x_i, x_{i-1}) eine *Rückwärtskante*.

Korollar 3.2.6. *Ein zulässiger Fluss φ auf einem Netzwerk ist genau dann ein Maximalfluss, wenn es keinen vergrössernden Weg bzgl. φ gibt.*

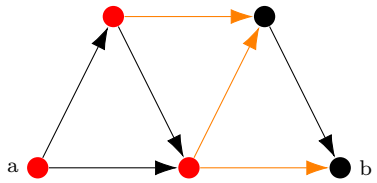
Beweis. Gibt es einen vergrössernden Weg, so gibt es nach Schritt 3 im Algorithmus von Ford-Fulkerson und Proposition 3.2.2 einen Fluss mit grösserem Wert. Gibt es keinen vergrössernden Weg, so kann man die Senke b nicht markieren, und daher folgt wie im Beweis von Satz 3.2.3, dass φ bereits den maximalen Wert hat. □

Definition 3.2.7. Sei $((X,\Gamma),a,b,c)$ ein Netzwerk und $A \subset X$ eine Menge von Knoten mit $a \in A$ und $b \notin A$. Dann heisst

$$\sum_{\gamma \in A_{\rightarrow}} c(\gamma)$$

die *Schnittkapazität* von A. Als *Schnitt* bezeichnet man die Zerlegung von X in die disjunkten Knotenmengen A und A^c.

Die Knotenmenge X wird also disjunkt zerlegt in A und A^c, wobei $a \in A$ ist und $b \in A^c$. Die Schnittkapazität ist die Summe der Kapazitäten (also die 'Gesamtkapazität') aller Kanten $\gamma \in A_{\rightarrow}$.



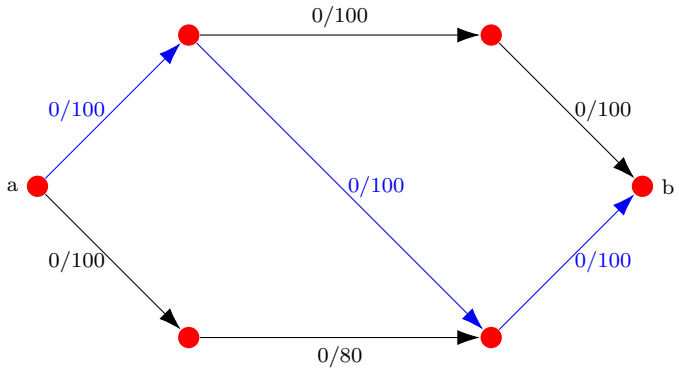
Erlaubt man A zu variieren, erhält man folgende Aussage, bekannt als das *Max flow-min cut-Theorem*.

Korollar 3.2.8. *Der Wert eines Maximalflusses in einem Netzwerk ist gleich der minimalen Schnittkapazität.*

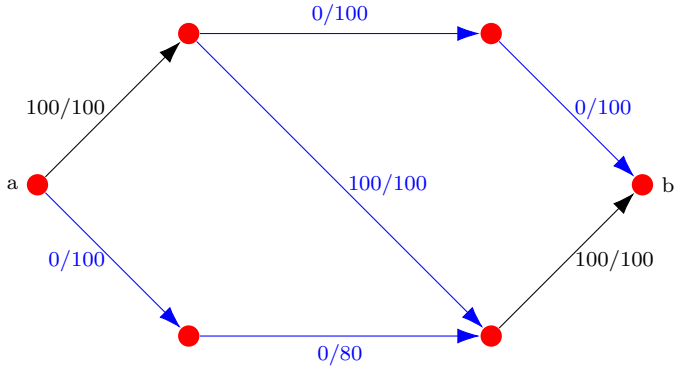
Beweis. Nach Korollar 3.1.8 ist der Wert eines beliebigen zulässigen Flusses stets von oben beschränkt durch jede mögliche Schnittkapazität, also insbesondere durch die minimal mögliche. Ist nun φ ein Maximalfluss, so gibt es nach Definition keinen vergrössernden Weg bzgl. φ. Sei nun A die Menge aller markierten Punkte (bzgl. φ) wie in Schritt 2 des Algorithmus von Ford-Fulkerson. Dann ergibt sich wie im Beweis von Satz 3.2.3, dass $\varphi(b, a)$ gleich der Schnittkapazität $\sum_{\gamma \in A_{\rightarrow}} c(\gamma)$ von A ist. □

Bemerkung 3.2.9. Der Algorithmus von Ford-Fulkerson liefert damit also auch ein konstruktives Verfahren zur Bestimmung des Schnittes mit minimaler Kapazität.

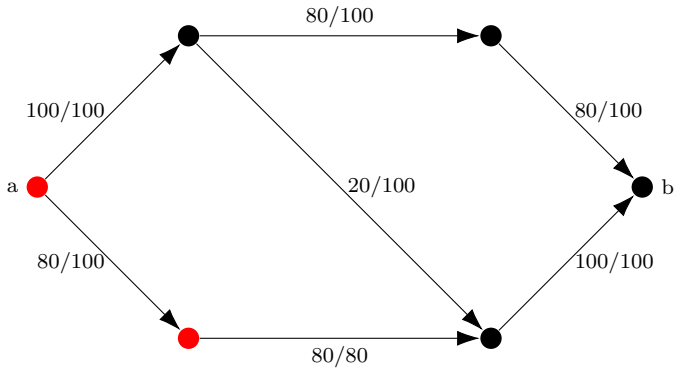
Beispiel 3.2.10. Wir betrachten als weiteres Beispiel folgendes Netzwerk mit dem Ausgangsfluss $\varphi_0 \equiv 0$:



Alle Knoten können markiert werden, und wir finden einen (blauen) Weg von a nach b entlang markierter Kanten. Für diesen Weg ist $\delta = 100$, das ergibt einen neuen Fluss φ_1 mit



Wieder können alle Knoten markiert werden, und wir können den eingezeichneten (blauen) Weg wählen, nun mit einer Rückwärtskante, für diese tritt Fall (β) ein. Diesmal ist $\delta = 80$, und wir bekommen einen neuen Fluss φ_2 mit



Der Knoten b kann nicht mehr markiert werden, also ist φ_2 ein Maximalfluss. Sein Wert ist 180.

In diesem Beispiel haben wir nicht nur entlang geeigneter Kanten vergrössert, sondern den Fluss auf der Rückwärtskante auch verkleinert, um seinen Wert zu vergrössern. Dieses Beispiel nutzt also beide Mechanismen.

3.2.3 Zyklenzerlegung für Flüsse

Wir schauen uns die mögliche Struktur von Netzwerken und Flüssen noch genauer an. Das erlaubt unter anderem eine alternative Formulierung des Max flow-min cut-Theorems.

Definition 3.2.11.

(i) Ein Graph $G_1 = (X_1, \Gamma_1)$ heisst ein *Teilgraph* eines Graphen $G = (X, \Gamma)$, falls

$$X_1 \subset X \quad \text{und} \quad \Gamma_1 \subset \Gamma.$$

(ii) Sei $G = (X, \Gamma)$ ein Graph. Ein Teilgraph $K = (X', \Gamma')$ von G heisst *Zyklus* (oder *Kreis*), falls es Punkte $a_1, ..., a_k \in X$ gibt mit

$$X' = \{a_1, ..., a_k\}$$

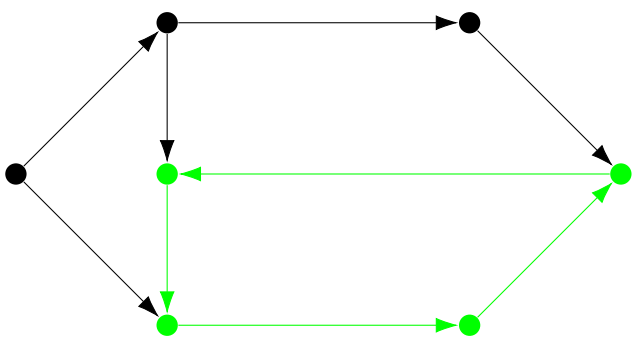
und

$$\Gamma' = \{(a_1, a_2), (a_2, a_3), ..., (a_{k-1}, a_k), (a_k, a_1)\}.$$

(iii) Ist $N = ((X, \Gamma), a, b, c)$ ein Netzwerk und φ ein zulässiger Fluss auf N, so heisst φ *Fluss entlang eines Zyklus* $K = (X', \Gamma')$, falls $K = (X', \Gamma')$ ein Zyklus ist und

$$\varphi(\gamma) = 0 \quad \text{für alle } \gamma \in \Gamma \setminus \Gamma'.$$

Hier ist ein Beispiel für einen Zyklus $K = (X', \Gamma')$:



Aus der Kirchhoff-Regel folgt, dass ein Fluss φ entlang eines Zyklus $K = (X', \Gamma')$ konstant sein muss auf dem Zyklus, d.h. es gibt eine Zahl $c_{\Gamma'} \in \mathbb{N}$ mit

$$\varphi(\gamma) = c_{\Gamma'}, \quad \gamma \in \Gamma'.$$

Man kann nun folgende strukturelle Beobachtung machen:

Satz 3.2.12. *Für jeden zulässigen Fluss in einem Netzwerk N existieren zulässige Flüsse $\varphi_1, ..., \varphi_m$ längs Zyklen, sodass gilt*

$$\varphi = \varphi_1 + ... + \varphi_m.$$

Diesen Satz nennt man auch *Zyklenzerlegung für Flüsse.*

Beweis. Sei $N = ((X, \Gamma), a, b, c)$. Wir nutzen vollständige Induktion über die Anzahl der Kanten $\gamma \in \Gamma$, für welche $\varphi(\gamma) > 0$ gilt. Da für $n < 3$ kein zulässiger Fluss $\varphi \neq 0$ auf einem gerichteten Graphen möglich ist (Frage: Warum ?), betrachten wir als Induktionsanfang den Fall $n = 3$, und für diesen ist die Aussage klar, denn nach den Kirchhoff-Regel muss dann φ selbst ein Fluss entlang eines einzelnen Zyklus sein. V17

Für den Induktionsschritt nehmen wir an, die Behauptung sei richtig für alle $i \leq n$. Sei nun φ ein zulässiger Fluss auf N mit $\varphi(\gamma) > 0$ auf $n + 1$ Kanten γ.

Sei

$$\gamma_1 =: (a_1, a_2) \in \Gamma$$

eine Kante mit $\varphi(\gamma_1) > 0$. Nach der Kirchhoffschen Regel gibt es $\gamma_2 \in (a_2)_{\rightarrow}$ mit

$$\gamma_2 =: (a_2, a_3) \in \Gamma \quad \text{mit } \varphi(\gamma_2) > 0.$$

Wieder nach der Kirchhoffschen Regel gibt es

$$\gamma_3 =: (a_3, a_4) \in \Gamma \quad \text{mit } \varphi(\gamma_3) > 0,$$

und rekursiv erhalten wir daraus eine Folge von Punkten

$$a_1, a_2, a_3, a_4, ...$$

Da X endlich ist, müssen sich manche dieser Punkte wiederholen. Wie beim Induktionsanfang muss der 'Abstand' zwischen sich wiederholenden Punkten mindestens drei sein, genauer: es gibt $k > 3$ minimal mit

$$a_k \in \{a_1, a_2, ..., a_{k-3}\}.$$

Sei nun $l \leq k - 3$ so gewählt, dass $a_l = a_k$. Setze

$$X_1 := \{a_l, a_{l+1}, ..., a_{k-1}\}$$

und

$$\Gamma_1 := \{(a_l, a_{l+1}), (a_{l+1}, a_{l+2}), ..., (a_{k-1}, \underbrace{a_k}_{=a_l})\}.$$

Dann ist $K_1 = (X_1, \Gamma_1)$ ein Zyklus in (X, Γ), und nach Konstruktion ist $\varphi(\gamma) > 0$ für alle $\gamma \in \Gamma_1$.

Setze $\alpha_1 := \min_{\gamma \in \Gamma_1} \varphi(\gamma) > 0$ und

$$\varphi_1(\gamma) := \begin{cases} \alpha_1 & \text{falls } \gamma \in \Gamma_1 \\ 0 & \text{falls } \gamma \in \Gamma \setminus \Gamma_1. \end{cases}$$

Dann ist φ_1 ein zulässiger Fluss entlang K_1 und

$$\varphi' := \varphi - \varphi_1$$

ist ein zulässiger Fluss in N. Da ein $\widetilde{\gamma} \in \Gamma_1$ existiert mit

$$\varphi(\widetilde{\gamma}) = \min_{\gamma \in \Gamma_1} \varphi(\gamma) = \alpha_1,$$

ist $\varphi'(\widetilde{\gamma}) = 0$. Also ist die Anzahl der Kanten $\gamma \in \Gamma$ mit $\varphi'(\gamma) > 0$ kleiner gleich n.

Nach Induktionsvoraussetzung existieren also zulässige Flüsse $\varphi_2, ..., \varphi_m$ entlang Zyklen in N mit

$$\varphi' = \varphi_2 + ... + \varphi_m,$$

und somit folgt

$$\varphi = \varphi_1 + \varphi_2 ... + \varphi_m.$$

□

Bemerkung 3.2.13.

(i) Es folgt insbesondere: Gibt es in einem Netzwerk einen zulässigen Fluss $\varphi \neq 0$, so gibt es einen Zyklus.

(ii) Der Induktionsschritt liefert ein konstruktives Verfahren zur Zerlegung eines Flusses in Flüsse entlang Zyklen.

(iii) Ist $\varphi = \varphi_1 + ... + \varphi_m$, so folgt

$$\varphi(b, a) = \varphi_1(b, a) + ... + \varphi_m(b, a),$$

d.h. der Wert von φ ist die Summe der Werte der Flüsse φ_i entlang derjenigen Zyklen, die die Kontrollkante (b, a) enthalten.

Insbesondere kann man folgende Aussage ber Maximalflüsse festhalten:

Korollar 3.2.14. *Sei $N = (G, a, b, c)$ ein Netzwerk. Dann existieren Zyklen $K_i = (X_i, \Gamma_i)$ in G mit $(b, a) \in \Gamma_i$ und zulässige Flüsse φ_i entlang K_i, $1 \leq i \leq n$, sodass*

$$\varphi = \varphi_1 + ... + \varphi_n$$

ein Maximalfluss ist.

Beweis. Nach Satz 3.2.3 gibt es einen Maximalfluss φ_0, und o.E. dürfen wir annehmen, dass $\varphi_0 \neq 0$. Dann gilt nach Satz 3.2.12, dass

$$\varphi_0 = \varphi_1 + ... + \varphi_m$$

mit zulässigen Flüssen φ_i entlang Zyklen K_i, $1 \leq i \leq m$. Seien nun $\varphi_1, ..., \varphi_n$, $n \leq m$, die Flüsse entlang derjenigen Zyklen, die (b, a) enthalten. Dann ist auch

$$\varphi := \varphi_1 + ... + \varphi_n \leq \varphi_0$$

ein zulässiger Fluss, und sein Wert ist

$$\begin{aligned} \varphi(b, a) &= \varphi_1(b, a) + ... + \varphi_n(b, a) \\ &= \varphi_1(b, a) + ... + \varphi_n(b, a) + \underbrace{\varphi_{n+1}(b, a) + ... + \varphi_m(b, a)}_{=0} = \varphi_0(b, a). \end{aligned}$$

Das bedeutet aber, φ ist ein Maximalfluss. □

Definition 3.2.15. Sei $((X,\Gamma),a,b,c)$ ein Netzwerk. Wir nennen einen Weg

$$a = x_0,\ x_1,\ ...,\ x_m = b$$

von a nach b *gerichtet*, falls für alle $i = 1,...,m$ gilt, dass $(x_{i-1},x_i) \in \Gamma' = \Gamma \setminus \{(b,a)\}$ (falls er also mit den Richtungen aller darin vorkommenden Kanten kompatibel ist).

Man kann nun das Max flow-min cut-Theorem (Korollar 3.2.8) wie folgt reformulieren.

Satz 3.2.16. *Sei $N = ((X,\Gamma),a,b,c)$ ein Netzwerk und $\mathcal{T}$ die Familie aller Teilmengen $T \subset \Gamma$ derart, dass jeder gerichtete Weg von a nach b mindestens eine Kante von T enthält. Dann gilt für den Wert α eines Maximalflusses die Gleichheit*

$$\alpha = \min_{T\in\mathcal{T}} \sum_{\gamma\in T} c(\gamma).$$

Beweis. Ist $A \subset X$ mit $a \in A$ und $b \notin A$, dann enthält jeder gerichtete Weg von a nach b eine Kante in $A_\rightarrow$. Also ist $A_\rightarrow \in \mathcal{T}$ und daher

$$\sum_{\gamma\in A_\rightarrow} c(\gamma) \geq \min_{T\in\mathcal{T}} \sum_{\gamma\in T} c(\gamma).$$

Weil das aber für alle $A \subset X$ mit $a \in A$ und $b \notin A$ gilt und nach Korollar 3.2.8

$$\alpha = \min\{ \sum_{\gamma\in A_\rightarrow} c(\gamma) :\ A \subset X, a \in A, b \notin A\}$$

gilt (α ist maximale Schnittkapazität), so erhalten wir

$$\alpha \geq \min_{T\in\mathcal{T}} \sum_{\gamma\in T} c(\gamma).$$

Umgekehrt gibt es nach Korollar 3.2.14 zulässige Flüsse φ_i entlang Zyklen $K_i = (X_i,\Gamma_i)$ mit $(b,a) \in \Gamma_i$, $1 \leq i \leq m$, sodass $\varphi = \varphi_1 + ... + \varphi_m$ ein Maximalfluss ist, also $\varphi(b,a) = \alpha$. Sei nun $T =: \{\gamma_1,...,\gamma_n\}$ ein Element von $\mathcal{T}$. Da wegen $(b,a) \in \Gamma_i$ jeder Zyklus K_i auch einen gerichteten (denn es gibt einen zulässigen Fluss darauf) Weg von a nach b enthalten muss, enthält er also auch eine Kante aus T , d.h. zu jedem $i \in \{1,...,m\}$ existiert ein $\gamma_{j_i} \in T$ mit $\gamma_{j_i} \in \Gamma_i$. Da Flüsse entlang Zyklen konstant auf deren Kanten sind, ist

$$\begin{aligned}\alpha = \varphi(b,a) = \sum_{i=1}^{m} \varphi_i(b,a) &= \sum_{i=1}^{m} \varphi_i(\gamma_{j_i}) \\ &\leq \sum_{i=1}^{m}\sum_{j=1}^{n} \varphi_i(\gamma_j) = \sum_{j=1}^{n}\sum_{i=1}^{m} \varphi_i(\gamma_j) = \sum_{j=1}^{n} \varphi(\gamma_j) \\ &\leq \sum_{j=1}^{n} c(\gamma_j) = \sum_{\gamma\in T} c(\gamma).\end{aligned}$$

Da $T \in \mathcal{T}$ beliebig war, folgt

$$\alpha \leq \min_{T\in\mathcal{T}} \sum_{\gamma\in T} c(\gamma).$$

□

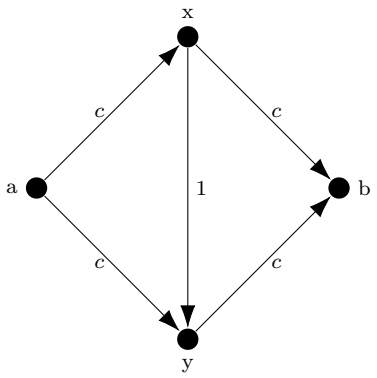
3.3 *Algorithmus von Edmond-Karp

Unser nächstes Ziel ist es nun, eine Abschätzung für die Anzahl der benötigten Schritte im Algorithmus von Ford-Fulkerson zu finden. Im Zuge dessen legen wir (ähnlich wie früher beim Simplexalgorithmus) weitere Regeln fest, die eine günstige Wahl von Wegen garantieren.

Sei $((X, \Gamma), a, b, c)$ ein Netzwerk und $\Gamma' := \Gamma \setminus \{(b, a)\}$.

Bemerkung 3.3.1. Der Algorithmus von Ford-Fulkerson vergrössert den Wert eines Flusses durch Abändern entlang eines vergrössernden Weges $a = x_0, x_1, ..., x_m = b$ von a nach b, und zwar durch Vergrösserung auf Vorwärts- und Reduktion auf Rückwärtskanten.

(i) Wählt man die vergrössernden Wege ungünstig, so ist die Anzahl der notwendigen Flussvergrösserungen nicht nur von dem Graphen (X, Γ), sondern auch von den Kapazitäten $c(\gamma)$ abhängig. Zum Beispiel kann für das Netzwerk



mit $c \geq 1$ der Maximalfluss mit Wert $2c$ mit 2 Flussvergrösserungen erreicht werden. Es ist aber auch möglich, als vergrössernde Wege abwechselnd

$$a,\ x,\ y,\ b \qquad \text{und} \qquad a,\ y,\ x,\ b$$

zu wählen, in diesem Falle wächst der Wert in jedem Schritt um 1, und man braucht $2c$ Schritte, um den Maximalfluss zu erhalten.

(ii) Diese Kapazitätsabhängigkeit wird vermieden, wenn man stets vergrössernde Wege kürzester Länge wählt. Die kürzeste Weglänge wird dabei erreicht, wenn man im Markierungsverfahren (Schritt 2 im Algorithmus)

- im ersten Teilschritt alle Punkte markiert, die von a aus markiert werden können und
- im jeweils $(k+1)$-ten Teilschritt *alle* Punkte markiert, die ausgehend von *allen* im k-ten Teilschritt markierten den Regeln nach erreicht werden können ('Breitensuche').

Nach dem ersten Erreichen von b (falls überhaupt erreichbar) liefert Rückverfolgung über einzelne 'Generationen' (aka Teilschritte) einen kürzesten vergrössernden Weg.

Implementiert man diese upgrades, erhält man den *Algorithmus von Edmond-Karp*.

Wir nehmen nun an, dass immer vergrössernde Wege kürzester Länge gewählt werden. (Das kann man anhand der obigen Instruktionen einfach implementieren, mathematisch passiert da nicht viel.) Für diese Situation berechnen wir mit 'worst-case'-Abschätzungen eine Laufzeitschranke für den Algorithmus, und das mathematische tool, welches wir dafür massgeblich nutzen, ist die Länge von vergrössernden Wegen.

Sei im folgenden φ_0 ein zulässiger Fluss (z.B. $\varphi_0 = 0$) und sei $\varphi_1, ..., \varphi_n$ eine endliche Folge von zulässigen Flüssen, bei denen φ_{k+1} aus φ_k durch Abänderung (nach Schritt 3 im Ford-Fulkerson-Algorithmus) längs eines vergrössernden Weges bzgl. φ_k kürzester Länge gewonnen ist.

Definition 3.3.2. Sei $0 \leq k \leq n$, $y, z \in X$, $y \neq z$. Ein *vergrössernder Weg von y nach z bzgl. φ_k* ist eine endliche Folge von Knoten

$$y = x_0,\ x_1,\ ...,\ x_l = z,$$

sodass für $1 \leq i \leq l$ entweder $(x_{i-1}, x_i) \in \Gamma'$ und $\varphi_k(x_{i-1}, x_i) < c(x_{i-1}, x_i)$ oder $(x_i, x_{i-1}) \in \Gamma'$ und $\varphi_k(x_i, x_{i-1}) > 0$. Die Zahl l nennen wir die *Länge* des Weges.

Wir schreiben $l_k(y, z)$ für die kürzeste Länge eines vergrössernden Weges von y nach z bzgl. φ_k, mit der Vereinbarung, dass $l_k(y, z) := +\infty$, falls kein solcher vergrössernder Weg existiert, und wir setzen $l_k(x, x) := 0$ für alle $x \in X$.

Durch 'Zwischenstopps' können keine Teilwege entstehen, die kürzer sind als die entsprechenden Teile eines kürzesten vergrössernden Weges:

Lemma 3.3.3. *Sei $0 \leq k \leq n$, $m := l_k(a, b)$ und sei*

$$a = x_0,\ x_1,\ ...,\ x_m = b$$

ein kürzester vergrössernder Weg von a nach b bzgl. φ_k. Dann gilt für jedes $x = x_j$, $0 \leq j \leq m$, dass

$$l_k(a, x) = j \qquad \textit{und} \qquad l_k(x, b) = m - j.$$

Beweis. Klar ist, dass $m_1 := l_k(a, x) \leq j$ und $m_2 := l_k(x, b) \leq m - j$ sein müssen, also auch $m_1 + m_2 \leq m$.

Seien nun $a = y_0,\ y_1,\ ...,\ y_{m_1} = x$ und $x = z_0,\ z_1,\ ...,\ z_{m_2} = b$ vergrössernde Wege kürzester Länge bzgl. φ_k. Dann ist $a = y_0,\ y_1,\ ...,\ y_{m_1} = x = z_0,\ z_1,\ ...,\ z_{m_2} = b$ ein vergrössernder Weg von a nach b. Setzen

$$\mu := \min\{1 \leq i \leq m :\ y_i \in \{z_0, ..., z_{m_2}\}\}.$$

Die Menge auf der rechten Seite ist immer nichtleer, denn $y_{m_1} = z_0$. Sei nun $\nu \in \{0, ..., m_2\}$ so gewählt, dass $y_\mu = z_\nu$. Dann ist

$$a = y_0,\ ...,\ y_\mu = z_\nu, z_{\nu+1},\ ...,\ z_{m_2} = b$$

ein vergrössernder Weg bzgl. φ_k von a nach b. Für seine Länge L gilt natürlich $L \geq l_k(a, b) = m$. Andererseits haben wir wegen $m_1 \leq j$ und $m_2 \leq m - j$ auch

$$L \leq m_1 + m_2 \leq j + (m - j) = m,$$

also $m_1 + m_2 = m$, und das geht dann aber nur mit $m_1 = j$ und $m_2 = m - j$. □

Korollar 3.3.4. *Sei $0 \leq k \leq n$ und $\gamma = (x, y) \in \Gamma'$.*

(i) Ist γ beim Übergang von φ_k zu φ_{k+1} eine Vorwärtskante, so gilt

$$l_k(a,x)+1=l_k(a,y).$$

(ii) Ist γ beim Übergang von φ_k zu φ_{k+1} eine Rückwärtskante, so gilt

$$l_k(a,y)+1=l_k(a,x).$$

Beweis. Sei $a=x_0,\ ...,\ x_m=b$ der beim Übergang von φ_k zu φ_{k+1} verwendete vergrössernde Weg von a nach b bzgl. φ_k. Ist γ eine Vorwärtskante, so gilt $x=x_j$ und $y=x_{j+1}$ für ein geeignetes j und somit nach Lemma 3.3.3 $l_k(a,y)=j+1=l_k(a,x)+1$. Ist γ eine Rückwärtskante, so gilt $x=x_{j+1}$ und $y=x_j$ für ein geeignetes j und somit $l_k(a,x)=j+1=l_k(a,y)+1$. □

Ein upgrade von φ_k zu φ_{k+1} kann Teilwege höchstens verlängern:

Lemma 3.3.5. *Sei $x\in X$ und $0\le k<n$. Dann gilt*

$$l_{k+1}(a,x)\ge l_k(a,x)\quad \textit{und}\quad l_{k+1}(x,b)\ge l_k(x,b).$$

Beweis. Wir beweisen die erste Ungleichung, die zweite folgt analog.

Sei $m:=l_{k+1}(a,x)$ und

$$a=x_0,\ x_1,\ ...,\ x_m=x$$

ein vergrössernder Weg kürzester Länge bzgl. φ_{k+1}.

Wir behaupten: Für $1\le i\le m$ gilt

$$l_k(a,x_i)\le l_k(a,x_{i-1})+1.$$

Das impliziert dann

$$l_k(a,x)=l_k(a,x_m)\le l_k(a,x_{m-1})+1\le ...\le \underbrace{l_k(a,x_0)}_{=l_k(a,a)=0}+m=m=l_{k+1}(a,x).$$

Um die Behauptung zu zeigen, sei nun $1\le i\le m$. Ist γ Vorwärtskante beim Übergang von φ_{k+1} zu φ_{k+2}, d.h. $\gamma=(x_{i-1},x_i)\in\Gamma'$ und $\varphi_{k+1}(\gamma)<c(\gamma)$, dann gibt es zwei Möglichkeiten für φ_k:

(1a) Es könnte sein, dass $\varphi_k(\gamma)<c(\gamma)$. In diesem Falle sei

$$a=x'_0,\ ...,\ x'_l=x_{i-1}$$

ein vergrössernder Weg kürzester Länge bzgl. φ_k von a nach x_{i-1}. Dann ist

$$a=x'_0,\ ...,\ x'_l=x_{i-1},\ x_i$$

ein vergrössernder Weg kürzester Länge bzgl. φ_k von a nach x_i, und somit wegen Korollar 3.3.4

$$l_k(a,x_i)\le l+1=l_k(a,x_{i-1})+1.$$

(1b) Andernfalls muss $\varphi_k(\gamma)=c(\gamma)$ sein. Dann muss γ eine Rückwärtskante beim Übergang von φ_k nach φ_{k+1} gewesen sein, und nach Korollar 3.3.4 $l_k(a,x_i)+1=l_k(a,x_{i-1})$, also

$$l_k(a,x_i)=l_k(a,x_{i-1})-1\le l_k(a,x_{i-1})+1.$$

Ist γ Rückwärtskante beim Übergang von φ_{k+1} zu φ_{k+2}, d.h. $\gamma = (x_i, x_{i-1}) \in \Gamma'$ und $\varphi_{k+1}(\gamma) > 0$, dann hat man wieder zwei Möglichkeiten:

(2a) Es könnte sein, dass $\varphi_k(\gamma) > 0$. In diesem Falle kann man wie in (1a) argumentieren.

(2b) Andernfalls muss $\varphi_k(\gamma) = 0$ sein. Dann muss γ beim Übergang von φ_k to φ_{k+1} eine Vorwärtskante gewesen sein, und es folgt, dass $l_k(a, x_i) + 1 = l_k(a, x_{i-1})$, also insbesondere $l_k(a, x_i) \leq l_k(a, x_{i-1}) + 1$.

□

Eine spezifische Kante kann in einem vergrössernden Weg eine derjenigen Kanten sein, die festlegen, wie sehr der Wert des Flusses erhöht werden kann. Das ist dann der Fall, wenn im betreffenden Vergrösserungsschritt der Fluss auf dieser Kante 'voll aufgedreht' oder 'ganz ausgetrocknet' wird. Diesen Umstand formalisiert man mit dem Begriff des 'Flaschenhalses' (bottleneck), die Intuition dabei ist, dass so eine Kante dann 'eine besonders dünne Stelle' ist, die die mögliche Flussvergrösserung limitiert.

Definition 3.3.6. Eine Kante $\gamma \in \Gamma'$ heisst *Flaschenhals* beim Übergang von φ_k zu φ_{k+1}, falls

$$\varphi_k(\gamma) < c(\gamma) \quad \text{und} \quad \varphi_{k+1}(\gamma) = c(\gamma)$$

oder

$$\varphi_k(\gamma) > 0 \quad \text{und} \quad \varphi_{k+1}(\gamma) = 0.$$

Diese Definition ist ein gutes tool, denn sie erlaubt ein Abzählargument. Wir klären dazu, wie sich vergrössernde Wege mindestens verlängern müssen, wenn eine Kante in verschiedenen Vergrösserungsschritten als Flaschenhals auftacht.

Satz 3.3.7. *Sei $0 \leq k < m < n$ und $\gamma \in \Gamma'$ ein Flaschenhals, sowohl für den Übergang von φ_k nach φ_{k+1}, als auch für den von φ_m nach φ_{m+1}. Dann gilt*

$$l_m(a, b) \geq l_k(a, b) + 2.$$

Beweis. Sei $\gamma = (x, y)$.

Der erste Fall, der eintreten kann, ist, dass γ beim Übergang von φ_k nach φ_{k+1} eine Vorwärtskante ist, also $\varphi_k(\gamma) < c(\gamma)$ und $\varphi_{k+1}(\gamma) = c(\gamma)$. Da γ ein Flaschenhals beim Übergang von φ_m zu φ_{m+1} ist, kann es nicht passieren, dass

$$\varphi_p(\gamma) = c(\gamma) \quad \text{für alle } k + 2 \leq p \leq m + 1.$$

Sei also p gegeben mit $k + 1 \leq p \leq m$ und

$$\varphi_p(\gamma) = c(\gamma) \quad \text{und} \quad \varphi_{p+1}(\gamma) < c(\gamma),$$

γ ist dann Rückwärtskante beim Übergang von φ_p zu φ_{p+1}. Nach Korollar 3.3.4 dann also

$$l_k(a, x) + 1 = l_k(a, y) \quad \text{und} \quad l_p(a, y) + 1 = l_p(a, x).$$

Mit Lemma 3.3.5 und Lemma 3.3.3 folgt dann

$$\begin{aligned} l_m(a, b) \geq l_p(a, b) &= l_p(a, x) + l_p(x, b) = l_p(a, y) + 1 + l_p(x, p) \\ &\geq l_k(a, y) + 1 + l_k(x, b) = l_k(a, x) + 2 + l_k(x, b) = l_k(a, b) + 2. \end{aligned}$$

Der zweite Fall, der eintreten kann, ist, dass γ beim Übergang von φ_k nach φ_{k+1} eine Rückwärtskante ist, also $\varphi_k(\gamma) > 0$ und $\varphi_{k+1}(\gamma) = 0$. Dann kann man wie oben schlussfolgern, dass es ein $k+1 \leq p \leq m$ gibt mit

$$\varphi_p(\gamma) = 0 \quad \text{und} \quad \varphi_{p+1}(\gamma) > 0,$$

d.h. γ ist Vorwärtskante beim Übergang von φ_p zu φ_{p+1}. Die gleiche Rechnung wie im ersten Fall, nur mit x und y vertauscht, liefert wieder

$$l_m(a,b) \leq l_k(a,b) + 2.$$

□

Wir erhalten den folgenden *Satz von Edmond-Karp.*

Satz 3.3.8. *Führt man den Algorithmus von Ford-Fulkerson in einem Netzwerk* $N = ((X,\Gamma), a, b, c)$ *mit kürzesten vergrössernden Wegen durch, so wird nach höchstens*

$$\frac{1}{2} \cdot \#X \cdot \#\Gamma'$$

Flussvergrösserungen ein Maximalfluss erreicht.

Dieser Satz liefert also eine Laufzeitschranke, die nur von der Anzahl der Knoten und der Anzahl der Kanten in $G = (X,\Gamma)$ abhängt.

Beweis. Sei $l_k(a,b)$ wie oben definiert durch eine Folge von Flüssen $\varphi_0, \varphi_1, ..., \varphi_n$, wobei φ_n ein Maximalfluss ist und φ_{k+1} aus φ_k ermittelt wird längs eines kürzesten vergrössernden Weges, und sei $\gamma \in \Gamma'$. Nach Satz 3.3.7 erhöht sich $l_k(a,b)$, $k = 0, ..., n-1$ zwischen je zwei Schritten, in denen γ Flaschenhals ist, um mindestens 2. Andererseits ist trivialerweise

$$l_k(a,b) \leq \#X - 1$$

für alle k. Also ist die Anzahl der Flussvergrösserungen mit γ als Flaschenhals beschränkt durch $\frac{1}{2}\,\#X$. Da es $\#\Gamma'$ Kanten verschieden von (b,a) gibt, und bei jeder Flussvergrösserung mindestens eine davon Flaschenhals ist, folgt die Behauptung. □

Bemerkung 3.3.9.

(i) In Beispiel 3.1.1 ist $\#X = 5$, $\#\Gamma' = 8$, also ergibt sich die obere Schranke $\frac{1}{2} \cdot 5 \cdot 8 = 20$ für die Laufzeit. Tatsächlich hatten wir nur 4 Schritte benötigt.

(ii) Sieht man die Vergrösserung des Flusses auf einer Kante auch als einen Rechenschritt an, so ist der Aufwand für eine einzelne Flussvergrösserung proportional zur Anzahl der Kanten in dem vergrössernden Weg, und man erhält einen Gesamtaufwand der Grössenordnung

$$O(\#X \cdot (\#\Gamma')^2).$$

Bemerkung 3.3.10. Ein weiterer Algorithmus zur Auffindung eines Maximalflusses ist der von *Goldberg-Tarjan* (push-relabel-Algorithmus), er ist der effizienteste bekannte Algorithmus für das Flussproblem. Schreiben wir $n = \#X$ und $m = \#\Gamma'$, so hat man Schranken der Ordnung $O(n \cdot m^2)$ für Ford-Fulkerson (bzw. Edmond-Karp) und $O(n^2 \cdot m)$ für Goldberg-Tarjan. In vielen Anwendungen ist $m \sim n^2$ (ansonsten wäre der Graph 'sparse'), und es ergibt sich $O(n^5)$ für Ford-Fulkerson und $O(n^4)$ für Goldberg-Tarjan. Man kann durch weitere Verbesserungen mit Goldberg-Tarjan sogar $O(n^3)$ erreichen.

Kapitel 4

Spieltheorie

4.1 Einleitung

Was wollen wir im Allgemeinen unter einem Spiel verstehen?

- Ein Spiel ist ein mathematisches Modell für eine Konfliktsituation zwischen mehreren Beteiligten.
- Die Beteiligten haben Auswahlmöglichkeiten für Handlungen. Diese erbringen einen gewissen Nutzen.
- Wir befassen uns mit strategischen Spielen, bei denen der Ausgang vom Verhalten der Beteiligten abhängt (im Gegensatz zu Glücksspielen).
- Eine Strategie eines Spielers ist eine vollständige Voraus-Festlegung seiner Handlungen für alle denkbaren Spielkonstellationen.
- Ziel der Spieltheorie ist es zum Beipiel Vorhersagen, Erklärungen, Untersuchungen und Anweisungen für Spiele zu liefern

Was für Arten von Spielen gibt es?

- Zahl der Spieler: 2-PS (Zweipersonenspiele) n-PS (n-Personenspiele)
- kooperative und nicht kooperative Spiele
- Spiele mit und ohne Seitenzahlungen (Bestechung)
- Nullsummen- und Nichtnullsummenspiele

4.2 Minimax-Stratgien

Definition 4.2.1. Ein *Zweipersonen-Nullsummenspiel (2-PNSS)* liegt in Matrixform vor. Die Zeilen stehen für die *Strategien* des Spielers I, die Spalten für die Strategien des Spielers II. Ein Eintrag a_{ij} der Matrix gibt den Gewinn von I und den Verlust von II an für den Fall, dass I die Strategie der Zeile i und II die Strategie der Spalte j spielt.

Die Auswahl von Strategien also Auswahl einer Zeile i und einer Spalte j nennen wir Strategiepaar (i, j).

Definition 4.2.2. Eine Strategiepaar (i, j) ist im *Gleichgewicht*, falls kein Spieler durch Änderung seiner Strategie einen Auszahlungsvorteil erhalten kann. Die Auszahlung a_{ij} heißt dann Sattelpunkt.

Korollar 4.2.3. *Eine Strategiepaar (i, j) einer $m \times n$-Auszahlungsmatrix $A = (a_{ij})$ ist im Gleichgewicht, genau dann wenn*

$$a_{ij} = \min\{a_{ik} \,|\, 1 \leq k \leq n\}$$

und

$$a_{ij} = \max\{a_{lj} \,|\, 1 \leq l \leq m\}$$

gilt.

Beispiel 4.2.4. In der folgenden Matrix ist $a_{22} = 2$ ein Sattelpunkt.

$$\begin{pmatrix} 5 & 1 & 3 \\ 3 & 2 & 4 \\ -3 & 0 & 4 \end{pmatrix}$$

Proposition 4.2.5. *Sind (i, j) und (i', j') Sattelpunkte in A, dann sind auch (i, j') und (i', j) Sattelpunkte in A und es gilt $a_{ij} = a_{ij'} = a_{i'j} = a_{i'j'}$.*

Beweis. Aus den Sattelpunktseigenschaften folgen

$$a_{ij} \;\leq\; a_{ij'} \;\leq\; a_{i'j'}$$

und

$$a_{i'j'} \;\leq\; a_{i'j} \;\leq\; a_{ij}.$$

Also $a_{ij} = a_{ij'} = a_{i'j} = a_{i'j'}$. Wegen dieser Gleichheiten gilt: mit $a_{i'j'}$ auch $a_{i'j}$ Minimum seiner Zeile. Genauso ist mit a_{ij} auch $a_{i'j}$ Maximum seiner Spalte. Also ist auch (i', j) ein Sattelpunkt. Analog zeigt man, dass (i, j') ein Sattelpunkt ist. □

Beispiel 4.2.6. In der folgenden Matrix gibt es keinen Sattelpunkt.

$$\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$$

Spieler II kann höchstens $3 = \min_j \max_i a_{ij}$ verlieren, Spieler I kann mindestens $2 = \max_i \min_j a_{ij}$ gewinnen.

Definition 4.2.7. Bei einem 2-PNSS besiztz Spieler I den *Gewinnsockel*

$$v_I^* := \max_i \min_j \{a_{ij}\}$$

uns Spieler II den *Verlustdeckel*

$$v_{II}^* := \min_j \max_i \{a_{ij}\}.$$

Lemma 4.2.8. *Es gilt stets*

$$v_I^* \;\leq\; v_{II}^*.$$

Beweis. Sei $a_{pq} = v_I^*$ und $a_{st} = v_{II}^*$. Dann ist a_{pq} ein Minimum in der p-ten Zeile und a_{st} ein Maximium in der t-ten Spalte. Daher gilt

$$v_{II}^* = a_{st} \geq a_{pt} \geq a_{pq} = v_I^*.$$

□

Lemma 4.2.9. *Genau dann ist a_{ij} ein Sattelpunkt von A, wenn $v_I^* = v_{II}^* = a_{ij}$ gilt.*

Beweis. Es sei a_{ij} ein Sattelpunkt. Die Minima der Zeilen seien $z_p := \min_q\{a_{pq}\}$. Also gilt $a_{ij} = z_i$. Angenommen es gibt ein z_s mit $z_s > z_i$. Dann folgt aus $z_s = \min_q\{a_{sq}\} > z_i$ auch $a_{s_j} > z_i = a_{ij}$. Daher ist a_{ij} nicht das Maximum seiner Spalte, also kein Sattelpunkt. Die Annahme ist demnach falsch, es gilt also $z_s \leq z_i$ für alle s und man berechnet

$$v_I^* = \max_i \min_j\{a_{ij}\} = \max_p z_p = z_i = a_{ij}.$$

Für v_{II}^* argumentiert man analog.

Nun gelte $v_I^* = a_{pq} = a_{st} = v_{II}^*$. Da a_{pq} das Maximum von Zeilenminima ist, ist a_{pq} selbst das Minimum seiner Zeile, also gilt $a_{pq} \leq a_{pt}$. Da a_{st} das Minimum von Spaltenmaxima ist, gilt analog $a_{st} \geq a_{pt}$. Insgesamt folgt

$$a_{pq} = a_{pt} = a_{st}$$

also ist a_{pt} das Maximum seiner Spalte und das Minimum seiner Zeile. □

Beispiel 4.2.10. Wir betrachten die Auszahlungsmatrix für das Spiel Stein-Schere-Papier:

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

Es gibt keinen Sattelpunkt, es gilt $v_I^* = -1$ und $v_{II}^* = 1$.

4.3 Gemischte Strategien

Definition 4.3.1. In einem 2-PNSS habe ein Spieler $m \in \mathbb{N}$ Strategien zur Wahl. Eine *gemischte Strategie* ist ein Vektor $x \in \mathbb{R}^m$ mit

$$x \geq 0 \quad \text{und} \quad \sum_{i=1}^{m} x_i = 1.$$

Die x_i benennen also die Wahrscheinlichkeit, mit der ein Spieler Strategie i spielt.

Lemma 4.3.2. *Bei einem 2-PNSS mit Auszahlungsmatrix $A = (a_{ij})$ seien $x \in \mathbb{R}^m$ und $y \in \mathbb{R}^n$ gemischte Strategien für Spieler I bzw. II. Wählen die Spieler die Strategien unabhängig voneinander, so beträgt die erwartete Auszahlung*

$$\sum_{i=1}^{m} \sum_{j=1}^{n} x_i a_{ij} y_j = x^T A y \, .$$

Beweis. Wegen der stochastischen Unabhängigkeit hat man die Multiplikativität:

$$P(\text{ I spielt } i \text{ und II spielt } j) = P(\text{ I spielt } i) \cdot P(\text{ II spielt } j) = x_i \cdot x_j.$$

Alles weitere ist Gewichtung der Spielausgänge. □

Spielt Spieler I die Strategie x und Spieler II weiß das, so wird Spieler II seine Strategie y so wählen, dass die erwartete Auszahlung $x^T Ay$ minimal ist. Dann ist

$$v_I(x) := \min\{x^T Ay \,|\, y \text{ Strategie für II}\}$$

die Auszahlung für I. Für festes x liegt also ein lineares Optimierungsproblem in y mit Nebenbedingungen $y \geq 0$ und $\sum y_j = 1$ vor. Diese liefern eine kompakte Teilmenge des $\mathbb{R}^n$, daher wird das Minimum angenommen. Betrachten wir als zulässigen Bereich das Polyeder

$$P := \{y \in \mathbb{R}^n \,|\, y \geq 0,\ \sum_{j=1}^{n} y_j \leq 1\}.$$

Das Minimum wird in einer Ecke von P angenommen. Nun ist P aber das Einheitssimplex mit den Einheitsvektoren e_j, $1 \leq j \leq n$ als Ecken. Also gilt

$$v_I(x) := \min\{x^T Ae_1, \ldots, x^T Ae_n\}.$$

Mit gemischten Strategien kann sich Spieler I demnach

$$\begin{aligned} v_I &:= \max\{v_I(x) \,|\, x \text{ Strategie für I}\} \\ &= \max\{\min\{x^T Ae_1, \ldots, x^T Ae_n\} \,|\, x \text{ Strategie für I}\} \end{aligned} \tag{4.1}$$

sichern. Umgekehrt kann sich Spieler II gegen einen höheren Verlust als

$$v_{II} := \min\{\max\{e_1^T Ay, \ldots, e_m^T Ay\} \,|\, y \text{ Strategie für II}\}$$

absichern. Betrachten wir nun die Menge

$$P_I := \left\{(x, \lambda) \in \mathbb{R}^m \times \mathbb{R} \,|\, x \geq 0, v_I(x) \geq \lambda, \sum_{i=1}^{m} x_i = 1\right\}$$

und die lineare Funktion

$$f : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}, \quad f(x, \lambda) := \lambda.$$

Dann ist

$$v_I = \max\{f(x, \lambda) \,|\, (x, \lambda) \in P_I\} = f(x^*, v_I),$$

denn wir wissen bereits dass v_I als Maximum in (4.1) auftritt. Nun läßt sich P_I auch schreiben als

$$P_I = \left\{(x, \lambda) \in \mathbb{R}^m \times \mathbb{R} \,|\, x \geq 0, \forall j \in \{1, \ldots, n\} : x^T Ae_j \geq \lambda, \sum_{i=1}^{m} x_i = 1\right\}$$

und man erhält ein lineares Optimierungsproblem für Spieler I:

$$f(x, \lambda) := \lambda \to \max, \quad (x, \lambda) \in P_I.$$

Das dazu duale Problem lautet

$$g(y, \mu) := \mu \to \min, \quad (y, \mu) \in P_{II}$$

wobei

$$P_{II} = \left\{(y, \mu) \in \mathbb{R}^n \times \mathbb{R} \,|\, y \geq 0, \forall i \in \{1, \ldots, m\} : e_i^T Ay \leq \mu, \sum_{j=1}^{n} y_j = 1\right\}$$

gilt und es gilt

$$v_{II} = \min\{g(y, \mu) \,|\, (y, \mu) \in P_{II}\} = g(y^*, v_{II}).$$

Satz 4.3.3. *Für 2-PNSS gilt $v_I = v_{II}$. Diese Werte und Strategien für Spieler I und II können mit einem linearen Optimierungsprobelem ermittelt werden.*

Beweis. Die zueinander dualen Optimierungsprobleme haben wir schon hergeleitet. Da $\lambda, \mu \in \mathbb{R}$ frei wählbar sind, ist sowohl P_I als auch P_{II} nichtleer. Aus dem Dualitätssatz folgt nun die Behauptung. □

4.4 Bimatrixspiele

Für ein Spiel mit zwei Spielern habe Spieler i genau $k_i \in \mathbb{N}$ Strategien zur Verfügung. Die Auszahlungsfunktion für Spieler i können wir als eine $k_1 \times k_2$ Matrix A_i definieren. Spielt Spieler j die Strategie l_j, so erhält Spieler i die Auszahlung $(A_i)_{(l_1,l_2)}$.

Das Tupel (A_1, A_2) nennen wir *Bimatrixspiel*.

Definition 4.4.1. Ein Paar (x^*, y^*) von gemischten Strategien für ein Bimatrixspiel (A_1, A_2) nennt man im *Gleichgewicht*, falls für alle anderen gemischten Strategien (x, y) gilt:

$$x^T A_1 y^* \leq (x^*)^T A_1 y^* \quad \text{und} \quad (x^*)^T A_2 y \leq (x^*)^T A_2 y^*.$$

Bemerkung 4.4.2. Sind (x^*, y^*) im Gleichgewicht, so kann keiner der Spieler seine erwartete Auszahung durch eine Änderung seiner Strategie verbessern.

Satz 4.4.3. *(Fixpunktsatz von Brouwer) Sei f eine stetige Abbildung von einer nichtleeren, kompakten, konvexen Teilmenge eines endlichdimensionalen Banachraumes in sich selbst. Dann hat f einen Fixpunkt.*

Satz 4.4.4. *Jedes Bimatrixspiel besitzt ein Strategiepaar im Gleichgewicht.*

Beweis. Es seien A_1 und A_2 die Matrizen des Spiels. Es seien

$$S_1 = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i = 1\}$$

$$S_2 = \{y = (y_1, \ldots, y_m) \in \mathbb{R}^m \mid \sum_{j=1}^{m} y_j = 1\}$$

die Mengen der gemischten Strategien. Zu $(x, y) \in S_1 \times S_2$ definieren wir für $i \in \{1, \ldots, n\}$ und $j \in \{1, \ldots, m\}$

$$c_i(x, y) := \max\{0, e_i^T A_1 y - x^T A_1 y\},$$
$$d_j(x, y) := \max\{0, x^T A_2 e_j - x^T A_2 y\}.$$

Jedes $z \in S_1$ läßt sich als Konvexkombination der Ecken des Simplex S_1 dartstellen, also

$$z = \sum_{i=1}^{n} z_i e_i.$$

Angenommen für alle $i \in \{1, \ldots, n\}$ gilt $c_i(x, y) = 0$. Dann folgt für jede Strategie $z \in S_1$:

$$z^T A_1 y = \sum_{i=1}^{n} z_i e_i^T A_1 y \leq \sum_{i=1}^{n} z_i x^T A_1 y = x^T A_1 y.$$

Falls $d_j(x,y) = 0$ für alle $j \in \{1,\dots,m\}$ gilt erhalten wir für jede Strategie $z \in S_2$:

$$x^T A_2 z = \sum_{i=1}^{m} z_j x^T A_2 e_j \leq \sum_{i=1}^{m} z_j x^T A_2 y = x^T A_2 y.$$

Das bedeutet, dass (x,y) in diesem Fall ein Gleichgewichtspunkt ist.
Wir definieren nun eine Abbildung

$$T : S_1 \times S_2 \to S_1 \times S_2, \quad T(x,y) := \left(\frac{x + c(x,y)}{1 + \sum_{i=1}^{n} c_i(x,y)}, \frac{y + d(x,y)}{1 + \sum_{j=1}^{m} d_j(x,y)} \right)$$

Damit ist T wohldefiniert und stetig. Wir wollen nun zeigen, dass folgendes gilt:

$$T(x,y) = (x,y) \quad \Leftrightarrow \quad \forall i : c_i(x,y) = 0 \wedge \forall j : d_j(x,y) = 0.$$

Die Richtung $\Leftarrow$ ist trivial. Für die andere Richtung sei nun (x,y) ein Fixpunkt von T. Angenommen es sind alle $c_i(x,y) > 0$. Da sich x als Konvexkombination darstellen läßt ergibt sich daraus der Widerspruch $x^T A_1 y > x^T A_1 y$ aus der Definition von $c_i(x,y)$. Also gibt es einen Index l mit $c_l(x,y) = 0$. Für die l-te Koordinate des Fixpunktes ergibt sich daraus

$$x_l = \frac{x_l + c_l}{1 + \sum_{i=1}^{n} c_i(x,y)} = \frac{x_l}{1 + \sum_{i=1}^{n} c_i(x,y)}.$$

Daher muss $\sum_{i=1}^{n} c_i(x,y) = 0$ und somit $c_i(x,y) = 0$ für alle i gelten. Analog sieht man das für d und die Richtung $\Rightarrow$ ist gezeigt.

Es ergibt aich also:

$$T(x,y) = (x,y) \quad \Leftrightarrow \quad (x,y) \text{ ist im Gleichgewicht.}$$

Da f eine stetige Selbstabbildung auf der konvexen und kompakten Menge $S_1 \times S_2$ ist, liefert der Fixpunktsatz von Brouwer den gewünschten Fixpunkt von T. □

Beispiel 4.4.5 (Battle)**.**

$$A_1 = \begin{pmatrix} 4 & 0 \\ -1 & 1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 1 & 0 \\ -1 & 4 \end{pmatrix}$$

Gleichgewichtspaare für Spieler 1 und 2 sind:

- $(1,0)$ und $(1,0)$ mit Auszahlung $(4,1)$.
- $(0,1)$ und $(0,1)$ mit Auszahlung $(1,4)$.

Wegen der unterschiedlichen Auszahlungen sind diese Paare nicht besonders attraktiv, denn wer sich zuerst festlegt zwingt den anderen in das Gleichgewichtspaar. Keine Gleichgewichtspaare sind hingegen:

- $(1,0)$ und $(0,1)$ mit Auszahlung $(0,0)$.
- $(0,1)$ und $(1,0)$ mit Auszahlung $(-1,-1)$.

Wir betrachten die gemischten Strategien. Spieler 1 spiele mit $(p, 1-p)$ und Spieler 2 mit $(q, 1-q)$. Die erwarteten Auszahlungen für Spieler 1 bzw. 2 sind

$$f_1(p,q) = p(6q-1) + 1 - 2q, \quad f_2(p,q) = q(6p-5) + 4(1-p). \tag{4.2}$$

Für $q = \frac{1}{6}$ kann Spieler 1 seine Auszahlung nicht mehr beeinflussen. Für $p = \frac{5}{6}$ trifft das auf Spieler 2 zu. Daher ist

$$(p, 1-p) = (\tfrac{5}{6}, \tfrac{1}{6}), \quad (q, 1-q) = (\tfrac{1}{6}, \tfrac{5}{6})$$

ein Gleichgewichtspaar. Die erwartete Auszahlung beträgt dann $\frac{24}{36}$ für jeden.

Also ergibt sich bei der radikalen Lösung jeweils etwas besseres als bei dem Kompromiss.

Spielt aber Spieler 2 mit $q > \frac{1}{6}$, so ist $6q - 1 > 0$ und Spieler 2 kann für $p = 1$ seine Auszahlung maximieren. Er wählt also die reine Strategie von oben. Spielt Spieler 2 mit $q < \frac{1}{6}$, so ist $6q - 1 < 0$ und Spieler 1 whlt $p = 0$.

Zum Vergleich berechnen wir nun die Minimax-Strategien für beide Spieler. Für Spieler 1 verwenden wir (4.1) mit Auszahlungsmatrix A_1 und die Strategie $x = (p, 1-p)$:

$$v_1(x) = \min\{x^T A_1 e_1, x^T A_1 e_2\} = \min\{5p - 1, 1 - p\}$$

also

$$v_1 = \max\{v_1(x) \mid p \in [0,1]\} = \tfrac{2}{3} =: p_0.$$

Analog erhalten wir für Spieler 2 mit Strategie $y = (q, 1-q)$ und Matrix A_2:

$$v_2 = \max\left\{\min\{e_1^T A_2 y, e_2^T A_2 y\} \mid q \in [0,1]\right\} = \tfrac{2}{3} =: q_0.$$

Die Auszahlungen gemäß (4.2) berechnen sich zu

$$f_1(p_0, 1-p_0) \;=\; \tfrac{2}{3} \;=\; f_2(q_0, 1-q_0).$$

Man erreicht also genauso viel wie beim obigen Gleichgewichtspunkt. Allerdings handelt es sich hier nicht um einen solchen.

Bemerkung 4.4.6.

- Gleichgewichtspunkte sind nicht automatisch Minimax-Strategien und umgekehrt.
- Verschiedene Gleichgewichtspunkte können verschiedene Auszahlungen haben.

Bemerkung 4.4.7. Die Methode im Beweis von (4.4.4) kann man zum Berechnen von Gleichgewichtspunkten verwenden.

4.5 Kooperative Spiele

Wir betrachten ein Bimatrixspiel. Ein Auszahlungspaar $(x_1, x_2) \in \mathbb{R}^2$ nennen wir *Garantiepunkt*, falls sich Spieler i mindestens x_i sichern kann. Mit gemischten Strategien können sich die beiden Spieler mit ihrer Maxmin-Strategie (4.1) jeweils mindestens v_1 bzw. v_2 an Auszahlung sichern. Daher ist (v_1, v_2) ein Garantiepunkt.

Definition 4.5.1. Die Menge

$$\begin{aligned}\mathcal{A} \;:=\; \big\{&(x, A) \mid x \in \mathbb{R}^2, A \subset \mathbb{R}^2 \text{ konvex und kompakt},\\ &\forall a \in A : x \le a, \exists a' \in A : x < a'\big\}\end{aligned}$$

Bemerkung 4.5.2.

- Die Menge $\mathcal{A}$ ist die Menge der Auszahlungskombinationen, die die Spieler für realitisch halten. Ein Element $(x, A) \in \mathcal{A}$ nennen wir *Verhandlungssituation*. Die Menge A selbst *Verhandlungsmenge*.
- Dabei ist x ein Garantiepunkt. Aber es gibt für beide eine echte Verbesserung $a' \in A$.
- Gesucht ist nun eine Verhandlungslösung zu jedem $(x, A) \in \mathcal{A}$. Der Garantiepunkt (v_1, v_2) ist eine solche und kann als *Konfliktlösung* betrachtet werden
- Die Konvexität von A ergibt sich aus der Annahme, dass die Spieler aus einer Verhandlungsmenge durch eine Lotterie neue Verhandlungsmengen erzeugen können bzw. wollen.

Definition 4.5.3. Eine *Verhandlungslösung* auf A ist eine Abbildung

$$\varphi : \mathcal{A} \to A \subset \mathbb{R}^2,$$

die jedem (x, A) eine Auszahlungspunkt $\varphi(x, A) \in A$ zuordnet.

Man versucht nun vernünfitge Anforderungan an φ zu stellen, so dass eine Lösung eindeutig festgelegt ist. Die folgenden hat Nash 1950 formuliert.

Axiome 4.5.4.

R1 „Individuelle Rationalität“

$$\forall\, (x, A) \in \mathcal{A} \;:\; \varphi(x, A) \;\geq\; x.$$

P1 „Pareto-Optimalität“

$$\varphi(x, A) \in P_W(A) := \{a \in A \,|\; \nexists y \in A \,:\, y > a\}.$$

S1 „Symmetrie“

Für eine symmetrische Verhandlungssituation (x, A), das bedeutet

$$x_1 = x_2 \quad \text{und} \quad (a_1, a_2) \in A \Leftrightarrow (a_2, a_1) \in A$$

gilt

$$(\varphi(x, a))_1 \;=\; (\varphi(x, a))_2 \,.$$

T2 „Unabhängigkeit von positiven lineraen Transformationen“

Sind $\alpha_1, \alpha_2 > 0$ und $\beta_1, \beta_2 \in \mathbb{R}$ und $T : \mathbb{R}^2 \to \mathbb{R}^2$ definiert durch

$$T(a_1, a_2) := (\alpha_1 a_1 + \beta_1, \alpha_2 a_2 + \beta_2)$$

dann gilt

$$\forall\, (x, A) \in \mathcal{A} \;:\; \varphi\,(T(x), T(A)) = T(\varphi(x, A)).$$

U „Unabhägigkeit von irrelvanten Alternativen“

Sind $(x, A), (x, B) \in \mathcal{A}$ mit $\varphi(x, A) \in B \subset A$, dann gilt

$$\varphi(x, B) = \varphi(x, A).$$

Bemerkung 4.5.5.

R1 Niemand wird sich mit weniger begnügen, als man ohne Verhandlung erreichen kann.

P1 Es soll keine gleichzeitige echte Verbesserung geben.

S1 Symmetrische Verhandlungssituationen ergeben symmetrische Konsequenzen.

T2 Ersetzt ein Spieler seine Verhandlungssituation durch eine Äquivalente, so gewinnt oder verliert er dadurch nichts in der Verhandlungslösung.

U Eine anerkannte Lösung soll sich nicht ändern, wenn man einige Verhandlungssituationen, die man vorher nicht betrachtet hat weglässt. Dazu ein Beispiel. Es sei

$$A := \{(a_1, a_2) \in \mathbb{R}^2 \mid a_1 + a_2 \leq 1\}, \quad x = (0, 0).$$

Nach Axiom S1 und P1 muss die Lösung $(\frac{1}{2}, \frac{1}{2})$ sein. Schränkt man nun A wie folgt ein

$$B := A \cap \{a_1 \leq \tfrac{1}{2}\}$$

so bleibt $(\frac{1}{2}, \frac{1}{2})$ nach Axiom U die Lösung. Spieler I erhält sein Maximum aber Spieler II könnte mehr bekommen.

Wir wollen im folgenden zeigen, dass es nur eine Verhandlungslösung gibt, die den Axiomen genügt.

Lemma 4.5.6. *Sei $(x, A) \in \mathcal{A}$, $A \neq \emptyset$ eine Verhandlungssituation. Dann gibt es genau ein*

$$a^* := (a_1^*, a_2^*)^T \in A,$$

das die Funktion

$$f : A \to \mathbb{R}, \quad f\left((a_1, a_2)^T\right) := f(a_1, a_2) := (a_1 - x_1)(a_2 - x_2)$$

maximiert.

Beweis. Die Abbildung f ist stetig auf der kompakten Menge A, dager ist

$$M := \max\{f(a) \mid a\ A\} \in \mathbb{R}.$$

Da es definitionsgemäß ein $a' \in A$ mit $a' > x$ gibt, gilt $M > 0$. Angenommen es gibt verschiedene $(a_1', a_2')^T$ und $(a_1'', a_2'')^T$ aus A mit

$$M = f(a_1', a_2') = f(a_1'', a_2'').$$

Falls $a_1' = a_1''$ gilt, so folgt $a_2' = a_2''$ aus $M > 0$. Daher muss oBdA $a_1' < a_1''$ gelten. Da beide Tupel auf M abgebildet werden, ergibt sich daraus $a_2' > a_2''$. Nun betrachten wir die Konvexkombination

$$b := \tfrac{1}{2}(a' + a'') \in A.$$

Damit ergibt sich

$$\begin{aligned} f(b) &= \left(\tfrac{1}{2}(a_1' + a_1'') - x_1)\right)\left(\tfrac{1}{2}(a_2' + a_2'') - x_2)\right) \\ &= \tfrac{1}{2}(a_1' - x_1)(a_2' - x_2) + \tfrac{1}{2}(a_1'' - x_1)(a_2'' - x_2) + \tfrac{1}{4}(a_1' - a_1'')(a_2'' - a_2') \\ &= \tfrac{1}{2}f(a') + \tfrac{1}{2}f(a'') + \tfrac{1}{4}(a_1' - a_1'')(a_2'' - a_2') \\ &= M + \tfrac{1}{4}(a_1' - a_1'')(a_2'' - a_2') \\ &> M \end{aligned}$$

Also ist die Annahme falsch und die Maximalstelle somit eindeutig. □

Lemma 4.5.7. *Vorgelegt sei die Situation aus Lemma (4.5.6) und die Abbildung*

$$h(a_1, a_2) := (a_1^* - x_1)a_2 + (a_2^* - x_2)a_1.$$

Dann gilt für alle $(a_1, a_2) \in A$:

$$h(a_1, a_2) \leq h(a_1^*, a_2^*).$$

Beweis. Es sei $\epsilon \in (0, 1]$ und $\hat{a} \in A$. Dann ist auch die Konvexkombination

$$a^* + \epsilon(\hat{a} - a^*) = (1 - \epsilon)a^* + \epsilon\hat{a}$$

in A. Nach Lemma (4.5.6) maximiert a^*, also

$$\begin{aligned}
(a_1^* - x_1)(a_2^* - x_2) &\geq (a_1^* + \epsilon(\hat{a}_1 - a_1^*)(a_2^* + \epsilon(\hat{a}_2 - a_2^*)) && (4.3)\\
&= (a_1^* - x_1)(a_2^* - x_2) + \epsilon^2(\hat{a}_1 - a_1^*)(\hat{a}_2 - a_2^*) && (4.4)\\
&\quad +\epsilon(a_1^* - x_1)(\hat{a}_2 - a_2^*) + \epsilon(a_2^* - x_2)(\hat{a}_1 - a_1^*) && (4.5)
\end{aligned}$$

Umstellen ergibt

$$\epsilon(\hat{a}_1 - a_1^*)(\hat{a}_2 - a_2^*) \geq (a_1^* - x_1)(\hat{a}_2 - a_2^*) + (a_2^* - x_2)(\hat{a}_1 - a_1^*)$$

Nun liefert $\epsilon \to 0$ und ausrechnen die Behauptung

$$h(\hat{a}_1, \hat{a}_2) \leq h(a_1^*, a_2^*).$$

□

Satz 4.5.8. *Es gibt nur eine Verhandlungslösung, die die Axiome 1 bis 5 erfüllt. Diese Lösung ist der Punkt* a^* *aus Lemma (4.5.6) und wird* Nash-Lösung *genannt.*

Beweis. Wir zeigen zunächst, dass $\varphi((x, A)) := a^*$ die Axiome erfüllt.

Für Axiom R1 folgt dies direkt aus der Definition.

Gäbe es $y \in A$ mit $y > a^*$ würde dies der Maximalität von $f(a)$ widersprechen. Also gilt Axiom P1.

Für die Maximierung gilt die Unabhängigkeit von irrelevanten Alternativen offensichtlich, denn maximiert man über die Menge A und findet die Maximalstelle in $B \subset A$, so ist die Stelle auch für B maximale Stelle.

Für eine Transformation $T(a_1, a_2) := (\alpha_1 a_1 + \beta_1, \alpha_2 a_2 + \beta_2)$ berechnet man für die Verhandlungssituation $(T(x), T(A))$:

$$\begin{aligned}
f\left(T(a_1, a_2)\right) &= (\alpha_1 a_1 + \beta_1 - \alpha_1 x_1 - \beta_1)(\alpha_2 a_2 + \beta_2 - \alpha_2 x_2 - \beta_2)\\
&= \alpha_1\alpha_2(a_1 - x_1)(a_2 - x_2).
\end{aligned}$$

Wegen $\alpha_1, \alpha_2 > 0$ maximiert a^* auch $f \circ T$, daher gilt

$$\varphi(T(x), T(A)) = (\alpha_1 a_1^* + \beta_1, \alpha_2 a_2^* + \beta_2) = T(\varphi(x, A)).$$

Zur Symmetrie: es sei $x_1 = x_2$. Falls (a_1^*, a_2^*) die Funktion f maximiert, so maximiert damit auch (a_2^*, a_1^*). Aus der Eindeutigkeit folgt $a_1^* = a_2^*$.

Es bleibt zu zeigen, dass kein anderer Vektor aus A die Axiome erfüllen kann. Sei also a^* die Lösung zu (x, A). Wir betrachten die Menge

$$U := \{(u_1, u_2) \,|\, h(u_1, u_2) \leq h(a_1^*, a_2^*)\}.$$

Aus Lemma (4.5.7) wissen wir, dass $A \subset U$ gilt. Wir betrachten die positive lineare Transformation

$$T : U \to T(U) := V, \quad T(u_1, u_2) := \left(\frac{u_1 - x_1}{a_1^* - x_1}, \frac{u_2 - x_2}{a_2^* - x_2} \right) =: (u_1', u_2').$$

Es gilt

$$(u_1, u_2) \in U \Leftrightarrow u_1' + u_2' \leq 2,$$

daher ist

$$V = \{(u_1', u_2') \,|\, u_1' + u_2' \leq 2\}.$$

Offenbar gilt $T(x_1, x_2) = (0, 0)$. Die Lösung gemäß der Axiome der Verhandlungssituation $(V, 0, 0)$ bekommen wir so: wegen der Symmetrie muss diese die Form $(v, v) \in V$ haben. Aber nur $(1, 1)$ ist in V pareto-optimal. Durch Rücktransformation erhalten wir daher eine eindeutige Lösung (u_1^*, u_2^*) für (x, U) durch

$$(1, 1) = T(u_1^*, u_2^*).$$

Einsetzen in T liefert dann $u_1 = a_1^*, u_2 = a_2^*$. Wegen $a^* \in A \subset U$ folgt aus der Unabhängigkeit von irrelevanten Alternativen, dass a^* auch die einzige Lösung für (x, A) ist. □

Beispiel 4.5.9. Wir betrachten Verhandlungen zwischen einem Unternehmen und Arbeitskräften. Das Unternehmen ist durch eine Produktionsfunktion f charakterisiert. Die Produktion f hänge nur von der Anazahl a der Arbeitskräfte ab. Sie sei monoton wachsend und konkav. Es sei w die Lohnhöhe einer Arbeitskraft. Das Verhandlungsziel ist eine Lohnhöhe und ein Beschäftigungsniveau (w, a) zu bestimmen. Bei einer anderen Firma wird mit Lohn w_0 vergütet. Der Gewinn der Firma ist $f(a) - wa$. Der Gesamtlohn der Arbeiskräfte beträgt $wa + (A - a)w_0$. Dabei ist A die Anzahl der insgesamt zur Verfügung stehenden Arbeitskräfte für die Firma. Die Menge

$$V' := \{(f(a) - wa, wa + (A - a)w_0) \,|\, f(a) \geq wa, 0 \leq a \leq A, w \geq w_0\} \subset \mathbb{R}^2$$

enthält demnach Elemente der Verhandlungsmenge. Führen die Verhandlungen zu keinem Ergebnis, so werden die Arbeitskräfte abwandern. Die Konfliktlösung ist daher $(0, w_0 A)$. Für $v = (v_1, v_2) \in V'$ gilt

$$v_1 + v_2 = f(a) + w_0(A - a).$$

Wegen der Stetigkeit nimmt der Ausdruck auf der rechten Seite in einem $a^* \in [0, A]$ ein Maximum an. Die Menge der Verhandlungslösungen können wir nun so darstellen:

$$V := \{(v_1, v_2) \,|\, v_1 + v_2 \leq f(a^*) + (A - a^*)w_0, v_1 \geq 0, v_2 \geq w_0 A\}.$$

Diese Menge zusammen mit der Konfliktlösung genügt der Definiton einer Verhandlungssituation.
Wir wollen das Verhandlungsergebnis berechnen. Wegen der Pareto-Optimalität muss für ein Verhandlungsergebnis in V Gleichheit $v_1 + v_2 = f(a^*) + (A - a^*)w_0$ eintreten. Daher muss das Beschäftigungsniveau a^* sein, denn für ein anderes a wird ja $v1 + v_2 = f(a) + w_0(A - a)$ nicht maximal. Nun müssen wir die Funktion

$$f(v_1, v_2) := (v_1 - 0)(v_2 - w_0 A)$$

maximieren. Wegen

$$v_1 = f(a^*) + (A - a^*)w_0 - v_2$$

müssen wir

$$\max_{v_2} (f(a^*) + (A - a^*)w_0 - v_2)(v_2 - w_0 A)$$

berechnen. Die Ableitung nach v_2 verschwindet für

$$v_2 = \tfrac{1}{2}(f(a^*) + 2w_0 A - a^* w_0).$$

Um die Lösung für den Lohn w zu erhalten lösen wir

$$\tfrac{1}{2}(f(a^*) + 2w_0 A - a^* w_0) = wa^* + (A - a^*)w_0$$

nach w auf und erhalten

$$w = \frac{1}{2}\left(\frac{f(a^*)}{a^*} + w_0\right).$$

Beispiel 4.5.10 (Drohungen)**.** Wir betrachten ein Bimatrixspiel gegeben durch

$$A = \begin{pmatrix} 4 & 8 \\ 0 & 6 \end{pmatrix} \qquad \text{und} \qquad B = \begin{pmatrix} 8 & 4 \\ -\frac{5}{2} & -2 \end{pmatrix}.$$

Die Verhandlungsmenge sei damit

$$k\left(\left\{(4,8)^T, (8,4)^T, (0,-\tfrac{5}{2})^T, (6,-2)^T\right\}\right).$$

Ein Gleichgewichtspunkt reiner Strategien ist $(4,8)^T$. Mit der Minimax-Strategie kann sich Spieler A gemäß seiner Auszahlungsmatrix mindestens 4 sichern, Spieler B hingegen nur -2. Vereinbart man $(4,-2)^T$ als Garantiepunkt muss man für die Nash-Lösung auf $y = 12 - x$ und $4 \leq x \leq 8$ die Funktion

$$(x-4)(y+2) = (x-4)(14-x) =: f(x)$$

maximieren. Das lokale Maximum von f liegt bei $x = 9$. Maximieren von f auf $[4,8]$ liefert daher die Nashlösung $(8,4)^T$. Das ist für Spieler A besser als vorher.

Wenn sich Spieler B auf -2 als Verhandlungsbasis einlässt, dann fährt er aber mit seiner anderen Strategie y_1, in der er $-\frac{5}{2}$ bekommt, nicht viel schlechter. Das würde aber Spieler A dazu bringen seine erste Strategie x_1 zu spielen, um 4 statt 0 zu erhalten. Daher *droht* Spieler B mit Strategie y_1.

Die Drohstrategie für A ist dagegen x_2, denn dort verringert sich die Auszahlung von B. Spielen beide nun diese gemischte Variante x_2 und y_1 so ergibt sich $(0,-\frac{5}{2})^T$ als Garantiepunkt. Für die Nashlösung muss also die Funktion

$$(x-0)(y+\tfrac{5}{2}) = x(14.5-x) := g(x)$$

maximiert werden. Das ergibt $(7.25, 4.75)^T$ als Nash-Lösung.

Für B hat sich das Drohen gelohnt, für A nicht.

4.6 n-Personenspiele

4.6.1 Kooperative n-Personenspiele

Falls bei einem Spiel mit n Spielern Koalitionen erlaubt sind, wird jeder Einzelspieler versuchen, einer für ihn bestmöglichen Koalition beizutreten.

Natürlich sollte die Teilnahme an einer Koalition mindestens so viel Auszahlung ergeben, wie man sich alleine sichern könnte.

Mit hilfe von Seitenzahlungen können Koalitionäre aus ihrer Koalition herausgelockt werden. Daher ergibt sich die Frage nach einer gewissen „Stabilität“von Koalitionen.

Beispiel 4.6.1. Drei Spieler sollen Koalitionen bilden, bei denen jeweils zwei ein Bündniss eingehen. Dies wied für beide jeweils mit Auszahlung 1 belohnt, der Außenseiter hingegen muss 2 zahlen. Es sind also folgende Auszahlungen möglich:

$$(-2,1,1)^T, \quad (1,-2,1)^T, \quad (1,1,-2)^T.$$

Falls keine Koalition zustande kommt, erhält jeder nichts also wird $(0,0,0)^T$ ausgezahlt. Es seien nun 2 und 3 verbündet. Jetzt zahlt 1 an 2 einen Betrag von 0.1 gegen das Versprechen mit ihm zu koalieren. Das ergäbe als Auszahlung $(0.9, 1.1, -2)^T$. Beide haben sich verbessert, 3 hat sich verschlechtert. Nun könnte Speiler 3 dem Spieler 2 mehr bieten, zum Beispiel 0.2, das ergäbe $(-2, 1.2, 0.8)^T$ als Auszahlung.

Die Koalitionen sind „instabil“.

Definition 4.6.2. Sei $N = \{1, \ldots, n\}$ die Menge der Spieler. Jede nicht-leere Teilmenge $K \subset N$ heißt eine *Koalition*.

Wir sind an den Auszahlungen für Koalitionen interessiert. Dazu abstrahieren wir von den Spielregeln.

Definition 4.6.3. Die *charakteristische Funktion* $\nu : 2^N \to \mathbb{R}$ eines n-Personenspiels mit Spielermenge $N = \{1, \ldots, n\}$ ist eine Abbildung, die die Bedingungen $\nu(\emptyset) = 0$ und

$$\nu(S) \ + \ \nu(T) \ \leq \ \nu(S \cup T) \quad \text{für alle} \quad S, T \subset N \text{ mit } S \cap T = \emptyset$$

erfüllt.

Bemerkung 4.6.4. Die Subadditivität in Definition 4.6.3 besagt, dass sich eine Koalition mindestens so viel sichern kann, wie sie sich einzeln sichern können.

Wir wollen eine charakteristische Funktion angeben. Dazu verallgemeinern wir die Terminologie der Bimatrixspiele. Es seien $S_1, \ldots, S_n$ Strategiemengen der n Spieler. Die Auszahlungsfunktionen seien für $k \in \{1, \ldots, n\}$ gegeben durch

$$a_k : S_1 \times \cdot \times S_n \to \mathbb{R}.$$

Die Strategiemengen sollen jeweils endlich (nur reine Strategien) oder gemischt (also Simplizes) sein.

Proposition 4.6.5. *Für $n \in \mathbb{N}$ sei $N = \{1, \ldots, n\}$ die Spielermenge und S_k deren Strategiemengen. Für $K \subset N$ sei S_K die Menge der Strategien der Koalitionäre aus K, also $S_K = \times_{k \in K} S_k$. Wir definieren $\nu : 2^N \to \mathbb{R}$ durch*

$$\nu(K) := \sup_{x \in S_K} \inf_{y \in S_{N-K}} \sum_{k \in K} a_k(x, y).$$

Dann ist ν eine charakteristische Funktion.

Beweis. Zunächst ist ν wohldefiniert, denn auf endlichen beziehungsweise kompakten Mengen werden Maximum und Minimum angenommen. Weiter gilt definitionsgemäß $\nu(\emptyset) = 0$. Zu zeigen bleibt die Subadditivität. Zur Abkürzung sei

$$f_K(x,y) := \inf_{y \in S_{N-K}} \sum_{k \in K} a_k(x,y)$$

also ist

$$\nu(K) = \sup_y f_K(x,y).$$

Sei $\epsilon > 0$. Zu $K \subset N$ gibt es $x_K \in S_K$ mit

$$f_K(x_K, y) \geq \nu(K) - \epsilon. \tag{4.6}$$

Betrachte nun eine disjunkte Zerlegung $K = K_1 \cup K_2$, $K_1 \cap K_2 = \emptyset$ und $x_{K_1} \in S_{K_1}$, $x_{K_2} \in S_{K_2}$, für die jeweils (4.6) gilt. Für $x_K := (x_{K_1}, x_{K_2})$ ergibt sich

$$\begin{aligned}
\nu(K) &\geq f(x_K, y) \\
&\geq \inf_{y \in S_{N-K}} \sum_{k \in K_1} a_k(x_K, y) \quad + \quad \inf_{y \in S_{N-K}} \sum_{k \in K_2} a_k(x_K, y) \\
&\geq f_{K_1}(x_{K_1}, y) + f_{K_2}(x_{K_2}, y) \\
&\geq \nu(K_1) - \epsilon + \nu(K_2) - \epsilon.
\end{aligned}$$

Für $\epsilon \to 0$ folgt nun die Behauptung. □

Definition 4.6.6.

(i) Ein kooperatives n-PS nennt man *Nullsummenspiel* falls

a) $\nu(N) = 0$

b) $\forall\, \emptyset \neq K \subset N \;:\; \nu(K) + \nu(N-K) = 0.$

(ii) Ein kooperatives n-PS nennt man *Konstantsummenspiel* falls

$$\forall\, \emptyset \neq K \subset N \;:\; \nu(K) + \nu(N-K) = \nu(N)$$

gilt.

Wir definieren zwei Klassen von kooperativen n-Personenspielen (N, ν).

Definition 4.6.7.

1. Ein Spiel (N, ν) heißt *wesentlich*, falls

$$\sum_{i \in N} \nu(i) \;<\; \nu(N)$$

gilt.

2. Ein Spiel (N, ν) heißt *unwesentlich*, falls

$$\sum_{i \in N} \nu(i) \;=\; \nu(N)$$

gilt.

Bemerkung 4.6.8. Unwesentliche Spiele sind aus Sicht von Koalitionen nicht interessant, denn schließen sich zwei Spieler zusammen, ist die Subadditivität von ν für sie eine Gleichheit.

4.6.2 Imputationen

Es stellt sich die Frage, wie die Koalitionäre die Auszahlung aufteilen sollten. Jeder sollte mindestens so viel bekommen, wie er ohne Koalition erhalten kann und eine große Koalition sollte den maximalen Gewinn des realisieren. Das führt zum Begriff der Imputation:

Definition 4.6.9. Ein Vektor $(z_1, \ldots, z_n)^T \in \mathbb{R}^n$ heißt *Imputation* oder *Zubilligungsvektor* zu einem n-Personenspiel (N, ν)), falls gilt:

1. „Individuelle Rationalität “
$$\forall\, i \in N \;:\; z_i \;\geq\; \nu(i).$$
2. „Kollektive Rationalität“
$$\sum_{i \in N} z_i \;=\; \nu(N).$$

Die Menge aller Imputationen heißt *Imputationsraum* (IR).

Definition 4.6.10. Eine Imputation z *dominiert* eine Imputation w bezüglich einer Koalition K, $K \neq \emptyset, K \neq N, K \subset N$, $|K| > 1$, falls gilt:

1. „Überlegenheit“
$$\forall i \in K \;:\; z_i \;>\; w_i.$$
2. „Zulässigkeit“
$$\sum_{i \in K} z_i \;\leq\; \nu(K).$$

Wir schreiben $z \overset{K}{\to} w$ und nennen K eine *effektive* Koalition für z.

Bemerkung 4.6.11. In Definition 4.6.10 ergäbe $|K| = 1$ keinen Sinn, denn dann wäre $\nu(1) \geq z_1 > w_1 \geq \nu(1)$. Der Fall $K = N$ kann wegen Definition 4.6.9.2 nicht eintreten.

Definition 4.6.12. Eine Imputation z ist *dominationsfähig* gegenüber w, falls es eine effektive Koalition K gibt mit $z \overset{K}{\to} w$. Wir schreiben dann $z \overset{f}{\to} w$.

Definition 4.6.13. Der *Kern* eines kooperativen n-Personenspiels ist

$$\left\{ z \in \text{IR} \mid \;\nexists w \in \text{IR} \text{ mit } w \overset{f}{\to} z \right\}.$$

Lemma 4.6.14. *Es sei (N, ν) ein n-Personenspiel und w eine Imputation. Dann gilt für alle $K \subset N$:*
$$\sum_{k \in K} w_k < \nu(K) \quad \Longleftrightarrow \quad \exists z \in IR \text{ mit } z \overset{K}{\to} w. \tag{4.7}$$

Beweis. Wir zeigen zunächst $\Leftarrow$. Es gelte $z \overset{K}{\to} w$. Da z eine Imputation ist, folgt sofort

$$\nu(K) \geq \sum_{k \in K} z_k \;>\; \sum_{k \in K} w_k$$

aus Zulässigkeit und Überlegenheit. Nun zu $\Rightarrow$. Dazu gelte

$$\sum_{k \in K} w_k < \nu(K).$$

Daraus definieren wir

$$d := \nu(K) - \sum_{k \in K} w_k \ > \ 0.$$

Weiter sei

$$z_k := \begin{cases} w_k + \frac{d}{|K|} & : \quad k \in K \\ \nu(k) + \frac{1}{|N|-|K|} \left(\nu(N) - \nu(K) - \sum_{j \in N \setminus K} \nu(j) \right) & : \quad k \in N \setminus K \end{cases}$$

Für $k \in K$ gilt dann $z_k > w_k \geq \nu(k)$, weil w eine Imputation ist. Die Superadditivität des Spiels liefert

$$\nu(K) + \sum_{j \in N \setminus K} \nu(j) \leq \nu(N).$$

Damit erhalten wir $z_k \geq \nu(k)$ für $k \in N \setminus K$. Das zeigt die individuelle Rationalität und Überlegenheit von z. Wegen

$$\sum_{k \in K} z_k = \sum_{k \in K} \left(w_k + \frac{d}{|K|} \right) = \sum_{k \in K} w_k \ + d = \nu(K)$$

ist z zulässig und

$$\begin{aligned} \sum_{k=1}^{n} z_k &= \sum_{k \in K} z_k \quad + \quad \sum_{k \in N \setminus K} z_k \\ &= \nu(K) + \sum_{k \in N \setminus K} \nu(k) + \nu(N) - \nu(K) - \sum_{j \in N \setminus K} \nu(j) \\ &= \nu(N) \end{aligned}$$

zeigt die kollektive Rationalität. Also ist K effektiv für z. □

Satz 4.6.15. *Für jedes kooperative n-Personenspiel (N, ν) gilt*

$$\text{Kern}((N,\nu)) = \left\{ z \in \mathbb{R}^n \mid \forall K \subset N \ : \ \sum_{k \in K} z_k \geq \nu(K) \ , \ \sum_{k=1}^{n} z_k = \nu(N) \right\}.$$

Beweis. Zunächst zeigen wir $\supset$. Es sei z ein Element der rechten Seite. Dann gilt $z_k \geq \nu(k)$, denn man kann die einelementigen Teilmenegen von N wählen. Zusammen mit der zweiten Bedingung erhält man $z \in I(v\nu)$. Angenommen es gibt eine Imputation w und eine Koalition K mit $w \xrightarrow{K} z$. Es folgt

$$\sum_{k \in K} w_k > \sum_{k \in K} z_k \geq \nu(K).$$

Also ist w unzulässig im Widerspruch zu $w \xrightarrow{K} z$. Daher gibt es so eine Imputation w nicht und es folgt $z \in I(\nu)$.

Nun zeigen wir $\subset$. Nun zeigen wir $\subset$. Dazu sei $z \in I(\nu)$ nicht dominierbar. Angenommen z ist nicht in der rechten Seite. Da z Imputation ist, kann

$$\sum_{k=1}^{n} z_k < \nu(N)$$

nicht gelten. Wir nehmen also an es gibt $K \subset N$ mit

$$\sum_{k \in K} z_k < \nu(K).$$

Dann ist aber z nach Lemma 4.6.14 dominierbar im Widerspruch zu $z \in \text{Kern}((N, \nu))$. □

Beweis. (ohne das Lemma). Zunächst zeigen wir $\supset$. Es sei z ein Element der rechten Seite. Dann gilt $z_k \geq \nu(k)$, denn man kann die einelementigen Teilmenegen von N wählen. Zusammen mit der zweiten Bedingung erhält man $z \in$IR. Angenommen es gibt eine Imputation w und eine Koalition K mit $w \xrightarrow{K} z$. Es folgt

$$\sum_{k \in K} w_k > \sum_{k \in K} z_k \geq \nu(K).$$

Also ist w unzulässig im Widerspruch zu $w \xrightarrow{K} z$. Daher gibt es so eine Imputation w nicht und es folgt $z \in I(\nu)$.
Nun zeigen wir $\subset$. Dazu sei $z \in$ IR nicht dominierbar. Angenommen z ist nicht in der rechten Seite. Da z Imputation ist, kann

$$\sum_{k=1}^{n} z_k < \nu(N)$$

nicht gelten. Wir nehmen also an es gibt $K \subset N$ mit

$$\sum_{k \in K} z_k < \nu(K).$$

Wir definieren

$$e := \nu(K) - \sum_{k \in K} z_k \quad > \ 0$$

und gemäß Subadditivität

$$d := \nu(N) - \nu(K) - \sum_{k \in N \setminus K} \nu(k) \quad \geq \ 0.$$

Weiter definieren wir

$$w_k := \begin{cases} z_k + \frac{e}{|K|} & : \quad k \in K \\ \nu(K) + \frac{d}{|N|-|K|} & : \quad k \in N \setminus K \end{cases}$$

Daraus ergibt sich $w_k > z_k \geq \nu(k)$ für alle $k \in K$ und $w_k \geq \nu(k)$ für alle $k \in N \setminus K$. Außerdem gilt

$$\begin{aligned} \sum_{k=1}^{n} w_k &= \sum_{k \in K} w_k + \sum_{k \in N \setminus K} w_k \\ &= \nu(K) + \sum_{k \in N \setminus K} \nu(k) + d \\ &= \nu(N). \end{aligned}$$

Also ist $w \in$ IR und $w \xrightarrow{K} z$ uns somit $z \notin \text{Kern}((N, \nu))$. Widerspruch! □

Der Kern kann leer sein. Wir betrachten folgendes

Beispiel 4.6.16. Drei Spieler wählen eine Zahl aus $\{0,1\}$. Die Spielermenge sei $N := \{1,2,3\}$. Die Auszahlungen $a_i(x,y,z)$ werden in folgender Tabelle festgelegt:

(x,y,z)	a_1	a_2	a_3	a_1+a_2
0,0,0	0	0	0	0
0,0,1	1	1	-2	2
0,1,0	1	-2	1	-1
1,0,0	-2	1	1	-1
1,1,0	1	1	-2	2
0,1,1	-2	1	1	-1
1,0,1	1	-2	1	-1
1,1,1	0	0	0	0

Die Summe der Auszahlungen bei jeder Wahl ist jeweils Null. Koalieren Spieler 1 und 2, so erhalten sie die Auszahlung $a_1 + a_2$. Es sei $K = \{1,2\}$. Dann wird diese Koalition K die Strategien 0,0 oder 1,1 spielen. Dies ergibt ein 2-PNSS zwischen der Koalition K und Spieler 3 mit Auszahlungsmatrix

$K \mid 3$	0	1
0,0	0	2
1,1	2	0

.

Die charakteristische Funktion ν wollen wir als den zu erwartenden Gewinn festlegen. Spielt die Koaltion K mit $(p, 1-p)$ und Spieler 3 mit $(q, 1-q)$ so ergibt sich als erwartete Auszahlung für K:

$$f(p,q) \;:=\; (p,1-p)\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} q \\ 1-q \end{pmatrix} \;=\; 2q(1-p)+2p(1-q).$$

Der Gradient von f verschwindet bei $p = q = \frac{1}{2}$. Dort ergibt sich die maximale Auszahlung

$$f(\tfrac{1}{2}, q) = 1$$

für K unabhängig von der Wahl von q. Wir legen daher $\nu(K) := 1$ fest. Da die Auszahlungen a_i sich zu Null summieren, definieren wir ν als Nullsummenspiel, also

$$\nu(N) := 0 \quad \text{und} \quad \forall S \subset N : \nu(S) + \nu(N \setminus S) = 0.$$

Wegen der Symmetrie in den Auszahlungen erhalten wir

$$\nu(\{2,3\}) = \nu(\{1,3\}) = 1.$$

Aus der Nullsummeneigenschaft ergibt sich

$$\nu(\{1\}) = \nu(\{2\}) = \nu(\{3\}) = -1.$$

Damit ist ν ein wohldefiniertes kooperatives 2 Personennullsummenspiel. Nun wenden wir uns dem Kern dieses Spiels zu. Nach Satz 4.6.15 müssen für ein Element (x_1, x_2, x_3) im Kern gelten:

$$x_1 + x_2 + x_3 = 0$$

$$x_1 \geq -1 \qquad x_2 \geq -1 \qquad x_3 \geq -1$$

$$x_1 + x_2 \geq 1 \qquad x_1 + x_3 \geq 1 \qquad x_3 + x_2 \geq 1$$

Addition der Ungleichungen ergibt

$$x_1 + x_2 + x_3 \geq \tfrac{2}{3}.$$

im Widerspruch zu $x_1 + x_2 + x_3 = 0$. Also ist der Kern leer.

Leider ist der Kern für eine ganze Klasse von Spielen leer.

Satz 4.6.17. *Für wesentliche Konstantsummenspiele ist der Kern leer.*

Beweis. Es sei z eine Imputation aus dem Kern. Dann gilt für jedes $i \in N$:

$$\sum_{k \in N \setminus \{i\}} z_k \geq \nu(N \setminus \{i\}) = \nu(N) - \nu(i).$$

Angenommen es gilt gilt $z_i > \nu(i)$. Dann folgt

$$\nu(N) = \sum_{k=1}^{n} z_k > \nu(i) + \sum_{k \in N \setminus \{i\}} z_k \geq \nu(N).$$

Daher gilt $\nu(k) = z_k$ für jedes k. Damit berechnen wir

$$\sum_{k \in N} z_k = \sum_{k \in N} \nu(k) < \nu(N).$$

Da das Spiel wesentlich ist, gilt die letzte Ungleichung. Dies ist aber ein Widerspruch, also ist der Kern leer. □

4.7 Der Shapley-Wert

4.7.1 Shapley's Funktion über Axiome

Es sei (N, ν) ein n-Personenspiel und $\sigma \in S_n$ eine Permutation. Zu einem Spieler $i \in N$ ist die Menge der *Vorgänger* definiert durch

$$P_\sigma(i) := \left\{ r \in \mathbb{N} \,|\, \sigma^{-1}(r) < \sigma^{-1}(i) \right\}.$$

Den *Anteil der Auszahlung* von Spieler i bezeichnen wir mit

$$m_i^\sigma := \nu\left(P_\sigma(i) \cup \{i\}\right) - \nu\left(P_\sigma(i)\right).$$

Den entsprechenden Auszahungsvektor schreiben wir als

$$m_\sigma := (m_1^\sigma, \ldots, m_n^\sigma).$$

Beispiel 4.7.1. Es sei $N := \{1, 2, 3, 4, 5\}$ und $\sigma \in S_5$ gegeben durch

$$\sigma := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{pmatrix}.$$

Dann hat man $P_\sigma(3) = \{4, 2, 1\}$.

Definition 4.7.2. Der *Shapley-Wert* eines n-Personenspiels (N, ν) ist definiert durch

$$\Phi((N,\nu)) := \frac{1}{n!} \sum_{\sigma \in S_n} m_\sigma.$$

Beispiel 4.7.3.

1. Für $N = \{1, 2\}$ ergibt sich

$$\begin{aligned} \Phi((N,\nu)) &= \frac{1}{2!} \sum_{\sigma \in S_2} m_\sigma = \tfrac{1}{2}(m_{id} + m_\tau) \\ &= \tfrac{1}{2}\left(\nu(N) + \nu(1) - \nu(2), \nu(N) - \nu(1) + \nu(2)\right) \end{aligned}$$

2. Für ein additives Spiel gilt $m_\sigma(i) = \nu(i)$ für jedes $i \in N$. Daher folgt

$$\Phi((N,\nu)) = (\nu(1), \ldots, \nu(n)).$$

Bemerkung 4.7.4. Für jede Komponente des Shapley-Wertes gilt definitionsgemäß

$$\Phi_i((N,\nu)) = \frac{1}{n!} \sum_{\sigma \in S_n} \nu\left(P_\sigma(i) \cup \{i\}\right) - \nu\left(P_\sigma(i)\right). \tag{4.8}$$

Wir erhalten daraus den folgenden

Satz 4.7.5. *Für die Komponenten des Shapley-Wertes eines n-Personenspiels (N, ν) gilt*

$$\Phi_i((N,\nu)) = \sum_{S \subset N, i \notin S} \frac{|S|!(n-1-|S|)!}{n!} \left(\nu(S \cup \{i\}) - \nu(S)\right),\ 1 \le i \le n.$$

Beweis. Wir betrachten in (4.8) eine Permutation

$$\sigma := \begin{pmatrix} 1 & 2 & 3 & \cdots & \sigma^{-1}(i) & \cdots & n \\ j_1 & j_2 & j_3 & \cdots & i & \cdots & j_n \end{pmatrix}.$$

Einerseits liefert jede Permutation ein geordnetes Tupel

$$(j_1, \ldots, j_k, i, j_{k+2}, \ldots, j_n), \quad 0 \le k \le n-1$$

andererseits kommt jedes Tupel dieser Art mit paarweise verschiedenen Komponenten als Bild einer Permutation vor. Umordnungen der Komponenten vor i beziehnungsweise nach i ergeben unter ν denselben Wert, da die Vorgängermenge $P_\sigma(i) = \{j_1, \ldots, j_k\}$ und $N \setminus (P_\sigma(i) \cup \{i\})$ ungeordnet sind. Für jede Menge $S \subset N$, die als Vorgängermenge auftaucht, gibt es daher $|S|!(n-1-|S|)!$ Permutationen, die denselben Wert unter ν liefern. □

Bemerkung 4.7.6. Mit Hilfe von (4.8) können wir eine stochastische Interpretation des Shapley-Wertes formulieren. Betrachte die Permutationen S_n als eine Urne, aus der eine Kugel gezogen wird. Die Spieler treten in der Reihenfolge $(\sigma(1), \ldots, \sigma(n))$ in einen Raum ein. Jedem Spieler wird bei Eintritt sein Anteil an seiner Koalition mit den bereits Eingetretenen ausgezahlt. Damit ist Φ_i die erwartete Auszahlung an Spieler i.

Definition 4.7.7. Es seien (N, ν) und (N, μ) jeweils n-Personenspiele.

(i) Eine Spieler i heißt *Nullspieler*, falls

$$\nu(S \cup \{i\}) = \nu(S)$$

für alle $S \subset N$ gilt.

(ii) Zwei verschiedene Spieler $i, j \in N$ heißen *symmetrisch*, falls

$$\nu(S \cup \{i\}) = \nu(S \cup \{j\})$$

für alle $S \subset N \setminus \{i, j\}$ gilt.

(iii) Die *Summe* $(N, \nu + \mu)$ der Spiele (N, ν) und (N, μ) ist definiert durch

$$(\nu + \mu)(S) := \nu(S) + \mu(S).$$

(iv) Für $\lambda > 0$ ist das n-Personenspiel $(N, \lambda\nu)$ definiert durch

$$(\lambda\nu)(S) := \lambda\nu(S).$$

Zu $N = \{1, \ldots, n\}$ sei $\mathcal{P}_N$ die Menge aller n-Personenspiele. Ein Element dieser Menge schreiben wir als $\nu \in \mathcal{P}_N$.

Definition 4.7.8. Eine Abbildung $\Phi : \mathcal{P}_N \to \mathbb{R}^n$ heißt *Shapley-Funktion*, wenn sie die folgenden Axiome 4.7.9 erfüllt. Der *Shapley-Wert* eines Spiels ist das Bild eines Spiels unter Φ. Der (Shapley-)Wert eines Spielers i ist die i-te Komponente des Shapley-Wertes des Spiels.

Axiome 4.7.9.

1. „Effizienz“
$$\forall \nu \in \mathcal{P}_N \ : \ \sum_{i=1}^{n} \Phi_i(\nu) = \nu(N).$$

2. „Nullspieler-Eigenschaft“
Für jeden Nullspieler $i \in N$ gilt:
$$\forall \nu \in \mathcal{P}_N \ : \ \Phi_i(\nu) = 0.$$

3. „Symmetrie“
Für je zwei symmetrische Spieler i, j gilt:
$$\forall \nu \in \mathcal{P}_N \ : \ \Phi_i(\nu) = \Phi_j(\nu).$$

4. „Additivität“
$$\forall \nu, \mu \in \mathcal{P}_N \ : \ \Phi(\nu + \mu) = \Phi(\nu) + \Phi(\mu).$$

Bemerkung 4.7.10. Ein Nullspieler trägt zu keiner Koalition etwas bei, daher wird ihm der Wert Null zugeordnet. Symmetrische Spieler tragen zu jeder Koalition dasselbe bei, daher sollen sie denselben Wert erhalten. Werden zwei Spiele hintereinander gespielt, so werden Spieler i die Werte $\Phi_i(\nu)$ und $\Phi_i(\mu)$ zugeordnet. Betrachtet man beide Spiele als nur ein Spiel, ergibt sich das Axiom zur Additivität.

Satz 4.7.11. *Der Shapley-Wert aus Definition 4.7.2 erfüllt die Shapley-Axiome 4.7.9.*

Beweis. Zunächst zur Effizienz. Es sei

$$\sigma := \begin{pmatrix} 1 & 2 & \cdots & n \\ j_1, & j_2 & \cdots & j_n \end{pmatrix} \in S_n$$

eine Permutation. Wegen

$$P_\sigma(j_k) = \{j_1, \ldots, j_{k-1}\} = P_\sigma(j_{k-1}) \cup \{j_{k-1}\}$$

gilt

$$\sum_{i=1}^{n} m_i^\sigma = \sum_{k=1}^{n} m_{j_k}^\sigma = \nu(N) - \nu(\emptyset) = \nu(N),$$

denn die zweite Summe ist eine Teleskopsumme. Damit rechnen wir

$$\begin{aligned} \sum_{i=1}^{n} \Phi_i(\nu) &= \frac{1}{n!} \sum_{i=1}^{n} \sum_{\sigma \in S_n} m_i^\sigma \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \sum_{i=1}^{n} m_i^\sigma \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} \nu(N) = \nu(N) \end{aligned}$$

Zur Nullspielereigenschaft sei $i \in N$ ein solcher. Aus Satz 4.7.5 folgt sofort $\Phi_i(\nu) = 0$. Zur Symmetrie: es seien i, j symmetrische Spiler. Wir betrachten eine Menge $S \subset N$ mit $i \notin S$ und $j \in S$. Dann gilt

$$\nu(S) = \nu((S \setminus \{j\}) \cup \{j\}) = \nu((S \setminus \{j\}) \cup \{i\}) = \nu((S \cup \{i\}) \setminus \{j\}). \tag{4.9}$$

Wir definieren nun Mengen

$$\begin{aligned} \mathcal{S}_{ij} &:= \{S \subset N \mid i \notin S, j \in S\} \\ \mathcal{M}_{ij} &:= \{M \subset N \mid i \in M, j \in M\} \end{aligned}$$

und Abbildungen

$$\begin{aligned} \Gamma &: \mathcal{S}_{ij} \to \mathcal{M}_{ij}, \quad S \mapsto S \cup \{i\} \\ \Delta &: \mathcal{S}_{ji} \to \mathcal{M}_{ij}, \quad S \mapsto S \cup \{j\} \end{aligned}$$

Aus der Definition ergibt sich sofort, dass Γ und Δ bijektiv sind. Zur Abkürzung sei

$f(|S|) := |S|!(n-1-|S|)!$. Wir rechnen mit Satz 4.7.5:

$$\begin{aligned}
\Phi_i(\nu) &\overset{(4.9)}{=} \sum_{i\notin S, j\notin S} f(|S|)(\nu(S\cup\{i\})-\nu(S)) + \\
&\qquad \sum_{i\notin S, j\in S} f(|S|)(\nu(S\cup\{i\})-\nu((S\cup\{i\})\setminus\{j\})) \\
&= \sum_{i\notin S, j\notin S} f(|S|)(\nu(S\cup\{i\}-\nu(S)) + \\
&\qquad \sum_{S\in\mathcal{S}_{ij}} f(|S|)(\nu(\Gamma(S))-\nu(\Gamma(S)\setminus\{j\})) \\
&\overset{\Gamma(S)=M}{=} \sum_{i\notin S, j\notin S} f(|S|)(\nu(S\cup\{i\}-\nu(S)) + \\
&\qquad \sum_{M\in\mathcal{M}_{ij}} f(|\Gamma^{-1}(M)|)(\nu(M)-\nu(M\setminus\{j\})) \\
&\overset{\Delta(T)=M}{=} \sum_{i\notin S, j\notin S} f(|S|)(\nu(S\cup\{i\}-\nu(S)) + \\
&\qquad \sum_{T\in\mathcal{S}_{ji}} f(|\Gamma^{-1}\Delta(T)|)(\nu(T\cup\{j\})-\nu(T)) \\
&= \Phi_j(\nu)
\end{aligned}$$

Es bleibt noch die Additivität zu zeigen. Aus Satz 4.7.5 und der Definition der Summe zweier Spiele folgt für jede Komponente sofort $\Phi_i(\nu+\mu) = \Phi_i(\nu) + \Phi_i(\mu)$. □

Jedes Spiel läßt sich als Linearkombination von Spielen ausdrücken, die eine gewisse Einstimmigkeit beschreiben. Dazu definieren wir zu $T \subset N$ das Spiel $u_T \in \mathcal{P}_N$ durch

$$u_T(S) := \begin{cases} 0 & : \quad T \not\subset S \\ 1 & : \quad T \subset S \end{cases}.$$

Lemma 4.7.12. *Zu jedem $\nu \in \mathcal{P}_N$ gibt es Zahlen $c_T \in \mathbb{R}$, derart dass*

$$\nu = \sum_{T\subset N, T\neq\emptyset} c_T u_T$$

gilt.

Beweis. Im folgenden sei stets $t = |T|$, $s = |S|$ und $u = |U|$. Wir definieren

$$c_T := \sum_{S\subset T} (-1)^{t-s}\nu(S)$$

und werden zeigen, dass für jedes $U \subset N$

$$\nu(U) = \sum_{T\subset N} c_T u_T(U)$$

gilt. Wir rechnen

$$\sum_{T\subset N} c_T u_T(U) = \sum_{T\subset U} c_T u_T(U) \tag{4.10}$$

$$= \sum_{T\subset U}\sum_{S\subset T} (-1)^{t-s}\nu(S) \tag{4.11}$$

$$= \sum_{S\subset U}\left(\sum_{S\subset T\subset U} (-1)^{t-s}\right)\nu(S) \tag{4.12}$$

Die erste Gleichung (4.10) gilt nach Definition von u_T, in (4.12) tauschen wir die Summationsreihenfolge, denn beide Seiten summieren über alle Paare (S,T) mit $S\subset T\subset U$. Die innnere Summe in (4.12) läuft über Mengen T mit jeweils $t=|T|$ Elementen. Wegen $S\subset T\subset U$ gibt es

$$\binom{u-s}{t-s}$$

Möglichkeiten solche Mengen T mit $T\setminus S\subset U\setminus S$ auszuwählen. Daher gilt

$$\begin{aligned}
\sum_{S\subset T\subset U}(-1)^{t-s} &= \sum_{t=s}^{u}\sum_{T:S\subset T\subset U,t=|T|}(-1)^{t-s}\\
&= \sum_{t=s}^{u}(-1)^{t-s}\sum_{T:S\subset T\subset U,t=|T|}1\\
&= \sum_{t=s}^{u}(-1)^{t-s}\binom{u-s}{t-s}\\
&= \sum_{t=0}^{u-s}(-1)^t 1^{u-s-t}\binom{u-s}{t} = (1-1)^{u-s}
\end{aligned} \tag{4.13}$$

Nun ist (4.13) nur für $u=s$ nicht Null. Einsetzen in (4.12) ergibt dann die Behauptung. □

Lemma 4.7.13. *Es seien $\nu\in\mathcal{P}_N$, $c_T\in\mathbb{R}$ die Zahlen aus Lemma 4.7.12 und Ψ eine Shapley-Funktion. Dann gilt*

$$\Psi(\nu) = \sum_{c_T>0}\Psi(c_T u_T) - \sum_{c_T<0}\Psi(|c_T|u_T). \tag{4.14}$$

Beweis. Wir betrachten Spiele $\mu_1,\mu_2\in\mathcal{P}_N$, für die $\mu_1-\mu_2\in\mathcal{P}_N$ gilt. Aus der Additivität von Ψ ergibt sich

$$\Psi(\mu_1) = \Psi(\mu_2+\mu_1-\mu_2) = \Psi(\mu_2)+\Psi(\mu_1-\mu_2).$$

Also folgt

$$\Psi(\mu_1)-\Psi(\mu_2) = \Psi(\mu_1-\mu_2). \tag{4.15}$$

Mit Lemma 4.7.12 schreiben wir

$$\nu = \sum_{T\subset N,T\neq\emptyset} c_T u_T = \sum_{c_T>0} c_T u_T - \sum_{c_T<0}|c_T|u_T.$$

Hier sind $c_T u_T$ und $|c_T|u_T$ jeweils Spiele und somit auch deren Summen. Die Behauptung folgt damit aus (4.15). □

Satz 4.7.14. *Der Shapley-Wert aus Definition 4.7.2 ist die einzige Funktion auf $\mathcal{P}_N$, die die Shapley-Axiome 4.7.9 erfüllt.*

Beweis. Es sei Ψ eine weitere Shapley-Funktion. Wegen Lemma 4.7.13 reicht es zu zeigen:

$$\forall c > 0,\ T \subset N, T \neq \emptyset\ :\ \Psi(cu_T) = \Phi(cu_T). \tag{4.16}$$

Es seien also $c > 0, T \subset N, T \neq \emptyset$. Sei nun $i \in N \setminus T$ und $S \subset N$. Dann gilt für $T \subset S$:

$$u_T(S \cup \{i\}) = 1 = u_T(S).$$

Für $t \not\subset S$ hingegen gilt

$$u_T(S \cup \{i\}) = 0 = u_T(S).$$

Daher erhält man für jedes $S \subset N$:

$$u_T(S \cup \{i\}) - u_T(S) = 0$$

also ist Spieler i ein Nullspieler in cu_T. Aus der Nullspielereigenschaft ergibt sich daher

$$\forall i \notin T\ :\ \Psi_i(cu_T) = \Phi_i(cu_T) = 0. \tag{4.17}$$

Seien nun $i \neq j$, $i, j \in T$ und $S \subset N \setminus \{i, j\}$. Damit ist $T \notin S$, $T \notin S \cup \{i\}$ und $T \notin S \cup \{j\}$. Also $cu_T(S \cup \{i\}) = 0 = cu_T(S \cup \{j\})$, das heißt die Spieler sind symmetrisch und das Symmetrie-Axiom ergibt

$$\forall i \neq j \in T\ :\ \Psi_i(cu_T) = \Psi_j(cu_T),\ \Phi_i(cu_T) = \Phi_j(cu_T). \tag{4.18}$$

Aus (4.17), (4.18) und der Effizienz folgt

$$c = (cu_T)(N) = \sum_{i=1}^{n} \Phi_i(cu_T) = \sum_{i \in T} \Phi_i(cu_T). \tag{4.19}$$

Die Gleichung (4.19) gilt auch für Ψ. Aus $T \neq \emptyset$ und (4.18) erhalten wir

$$\forall i \in T\ :\ \Phi_i(cu_T) = \frac{c}{|T|} = \Psi_i(cu_T) \tag{4.20}$$

Insgesamt folgt nun die Behauptung (4.16) aus (4.20) und (4.17). □

Beispiel 4.7.15. Wir betrachten vier Parteien $N = \{1, 2, 3, 4\}$ mit jeweils 5, 20, 25 und 50 Sitzen im Parlament. Es sei $\nu(S) \in \{0, 1\}$ und $\nu(S) = 1$ für alle Koalitionen, die mehr als 50% der Mandate haben. Unter den Gewinnkoalitionen, tragen für die Spieler 1,2 und 3 jeweils $\{3, 4\}, \{2, 4\}$ und $\{1, 4\}$ etwas zum Shapley-Wert bei. Daher gilt

$$\Phi_i(\nu) = \frac{1!(3-1)!}{4!} = \frac{1}{12} \approx 0.083,\ i \in \{1, 2, 3\}.$$

Die Gewinnkoalitionen für Spieler 4 sind

$$\{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 4\}, \{2, 3, 4\}, \{1, 3, 4\}, \{1, 2, 3, 4\}, \tag{4.21}$$

also

$$\Phi_4(\nu) = \frac{6 + 6 + 6}{24} = \frac{9}{12} = 0.75$$

Damit haben Spieler 1 und 4 mehr Macht als Mandate, bei den Spielern 2 und 3 ist es umgekehrt.

4.7.2 Shapley's Funktion über die Betafunktion

Zu jedem Spieler $i \in N$ eines Spiels (N, ν) sei $x_i \in [0,1]$ die generelle Bereitschaft (die Wahrscheinlichkeit) an Koalitionen mitzuwirken. Für $S \subset N$ ist dann

$$\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i)$$

die Wahrscheinlichkeit $\mathbb{P}(S)$, dass die Koalition S zustande kommt. Die Zahl

$$\mathbb{E}(\nu) = \sum_{S \subset N} \nu(S) \cdot \mathbb{P}(S) = \sum_{S \subset N} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S)$$

können wir daher als Erwartungswert des Spiels interpretieren. Uns interessiert hier die Abhängigkeit von den x_i, also definieren wir

$$f(x_1, \ldots, x_n) := \sum_{S \subset N} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S).$$

Die partiellen Ableitungen von f wollen wir $\partial_k f$ nennen. Wir erhalten folgenden

Satz 4.7.16. *Für jedes $k \in n$ gilt*

$$\Phi_k(\nu) = \int_0^1 (\partial_k f)(t, \ldots, t)\, dt.$$

Beweis. Zunächst berechnen wir die partiellen Ableitungen von f. Dazu schreiben wir für $k \in N$

$$f(x_1, \ldots, x_n) = \sum_{S \subset N} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S) \tag{4.22}$$

$$= \sum_{S \subset N, k \in S} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S) + \sum_{S \subset N, k \notin S} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S) \tag{4.23}$$

$$= \sum_{S \subset N, k \in S} \left(x_k \prod_{i \in S \setminus \{k\}} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S) \tag{4.24}$$

$$+ \sum_{S \subset N, k \notin S} \left((1 - x_k) \prod_{i \in S} x_i \prod_{i \in N \setminus (S \cup \{k\})} (1 - x_i) \right) \nu(S) \tag{4.25}$$

$$\tag{4.26}$$

Differenzieren nach k ergibt

$$\partial_k f(x_1, \ldots, x_n)$$

$$= \sum_{S \subset N, k \in S} \left(\prod_{i \in S \setminus \{k\}} x_i \prod_{i \in N \setminus S} (1 - x_i) \right) \nu(S) - \sum_{S \subset N, k \notin S} \left(\prod_{i \in S} x_i \prod_{i \in N \setminus (S \cup \{k\})} (1 - x_i) \right) \nu(S)$$

$$= \sum_{S \subset N, k \notin S} \left[\prod_{i \in S} x_i \prod_{i \in N \setminus (S \cup \{k\})} (1 - x_i) \right] (\nu(S \cup \{k\}) - \nu(S))$$

und

$$\partial_k f(t,\dots,t) = \sum_{S\subset N, k\notin S} t^{|S|}(1-t)^{n-|S|-1}\left(\nu(S\cup\{k\})-\nu(S)\right)$$

also

$$\int_0^1 \partial_k f(t,\dots,t)\,dt = \sum_{S\subset N, k\notin S}\left(\int_0^1 t^{|S|}(1-t)^{n-|S|-1}\right)\left(\nu(S\cup\{k\})-\nu(S)\right).$$

Die Formel für das Beta-Integral

$$\int_0^1 t^{|S|}(1-t)^{n-|S|-1} = \frac{|S|!(n-1-|S|!}{n!}$$

liefert durch Einsetzen den Shapley-Wert des k-ten Spielers . □

Beispiel 4.7.17. Wir betrachten noch einmal das Beispiel 4.7.15. Die Gewinnkoalitionen sind diejenigen in (4.21), daher berechnen wir

$$\begin{aligned} f(x_1,x_2,x_3,x_4) &= x_1x_4(1-x_2)(1-x_3)+x_2x_4(1-x_3)(1-x_1)\\ &\quad +x_3x_4(1-x_2)(1-x_1)+x_1x_2x_4(1-x_3)\\ &\quad +x_1x_3x_4(1-x_2)+x_2x_3x_4(1-x_1)\end{aligned}$$

Differenzieren nach x_1 und einsetzen von $(t,\dots,t)$ ergibt

$$\partial_1 f(t,\dots,t) = t(1-t)^2.$$

Daher gilt

$$\Phi_1(\nu) = \int_0^1 t(1-t)^2\,dt = \tfrac{1}{2}t^2-\tfrac{2}{3}t^3+\tfrac{1}{4}t^4\Big|_0^1 = \tfrac{1}{12}.$$

Kapitel 5

Nichtlineare Optimierung

5.1 Nichtlineare Optimierungsprobleme

Betrachten allgemeine Minimierungsprobleme, bei welchen eine Funktion $f : S \to \mathbb{R}$ auf einer Menge $S \subset \mathbb{R}^n$ gegeben ist, und wir ihr Minimum auf S suchen. Dabei nehmen wir an, S sei durch *Gleichungs-* und *Ungleichungsrestriktionen* beschrieben.

Definition 5.1.1. Sei $X \subset \mathbb{R}^n$ offen und seien $f,\ g_1, ..., g_m, h_1, ..., h_p : X \to \mathbb{R}$ gegebene Funktionen. Das Problem, f zu minimieren unter den Restriktionen

$$g_i \leq 0, \quad i = 1, ..., m,$$

und

$$h_j = 0, \quad j = 1, ..., p,$$

nennen wir (allgemeines) *nichtlineares Optimierungsproblem* (MP). Die Menge

$$S := \{x \in X : \ g(x) \leq 0, \ h(x) = 0\}$$

heisst *zulässiger Bereich*, hierbei ist

$$g := \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix} \quad \text{und} \quad h := \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}.$$

Soll f nur unter den Restriktionen

$$g_i \leq 0, \quad i = 1, ..., m,$$

minimiert werden, sagen wir, dass (MP) sei *vom Ungleichungstyp* (MP$\leq$), in diesem Fall ist

$$S := \{x \in X : \ g(x) \leq 0\}$$

der *zulässige Bereich*.

Bemerkung 5.1.2. Sind in (MP$\leq$) alle g_i affin-linear, dann ist S ein Polyeder.

Wir verwenden für (MP) auch die Kurznotation

$$\begin{aligned} f &\stackrel{!}{=} \min, \\ g &\leq 0, \\ h &= 0. \end{aligned}$$

Beispiel 5.1.3. Sei $n = 2$ und $X = \mathbb{R}^2$. Wir suchen das Minimum von

$$f(x, y) = (x - 3)^2 + (y - 2)^2$$

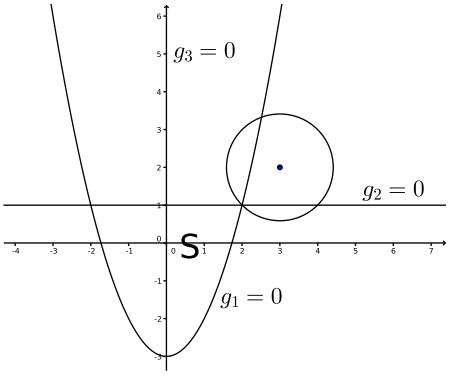
unter den Restriktionen

$$\begin{aligned} x^2 - y - 3 &\leq 0 \\ y - 1 &\leq 0 \\ -x &\leq 0. \end{aligned}$$

Hier sind also $g_1(x, y) = x^2 - y - 3$, $g_2(x, y) = y - 1$ und $g_3(x) = -x$, demnach also

$$S = \{(x, y) \in \mathbb{R}^2 : \ x^2 - 3 \leq y, \ y \leq 1, \ x \geq 0\}.$$

Die Funktion f ist das Quadrat des Abstandes zum Punkt $(3, 2)$.



Bemerkung 5.1.4. Wir machen ein paar heuristische Vorüberlegungen, dazu nehmen wir an, die betreffenden Funktionen seien in den betrachteten Punkten differenzierbar.

(i) Wie bei linearen Optimierungsproblemen ist $\nabla f(x) = 0$ nur für innere Punkte x von S eine notwendige Bedingung dafür, dass sie Minimalstellen sind. Minimalstellen sind aber oft Punkte des Randes ∂S von S.

(ii) Die Richtung des maximalen Abfallens von f ist gegeben durch $-\nabla f$. Ist $x \in \partial S$ eine Minimalstelle für f auf S, so muss $-\nabla f(x)$ 'nach aussen zeigen', denn hätte $-\nabla f(x)$ einen positiven Anteil in eine Richtung 'in S hinein' oder 'entlang ∂S', so könnte man in diese Richtung gehen und einen kleineren Wert von f finden, was der Minimalität widerspricht.

(iii) Sei Randpunkt x gegeben durch *eine* straffe Nebenbedingung $g_{i_0}(x) = 0$. Ist x Minimalstelle, dann zeigen $\nabla g_{i_0}(x)$ und $-\nabla f(x)$ in dieselbe Richtung, denn: Der Gradient $\nabla g_{i_0}(x)$ einer Funktion g_{i_0} steht senkrecht auf ihrer Niveaumenge, welche x enthält. Dass $\nabla g_{i_0}(x)$ dann nur nach aussen zeigen kann, folgt aus der Definition von S. Normalisieren wir $\nabla g_{i_0}(x)$, ergibt sich gerade die äussere Normale in x an S. Wegen (ii) muss dann auch $-\nabla f(x)$ in Richtung dieser äusseren Normalen zeigen. Sei nun x ein Randpunkt, der zwei straffe Nebenbedingungen $g_{i_1}(x) = g_{i_2}(x) = 0$

erfüllt. Ist x Minimalstelle, dann muss $-\nabla f(x)$ in dem konvexen Kegel liegen, der von $\nabla g_{i_1}(x)$ und $\nabla g_{i_2}(x)$ aufgespannt wird (dem sogenannten 'Normalenkegel' in x), und die Begründungen sind analog wie zuvor.

V25, 13.1.25

Definition 5.1.5. Sei $\emptyset \neq S \subset \mathbb{R}^n$ und $x \in S$. Ein Vektor $d \in \mathbb{R}^n$ heisst *zulässige Richtung* (bzgl. S) in x, falls es ein $\delta > 0$ gibt, sodass

$$[x, x + \delta d] \subset S.$$

Die Menge

$$D_S(x) := \{d : \ d \text{ zulässige Richtung bzgl. } S \text{ in } x\}$$

heisst der *Kegel der zulässigen Richtungen* in x.

Bemerkung 5.1.6.

(i) Offensichtlich ist $D_S(x)$ ein Kegel: Ist $d \in D_S(x)$ und $\lambda \geq 0$, dann ist $\lambda d \in D_S(x)$.

(ii) Ist $x \in S$ ein innerer Punkt von S, so gilt $D_S(x) = \mathbb{R}^n$.

Definition 5.1.7. Sei $X \in \mathbb{R}^n$ offen, $x \in X$ und $f : X \to \mathbb{R}$ differenzierbar in x. Ein Vektor $d \in \mathbb{R}^n$ heisst *Abstiegrichtung* für f in x, falls es ein $\delta > 0$ gibt mit

$$f(x + \lambda d) < f(x) \quad \text{für alle } \lambda \in (0, \delta).$$

Lemma 5.1.8. *Sei $X \subset \mathbb{R}^n$ offen, $x \in X$ und $f : X \to \mathbb{R}$ differenzierbar in x. Erfüllt $d \in \mathbb{R}^n$ die Ungleichung*

$$\nabla f(x) \cdot d < 0,$$

so ist d eine Abstiegsrichtung für f in x.

Beweis. Die Aussage folgt aus der Definition der Richtungsableitung von f in x in Richtung d: Man hat

$$0 > \nabla f(x) \cdot d = \lim_{\lambda \to 0} \frac{1}{\lambda}(f(x + \lambda d) - f(x)),$$

also für hinreichend kleine λ notwendigerweise $f(x + \lambda d) - f(x) < 0$. □

Definition 5.1.9. Ist $X \subset \mathbb{R}^n$ offen und $f : X \to \mathbb{R}$ differenzierbar in $x \in X$, so bezeichnen wir mit

$$F_f(x) := \{d \in \mathbb{R}^n : \ \nabla f(x) \cdot d < 0\}$$

den *Kegel der Abstiegsrichtungen* von f in x.

Bemerkung 5.1.10. Strenggenommen ist $F_f(x)$ selbst noch kein Kegel, sondern $F_f(x) \cup \{0\}$.

Ist $x \in S \subset X$ und besitzt f in x ein lokales Minimum in S, und ist $d \in D_S(x)$ eine zulässige Richtung in x, so folgt $x + \lambda d \in S$ für kleine $\lambda > 0$, und man hat

$$f(x + \lambda d) \geq f(x).$$

Somit kann d keine Abstiegsrichtung für f in x sein, also (nach Lemma 5.1.8) $\nabla f(x) \cdot d \geq 0$, also $d \notin F_f(x)$. Wir haben somit gezeigt, dass in einem lokalen Minimum (in S) keine zulässige Richtung eine Abstiegsrichtung sein kann:

Proposition 5.1.11. *Sei $X \subset \mathbb{R}^n$ offen, $S \subset X$ und $f : X \to \mathbb{R}$. Besitzt f in $x \in S$ ein lokales Minimum in S und ist f in x differenzierbar, so gilt*

$$D_S(x) \cap F_f(x) = \emptyset.$$

Bemerkung 5.1.12. Proposition 5.1.11 macht insbesondere Bemerkung 5.1.4 (ii) präzise.

Wir betrachten nun (MP$\leq$),

$$\begin{aligned} f &\stackrel{!}{=} \min, \\ g_i &\leq 0, \quad i = 1, ..., m. \end{aligned}$$

Dann ist $D_S(x)$ selbst schlecht handhabbar, man erhält aber Aussagen zu $D_S(x)$ mittels der Restriktionsfunktionen g_i. Für $x \in S = \bigcap_{i=1}^m \{g_i \leq 0\}$ sei

$$I(x) := \{i \in \{1, ..., m\} : \ g_i(x) = 0\}$$

die Menge der Indizes der in x *straffen* Restriktionen.

Bemerkung 5.1.13. Im Spezialfall, dass alle g_i affin-linear sind, sammelt man hier die Indizes der Restriktionshyperebenen, in welchen x liegt.

Lemma 5.1.14. *Sei (MP$\leq$) gegeben und $x \in S$. Weiter sei*

$$g_i \text{ in } x \begin{cases} \text{differenzierbar, falls } i \in I(x) \\ \text{stetig, falls } i \notin I(x). \end{cases}$$

Dann gilt für

$$G_S(x) := \{d \in \mathbb{R}^n : \ \nabla g_i(x) \cdot d < 0 \text{ für alle } i \in I(x)\}$$

die Inklusion

$$G_S(x) \subset D_S(x).$$

Beweis. Sei $d \in G_S(x)$. Da X offen ist, gibt es ein $\delta > 0$ mit $x+\lambda d \in X$ für alle $\lambda \in (0, \delta)$.

Für $i \notin I(x)$ ist $g_i(x) < 0$, und da g_i stetig ist in x, gibt es ein $\delta' > 0$ sodass $g_i(x + \lambda d) < 0$ für alle $\lambda \in (0, \delta')$.

Für $i \in I(x)$ ist $\nabla g(x) \cdot d < 0$, also gibt es ein $\delta'' > 0$ sodass $g_i(x + \lambda d) < g_i(x) = 0$ für alle $\lambda \in (0, \delta'')$.

Also ist $x + \lambda d \in S = \bigcap_{i=1}^m \{g_i \leq 0\}$ für alle $0 < \lambda < \min\{\delta, \delta', \delta''\}$, und das bedeutet, dass $d \in D_S(x)$ ist. □

Zusammen mit Proposition 5.1.11 ergibt sich sofort folgendes Resultat:

Satz 5.1.15. *Gegeben sei (MP≤) und $x \in S$. Ferner sei*

$$g_i \text{ in } x \begin{cases} \text{differenzierbar, falls } i \in I(x) \\ \text{stetig, falls } i \notin I(x). \end{cases}$$

Falls f in x ein lokales Minimum in S hat, so ist

$$G_S(x) \cap F_f(x) = \emptyset.$$

Diese Bedingung lässt sich in eine algebraische Aussage umformen, man nennt sie die *Fritz-John-Bedingung*:

Satz 5.1.16. *Seien in (MP≤) mit $x \in S$ die Funktionen g_i wie in Satz 5.1.15. Sei f differenzierbar in x und sei x eine lokale Minimalstelle für f in S. Dann gibt es Konstanten $\mu_0 \geq 0$ und $\mu_i \geq 0$, $i \in I(x)$, die nicht sämtlich null sind, sodass die folgende* Fritz-John-Bedingung *gilt:*

$$\mu_0 \nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0. \tag{FJB}$$

Bemerkung 5.1.17. Ist $\mu_0 > 0$, so sagt die (FJB), dass $-\nabla f(x)$ in dem konvexen Kegel liegt, der von den Gradienten $\nabla g_i(x)$ mit $i \in I(x)$ aufgespannt wird. Das macht insbesondere Bemerkung 5.1.4 (iii) präzise.

Für den Beweis des Satzes nutzen wir folgenden Alternativsatz von *Gordan*:

Satz 5.1.18. *Sei A eine reelle $(m \times n)$-Matrix. Dann gilt genau eine der folgenden Alternativen:*

(i) Es gibt ein $x \in \mathbb{R}^n$ mit $Ax > 0$.

(ii) Es gibt ein $y \in \mathbb{R}^m$ mit $A^T y = 0$, $y \geq 0$ und $y \neq 0$.

Beweis. (i) ist äquivalent dazu, dass es ein $\delta > 0$ gibt mit

$$Ax \geq \delta \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{=:j \in \mathbb{R}^m}.$$

Betrachten nun die $(m \times n+1)$-Matrix und die Vektoren

$$\overline{A} := (-A, j), \quad \overline{b} := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad \overline{x} := \begin{pmatrix} x \\ \delta \end{pmatrix}.$$

Dann ist (i) äquivalent dazu, dass es ein $\overline{x} \in \mathbb{R}^{n+1}$ gibt mit $\overline{A}\overline{x} \leq 0$ und $\overline{b}^T\overline{x} > 0$. Nach Satz 2.7.5 (Minkowski-Farkas) ist das die Alternative dazu, dass es ein $y \in \mathbb{R}^m$ gibt mit $\overline{A}^T y = \overline{b}$ und $y \geq 0$. Nun ist

$$\overline{A}^T y = \begin{pmatrix} -A^T \\ j^T \end{pmatrix} \cdot y = \begin{pmatrix} -A^T y \\ j^T y \end{pmatrix} = \begin{pmatrix} -A^T y \\ y_1 + ... + y_m \end{pmatrix},$$

also, wegen der Gestalt von $\overline{b}$,

$$A^T y = 0, \quad y \geq 0,$$

und da $y_1 + ... + y_m = 1$ ist, folgt auch $y \neq 0$. Das ist aber gerade (ii). □

Wir beweisen Satz 5.1.16.

Beweis. Nach Satz 5.1.15 ist $G_S(x) \cap F_f(x) = \emptyset$. Dann gibt es also kein $d \in \mathbb{R}^n$ sodass

$$\nabla f(x) \cdot d < 0 \quad \text{und} \quad \nabla g_i(x) \cdot d < 0, \quad i \in I(x).$$

Sei A die Matrix mit den Zeilen $\nabla f(x)$ und $\nabla g_i(x)$, $i \in I(x)$. Dann kann es keine Lösung $d \in \mathbb{R}^n$ geben zu

$$Ad < 0 \qquad (\Leftrightarrow -Ad > 0),$$

also folgt mit Satz 5.1.18, dass ein

$$z = \begin{pmatrix} z_0 \\ (z_i)_{i \in I(x)} \end{pmatrix} \quad \text{existiert mit } A^T z = 0,\ z \geq 0 \text{ und } z \neq 0.$$

Also erfüllen $\mu_0 := z_0$, $\mu_i := z_i$, $i \in I(x)$, die Bedingung

$$\mu_0 \nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0,$$

wobei $\mu_0, \mu_i \geq 0$ sind und nicht alle null. □

Sind alle g_i, $i = 1, ..., m$ in x differenzierbar (also nicht nur jene mit $i \in I(x)$ und die anderen nur stetig), so kann man FJB auch wie folgt umformulieren:

Korollar 5.1.19. *Sei (MP≤) gegeben und $x \in S$ eine lokale Minimalstelle für f in S. Sind f und alle g_i, $i = 1, ..., m$, differenzierbar in x, so gibt es $\mu_0, \mu_1, ..., \mu_m \geq 0$, die nicht alle gleich null sind, mit*

$$\mu_0 \nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) = 0 \quad \textit{und} \quad \mu_i g_i(x) = 0, \quad i = 1, ..., m.$$

Beweis. Für $i \notin I(x)$ setzen wir $\mu_i = 0$. (Da für solche i nach Definition $g_i(x) \neq 0$ ist, ist das sogar die einzige Möglichkeit, die zweite Bedingung nicht zu verletzen.) Diese Summanden entfallen also in der ersten Bedingung. Dass die zweite Bedingung für alle $i \in I(x)$ gilt, ist klar nach Definition. □

Bemerkung 5.1.20.

(i) Die Fritz-John-Bedinung (FJB) ist also eine *notwendige* Bedingung dafür, dass in x ein *lokales* Minimum für f auf S vorliegt, und somit insbesondere dafür, dass ein globales Minimum für f auf S vorliegt. Für bestimmte (genauer: 'pseudokonvexe') f schauen wir uns später auch hinreichende Bedingungen dafür an, dass sich in x ein globales Minimum für f auf S ergibt.

(ii) Die skalaren Faktoren μ_i heissen auch *Lagrange-Multiplikatoren*, die Gleichungen $\mu_i g_i(x) = 0$ *Bedingungen des komplementären Schlupfes*.

(iii) Ist in Satz 5.1.16 in der lokalen Minimalstelle x keine der Bedingungen straff (gilt also $g_i(x) < 0$, $i = 1, ..., m$ und somit $I(x) = \emptyset$), so liegt x im Inneren von S, und (FJB) reduziert sich wie zu erwarten auf

$$\nabla f(x) = 0.$$

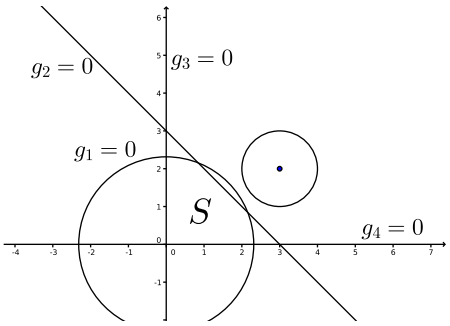
In diesem Fall ist x sogar lokale Minimalstelle für f auf einer Umgebung von S.

Oft kann man schon allein mit (FJB) eine Minimalstelle identifizieren.

Beispiel 5.1.21. Sei $n = 2$, $X = \mathbb{R}^2$. Wir wollen $(x, y) \mapsto (x-3)^2 + (y-2)^2$ minimieren unter

$$\begin{aligned} x^2 + y^2 &\leq 5 \\ x + y &\leq 3 \\ x, y &\geq 0. \end{aligned}$$

Also haben wir wieder $f(x, y) = (x-3)^2 + (y-2)^2$, nun mit



$$\begin{aligned} g_1(x, y) &= x^2 + y^2 - 5, \\ g_2(x, y) &= x + y - 3, \\ g_3(x, y) &= -x, \\ g_4(x, y) &= -y. \end{aligned}$$

Wie zuvor ist $S = \bigcap_{i=1}^4 \{g_i \leq 0\}$. Wir bemerken, dass in diesem Beispiel maximal zwei der g_i gleichzeitig straff sind in einem Punkt $(x, y) \in S$, und somit muss $I(x, y)$ eine der folgenden Mengen sein:

$$\emptyset,\ \{3\}, \{4\}, \{3, 4\}, \{1, 3\}, \{1, 4\}, \{2\}, \{1\}, \{1, 2\}.$$

Die Gradienten sind

$$\begin{aligned}\nabla f(x,y) &= (2(x-3), 2(y-2)),\\ \nabla g_1(x,y) &= (2x, 2y),\\ \nabla g_2(x,y) &= (1,1),\\ \nabla g_3(x,y) &= (-1,0),\\ \nabla g_4(x,y) &= (0,1).\end{aligned}$$

Wir checken die verschiedenen Möglichkeiten:

- Fall $I(x,y) = \emptyset$: Dann würde (FJB) $\nabla f(x,y) = 0$ erzwingen, also $(x,y) = (3,2)$. Dieser Punkt liegt aber nicht in S, kann also keine Minimalstelle in S sein.

- Fall $I(x,y) = \{3\}$, also g_3 straff in (x,y), also $x = 0$: Dann wird die Gleichung in (FJB) zu
$$\mu_0 \begin{pmatrix} -6 \\ 2(y-2) \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$
und das liefert $-6\mu_0 - 3\mu_3 = 0$, daher $\mu_0 = \mu_3 = 0$, und (FJB) nicht erfüllt.

- Fall $I(x,y) = \{4\}$ ist analog.

- Fall $I(x,y) = \{3,4\}$, also g_3 and g_4 straff in (x,y): Dann $x = y = 0$, die Gleichung in (FJB) ergibt
$$\mu_0 \begin{pmatrix} -6 \\ -4 \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$
das impliziert $\mu_0 = \mu_3 = \mu_4 = 0$, also (FJB) nicht erfüllt.

- Fall $I(x,y) = \{1,3\}$, also g_1 and g_3 straff in (x,y): Dann $(x,y) = (0,\sqrt{5})$, und
$$\mu_0 \begin{pmatrix} -6 \\ 2(\sqrt{5}-2) \end{pmatrix} + \mu_1 \begin{pmatrix} 0 \\ 2\sqrt{5} \end{pmatrix} + \mu_3 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
führt auf $\mu_0 = \mu_3 = 0$ und folglich auch auf $\mu_1 = 0$, (FJB) also nicht erfüllt.

- Fall $I(x,y) = \{1,4\}$ ist analog.

- Fall $I(x,y) = \{1,2\}$: Dann insbesondere g_2 straff in (x,y) und somit $y = 3 - x$. Es folgt, dass $g_1(x,y) \leq 0$ genau dann gilt, wenn $g_1(x, 3-x) \leq 0$. Das ist äquivalent zu $x^2 + (3-x)^2 - 5 = 2x^2 - 6x + 4 = 2(x-1)(x-2) \leq 0$, und das impliziert, dass $x \in [1,2]$ sein muss (um $(x,y) \in S$ zu garantieren) und dass g_1 ebenfalls straff ist in (x,y) falls $x = 1$ oder $x = 2$ gilt.

  (a) Ist $(x,y) = (2,1)$, dann $\nabla f(x,y) = (-2,-2)$, $\nabla g_1(x,y) = (4,2)$, $\nabla g_2(x,y) = (1,1)$, und daraus folgt insbesondere, dass
  $$\nabla f(x,y) + 2\nabla g_2(x,y) = 0.$$
  Somit ist hier (FJB) erfüllt mit $\mu_0 = 1$, $\mu_1 = 0$, $\mu_2 = 2$.

(b) Ist $(x,y) = (1,2)$, dann $\nabla f(x,y) = (-4,0)$, $\nabla g_1(x,y) = (2,4)$, $\nabla g_2(x,y) = (1,1)$. Die Gleichung in (FJB) ergibt $\mu_1 = \mu_2 = 0$ und $-4\mu_0 + 4\mu_1 + \mu_2 = 0$, somit auch $\mu_0 = 0$, (FJB) ist also nicht erfüllt.

- Fall $I = \{2\}$, also nur g_2 straff in (x,y): Dann ist $y = 3 - x$ und $x \in (1,2)$, und Gleichung ergibt

$$\mu_0 \begin{pmatrix} 2(x-3) \\ 2(1-x) \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

  Man kann offensichtlich nicht $\mu_0 = 0$ haben. Wegen $x \neq 2$ ist $2(x-3) \neq 2(1-x)$, und somit funktioniert $\mu_0 > 0$ auch nicht. Also (FJB) nicht erfüllt.

- Fall $I = \{1\}$, also nur g_1 straff in (x,y): Dann $x^2 + y^2 = 5$, wobei $0 < x < \sqrt{5}$, $x \neq 1, 2$. Das ergibt

$$2\mu_0 \begin{pmatrix} (x-3) \\ (y-2) \end{pmatrix} + 2\mu_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  und wegen $(x,y) \neq (0,0)$ ist $\mu_0 \neq 0$, also o.E. $\mu_0 = 1$. Dann also

$$x - 3 + \mu_1 x = 0 \quad \text{und} \quad y - 2 + \mu_1 y = 0,$$

  d.h.

$$1 + \mu_1 = \frac{3}{x} \quad \text{und} \quad (1+\mu_1)y = (1+\mu_1)\sqrt{5-x^2} = 2.$$

  Einsetzen ergibt

$$\frac{3}{x}\sqrt{5-x^2} = 2, \quad \text{und nach Quadrieren} \quad 45 - 9x^2 = 4x^2,$$

  also $13x^2 = 45$. Das ergibt (wegen $x > 0$)

$$x = \sqrt{\frac{45}{13}} \approx 1,86 \quad \text{und} \quad y = \sqrt{5 - \frac{45}{13}} = \sqrt{\frac{20}{13}} \approx 1,24.$$

  Dann ist aber $x + y > 3$, also $(x,y) \notin S$, also kann (x,y) keine Minimalstelle in S sein.

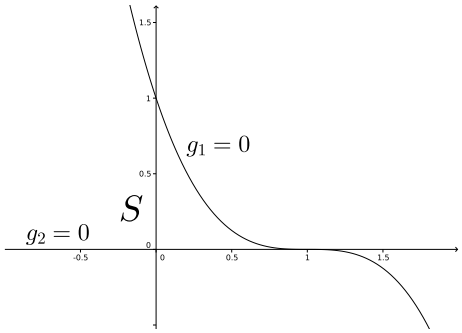
Insgesamt erfüllt also nur $(x,y) = (2,1)$ die Bedingung (FJB). Da es offensichtlich in jeder kleinen Umgebung von $(2,1)$ Punkte in S gibt, in denen f grösser ist, folgt, dass $(2,1)$ eine lokale Minimalstelle für f auf S sein muss. (Wie wir sehen, ist es im vorliegenden Beispiel sogar die einzige.) Das lokale Minimum, das sich ergibt, ist 2.

Beispiel 5.1.22. Sei $n = 2$, $X = \mathbb{R}^2$. Wir wollen $(x,y) \mapsto -x$ minimieren unter den Bedingungen

$$\begin{aligned} y - (1-x)^3 &\leq 0 \\ -y &\leq 0. \end{aligned}$$

Also $f(x,y) = -x$ und

$$\begin{aligned} g_1(x,y) &= y - (1-x)^3, \\ g_2(x,y) &= -y. \end{aligned}$$



Hier ist offensichtlich $w = (1, 0)$ eine Minimalstelle von f auf S, und man hat $f(w) = -1$. In w sind sowohl g_1 als auch g_2 straff. Des Weiteren hat man

$$\begin{aligned}\nabla f(x, y) &= (-1, 0),\\ \nabla g_1(x, y) &= (3(1 - x)^2, 1),\\ \nabla g_2(x, y) &= (0, -1),\end{aligned}$$

und insbesondere $\nabla g_1(w) = (0, 1)$ und $\nabla g_2(w) = (0, -1)$. Somit ist die Gleichung

$$\mu_0 f(w) + \mu_1 \nabla g_1(w) + \mu_2 \nabla g_2(w) = 0$$

in (FJB) hier für beliebige $\mu_1 = \mu_2 > 0$ erfüllt, falls $\mu_0 = 0$. Letzteres bedeutet aber, dass keine Information über f in die Gleichung in (FJB) eingeht, sie hat also bezüglich einer etwaigen Minimalität von f in w gar keine Aussagekraft.

Besser wäre es, man könnte hier irgendwie sicher stellen, dass $\mu_0 \neq 0$ ist und somit o.E. $\mu_0 = 1$. Man kann das erreichen, indem man zusätzliche Eigenschaften der Restriktionen fordert, sogenannte *constraint qualifications* (QC). Eine einfache Möglichkeit, das zu implementieren, ist eine *(notwendige) KKT-Bedingung* (nach *Karush, Kuhn und Tucker*).

Satz 5.1.23. *Im Problem (MP$\leq$) sei $x \in S$ ein Punkt, für den*

$$g_i \text{ in } x \begin{cases} \text{differenzierbar, falls } i \in I(x) \\ \text{stetig, falls } i \notin I(x). \end{cases}$$

und f in x differenzierbar ist. Weiter gelte

$$\text{Die Gradienten } \nabla g_i(x),\ i \in I(x), \text{ sind linear unabhängig.} \tag{LICQ}$$

Ist x ein lokales Minimum für f in S, so gibt es $\mu_i \geq 0$, $i \in I(x)$, die die folgende KKT-Bedingung erfüllen:

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0. \tag{KKT}$$

Falls alle g_i, $i = 1, ..., m$ in x differenzierbar sind, ist dies äquivalent zu

$$\nabla f(x) + \sum_{i=1}^{m} \mu_i \nabla g_i(x) = 0, \quad \mu_i \geq 0, \quad \mu_i g_i(x) = 0, \quad i = 1, ..., m.$$

Mit $\mu = (\mu_1, ..., \mu_m)^T$ und $g = (g_1, ..., g_m)^T$ kann man die letzte Identität im Satz auch wie folgt schreiben:

$$\nabla f(x) + \mu^T \cdot \nabla g(x) = 0, \quad \mu \geq 0, \quad \mu^T \cdot g(x) = 0.$$

Beweis. Unter den gegebenen Voraussetzungen is (FJB) erfüllt. Wäre dabei $\mu_0 = 0$, dann hätte man

$$\sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0$$

mit mindestens einem $\mu_i > 0$. Das widerspricht aber (LICQ). □

Das Kürzel LICQ steht für *linear independence constraint qualification*. Es gibt darüber hinau noch weitere (auch weniger strikte) CQs.

Definition 5.1.24. Sei $S \subset \mathbb{R}^n$. Ein Vektor $d \in \mathbb{R}^n$ heisst *Tangentialrichtung* für S in $x \in S$, wenn es Folgen $(x_k)_{k=1}^{\infty} \subset S$ und $(\lambda_k)_{k=1}^{\infty} \subset (0, +\infty)$ gibt, sodass

$$\lim_{k\to\infty} x_k = x, \quad \lim_{k\to\infty} \lambda_k = 0 \quad \text{und} \quad \lim_{k\to\infty} \frac{1}{\lambda_k}(x_k - x) = d.$$

Die Menge

$$T_S(x) := \{d \in \mathbb{R}^n : \ d \text{ Tangentialrichtung für } S \text{ in } x\}$$

aller Tangentialrichtungen für S in x heisst *Tangentialkegel für S in x*.

Bemerkung 5.1.25.

(i) Offensichtlich gilt $D_S(x) \subset T_S(x)$ für alle $x \in S$: Für $d \in D_S(x)$ folgt sofort mit $x_k := x + \lambda_k d \in S$, wobei $(\lambda_k)_k$ eine Nullfolge positiver Zahlen ist, dass $d = \lim_k \frac{1}{\lambda_k}(x_k - x)$.

(ii) Man kann zeigen, dass für konvexe S die Identität $\overline{D}_S(x) = T_S(x)$ gilt für alle $x \in S$.

Lemma 5.1.26. *Sei $X \subset \mathbb{R}^n$ offen, $S \subset X$ und $f : X \to \mathbb{R}$. Ist x ein lokales Minimum für f auf S und ist f differenzierbar in x, so gilt*

$$T_S(x) \cap F_f(x) = \emptyset.$$

Beweis. Sei $d \in T_S(x)$ und

$$d = \lim_{k\to\infty} \frac{1}{\lambda_k}(x_k - x) \quad \text{mit } \ (x_k)_k \subset S, \ \lim_{k\to\infty} x_k = x, \ (\lambda_k)_k \subset (0, +\infty), \ \lim_{k\to\infty} \lambda_k = 0.$$

Wegen der (totalen) Differenzierbarkeit von f in x gilt

$$f(x_k) - f(x) = \nabla f(x) \cdot (x_k - x) + \|x_k - x\| \, \varphi(x_k - x),$$

mit $\lim_{y\to 0} \varphi(y) = 0$.

Da x ein lokales Minimum für f auf S ist, gilt $f(x_k) - f(x) \geq 0$ für grosse k, also auch

$$\nabla f(x) \cdot (x_k - x) + \|x_k - x\| \, \varphi(x_k - x) \geq 0,$$

und Division durch die $\lambda_k > 0$ ergibt

$$\nabla f(x) \cdot \frac{1}{\lambda_k}(x_k - x) + \left\| \frac{1}{\lambda_k}(x_k - x) \right\| \varphi(x_k - x) \geq 0,$$

und für $k \to \infty$ gerade

$$\nabla f(x) \cdot d \geq 0,$$

also $d \notin F_f(x)$. □

Wir finden die folgende *(notwendige) KKT-Bedingung mit Abadie-CQ*:

Satz 5.1.27. *Sei in (MP$\leq$) $x \in S$, alle g_i, $i \in I(x)$ differenzierbar in x und*

$$T_S(x) = G'_S(x), \tag{ACQ}$$

wobei

$$G'_S(x) := \{d \in \mathbb{R}^n : \nabla g_i(x) \cdot d \leq 0 \ \ \textit{für alle } i \in I(x)\}.$$

Ist x ein lokales Minimum für f in S, so gibt es $\mu_i \geq 0$, $i \in I(x)$ mit

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0.$$

Bemerkung 5.1.28.

(i) Man sieht leicht, dass $T_S(x) \subset G'_S(x)$ gilt für alle $x \in S$. Es gilt also

$$G_S(x) \subset D_S(x) \subset T_S(x) \subset G'_S(x)$$

und daher

$$\overline{G_S(x)} \subset \overline{D_S(x)} \subset \underbrace{\overline{T_S(x)}}_{=T_S(x)} \subset \underbrace{\overline{G'_S(x)}}_{=G'_S(x)},$$

Folglich gilt in der Inklusionskette überall Gleichheit, falls $\overline{G_S(x)} = G'_S(x)$ *(Cottle-CQ)*. In diesem Falle gilt auch (ACQ).

(ii) Man kann zeigen, dass die Bedingung (LICQ) die Bedingung (ACQ) impliziert.

Wir beweisen Satz 5.1.27.

Beweis. Wegen (ACQ) und Lemma 5.1.26 gilt $G'_S(x) \cap F_f(x) = \emptyset$. Somit gibt es kein $d \in \mathbb{R}^n$ mit

$$\nabla f(x) \cdot d > 0 \quad \text{und} \quad \nabla g_i(x) \cdot d \leq 0, \quad i \in I(x).$$

Ist nun A die Matrix mit Zeilen $\nabla g_i(x)$, $i \in I(x)$, dann hat

$$A \cdot d \leq 0, \quad -\nabla f(x) \cdot d > 0$$

keine Lösung. Mit Satz 2.7.5 (Minkowski-Farkas) folgt dann, dass es eine Lösung $z \geq 0$ gibt zu

$$A^T z = -(\nabla f(x))^T.$$

Bezeichnet man nun die Komponenten von z mit μ_i, ergibt sich die Aussage des Satzes. □

5.2 Konvexe Funktionen

Wir wollen hinreichende Bedingungen für die Existenz von Minimalstellen formulieren. Dazu werden wir konvexe Funktionen betrachten.

Definition 5.2.1. Sei $X \subset \mathbb{R}^n$ konvex. Eine Funktion $f : X \to \mathbb{R}$ heisst *konvex*, falls

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \quad x, y \in X, \quad \lambda \in [0,1],$$

und *strikt konvex*, falls

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad x, y \in X, \quad \lambda \in (0,1).$$

Die Funktion f heisst *(strikt) konkav*, falls $-f$ (strikt) konvex ist.

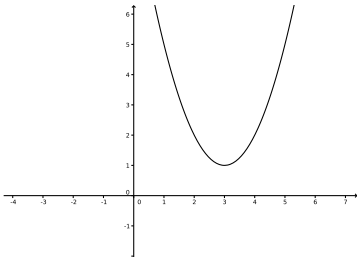


Abbildung 5.1: Strikt konvex.

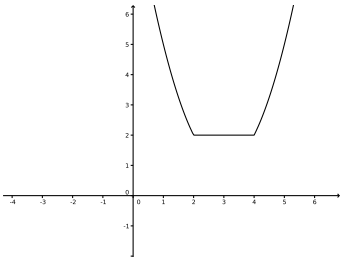


Abbildung 5.2: Konvex.

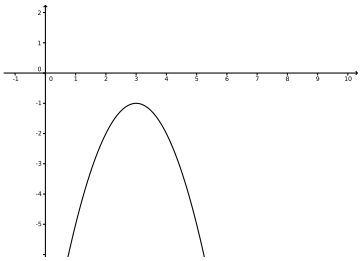


Abbildung 5.3: Strikt konkav.

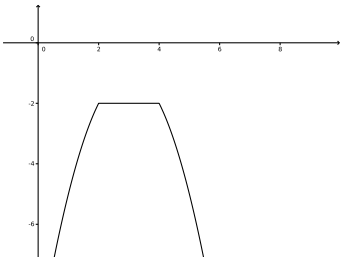


Abbildung 5.4: Konkav.

Beispiele 5.2.2.

(i) Die Funktion $f(x) = x^2$ ist strikt konvex auf $X = \mathbb{R}^n$. Die Funktionen $f(x) = e^x$ und $g(x) = e^{-x}$ sind strikt konvex auf $X = \mathbb{R}$.

(ii) Die Funktion $f(x) = -x^2$ ist strikt konkav auf $X = \mathbb{R}^n$. Die Funktion $f(x) = \log(x)$ ist strikt konkav auf $S = (0, +\infty)$.

(iii) Affin-lineare Funktionen, d.h. Funktionen der Form $f(x) = a^T x + b$ mit $a \in \mathbb{R}^n$ und $b \in \mathbb{R}$ sind gleichzeitig konvex und konkav auf $X = \mathbb{R}^n$, aber weder strikt konvex, noch strikt konkav, denn für alle $x, y \in \mathbb{R}^n$ und $\lambda \in [0, 1]$ hat man

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= a^T(\lambda x + (1-\lambda)y) + b \\ &= \lambda(a^T x + b) + (1-\lambda)(a^T y + b) \\ &= \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

Bemerkung 5.2.3. Eine konvexe Funktion $f : S \to \mathbb{R}$ muss nicht stetig auf der (konvexen) Menge) S sein, ihre Unstetigkeitsstellen sind aber in ∂S enthalten. Mit anderen Worten: Ist $f : S \to \mathbb{R}$ konvex, so ist f stetig auf dem Inneren von S.

Satz 5.2.4. *Sei $S \subset \mathbb{R}^n$ konvex und $f : S \to \mathbb{R}$. Ist S zusätzlich offen und f differenzierbar auf S, so gilt: Die Funktion f ist genau dann konvex, wenn*

$$f(x) \geq f(y) + \nabla f(y) \cdot (x - y)$$

für alle $x, y \in S$.

Beweis. Ohne.

Satz 5.2.5. *Sei $X \subset \mathbb{R}^n$ offen und konvex und sei $f : X \to \mathbb{R}$ differenzierbar.*

(i) Die Funktion f ist genau dann konvex, wenn für alle $x, y \in X$ die Ungleichung

$$(\nabla f(y) - \nabla f(x)) \cdot (y - x) \geq 0$$

gilt.

(ii) Ist $f : X \to \mathbb{R}$ zweimal differenzierbar, dann gilt: Die Funktion f ist genau dann konvex, wenn ihre Hesse-Matrix $H_f(y)$ ist in allen Punkten $y \in X$ positiv semidefinit ist.

Konvexität kann man auf verschiedene Weisen verallgemeinern. Für unsere Zwecke bietet sich folgender Begriff an.

Definition 5.2.6. Sei $X \subset \mathbb{R}^n$ offen und $f : X \to \mathbb{R}$. Die Funktion f heisst *pseudokonvex in $y \in X$*, falls f in y differenzierbar ist und für alle $x \in X$ die Implikation

$$\nabla f(y) \cdot (x - y) \geq 0 \quad \Rightarrow \quad f(x) \geq f(y)$$

gilt.

Wir werden in Kürze eine hinreichende Bedingung für das Vorliegen einer (globalen) Minimalstelle für pseudokonvexe Zielfunktionen f beweisen.

Bemerkung 5.2.7.

(i) Sei $S \subset \mathbb{R}^n$ konvex und offen. Dann ist jede konvexe Funktion $f : S \to \mathbb{R}$, die in $y \in S$ differenzierbar ist, auch pseudokonvex in y: Nach Satz 5.2.4 gilt

$$f(x) \geq f(y) + \underbrace{\nabla f(y) \cdot (x - y)}_{\geq 0} \geq f(y).$$

(ii) Ein Beispiel für eine Funktion, die pseudokonvex ist auf $\mathbb{R}$, aber nicht konvex, ist

$$f(x) = -\frac{1}{1 + x^2}, \quad x \in \mathbb{R}.$$

Die Funktion ist auf $\mathbb{R}$ zweimal stetig differenzierbar. Man hat

$$f'(x) = \frac{2x}{(1 + x^2)^2}, \quad x \in \mathbb{R}.$$

Ist $y = 0$, so ist $f(x) \geq -1 = f(y)$ für alle $x \in \mathbb{R}$. Ist $y > 0$, so impliziert $f'(y)(x - y) \geq 0$, dass $0 < y \leq x$ gilt und somit $f(x) \geq f(y)$. Analog für $y < 0$. Also ist f pseudokonvex in jedem Punkt $y \in \mathbb{R}$. Andererseits hat man

$$f''(x) = \frac{2(1 - 3x^2)}{(1 + x^2)^3}, \quad x \in \mathbb{R},$$

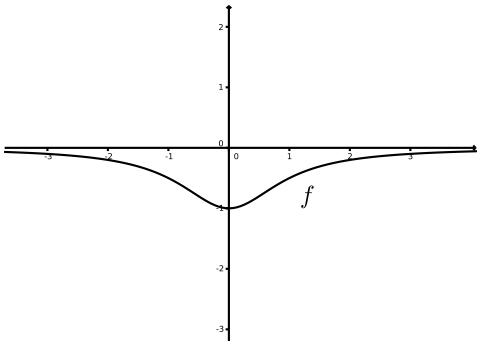
und das ist nichtnegativ für $|x| \leq 1/\sqrt{3}$, aber negativ für $|x| > 1/\sqrt{3}$. Also ist f nach Satz 5.2.5 (ii) nicht konvex auf $\mathbb{R}$.

Definition 5.2.8. Sei $X \subset \mathbb{R}^n$ konvex. Eine Funktion $f : X \to \mathbb{R}$ heisst *quasikonvex*, falls

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}, \quad x, y \in X,\ 0 \leq \lambda \leq 1.$$

Bemerkung 5.2.9.

(i) Es gilt also: Ist f konvex, so ist f auch quasikonvex.



(ii) Die Funktionen $f(x) = x^3$ und $f(x) = \sqrt{|x|}$ sind quasikonvex auf $\mathbb{R}$, aber nicht konvex.

Wir betrachten wieder (MP$\leq$), d.h.

$$\begin{aligned} f &\stackrel{!}{=} \min, \\ g_i &\leq 0, \quad i = 1, ..., m, \end{aligned}$$

wobei $f, g_i : X \to \mathbb{R}$ sind.

Wir können nun, wie gewünscht, eine *hinreichende KKT-Bedingung* dafür formulieren, dass ein Punkt eine globale Minimalstelle für f auf S ist:

Satz 5.2.10. *es seien $X \subset \mathbb{R}^n$ offen und konvex, $f, g_i : X \to \mathbb{R}$ und $x \in S := \bigcap_{i=1}^m \{g_i \leq 0\}$, sodass gilt:*

- *f ist in x pseudokonvex und*
- *für alle $i \in I(x)$ ist g_i quasikonvex und in x differenzierbar.*

Ist in x die KKT-Bedingung

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0 \quad \textit{für geeignete } \mu_i \geq 0$$

erfüllt, so ist x ein globales Minimum für f auf S.

Beweis. Sei $y \in S$ beliebig. Dann gilt für $i \in I(x)$, dass

$$g_i(y) \leq 0 = g_i(x).$$

Da die g_i quasikonvex sind, folgt für $0 \leq \lambda \leq 1$, dass

$$g_i(x + \lambda(y - x)) = g_i((1 - \lambda)x + \lambda y) \leq \max\{g_i(x), g_i(y)\} = g_i(x) = 0,$$

d.h. $y - x$ ist keine Abstiegsrichtung für $-g_i$ in x, also

$$\nabla g_i(x)(y - x) \leq 0.$$

Dann gilt aber auch

$$\sum_{i \in I(x)} \underbrace{\mu_i}_{\geq 0} \nabla g_i(x) \cdot (y - x) \leq 0,$$

und mit der KKT-Bedingung folgt, dass

$$\nabla f(x) \cdot (y - x) \geq 0.$$

Weil f pseudokonvex ist in x, folgt daraus nun $f(y) \geq f(x)$, und da $y \in S$ beliebig war, ist x somit eine globale Minimalstelle für f auf S. □

5.3 Lineare Restriktionen

Bemerkung 5.3.1. Wir können nun auch leicht *Probleme (MP) mit Gleichheitsbedingungen* betrachten, d.h. vom Typ

$$\begin{aligned} f &\stackrel{!}{=} \min, \\ g_i &\leq 0, \quad i = 1, ..., m, \\ h_j &= 0, \quad j = 1, ..., p. \end{aligned}$$

wobei $f, g_i, h_j : X \to \mathbb{R}$ sind. Das ist nämlich äquivalent zu

$$\begin{aligned} f &\stackrel{!}{=} \min, \\ g_i &\leq 0, \quad i = 1, ..., m, \\ h_j &\leq 0, \quad j = 1, ..., p, \\ -h_j &\leq 0, \quad j = 1, ..., p. \end{aligned}$$

In der KKT-Bedingung erscheinen für jede einzelne Gleichungsrestriktion $h_j(x) = 0$ dann die beiden Terme

$$\mu_j^+ \nabla h_j(x) + \mu_j^- \nabla(-h_j)(x) \quad \text{mit } \mu_j^+, \mu_j^- \geq 0,$$

bzw. ein Term $\lambda_j \nabla h_j(x)$ mit $\lambda_j := \mu_j^+ - \mu_j^-$. Die Restriktionen h_j und $-h_j$ sind in x immer straff. Insgesamt ergibt sich als KKT-Bedingung, dass

$$\nabla f(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) + \sum_{j=1}^{p} \lambda_j \nabla h_j(x) = 0$$

mit geeigneten $\mu_i \geq 0$ und $\lambda_j \in \mathbb{R}$ erfüllt sein muss.

Bemerkung 5.3.2. Eine besonders einfache Situation liegt vor, wenn die Restriktionen alle linear sind, wenn man also (wie früher bei linearen Optimierungsproblemen)

$$Ax \leq b \tag{5.1}$$

fordert, wobei A eine reelle $(m \times n)$-Matrix ist und $b \in \mathbb{R}^m$. Auch Gleichheitsbedingungen können (mittels obigem Trick) so geschrieben werden. Dann ist

$$S = \{Ax \leq b\},$$

und wenn $a_1, ..., a_m$ die Zeilen von A sind und

$$g_i(x) := a_i x - b_i,$$

so ist (5.1) äquivalent zu

$$g_i(x) \leq 0, \quad i = 1, ..., m.$$

In dieser Situation ist die Abadie-Bedingung (ACQ) stets erfüllt: Wir wissen bereits, dass $T_S(x) \subset G'_S(x)$ gilt, und weil hier

$$G'_S(x) = \{d \in \mathbb{R}^n : \nabla g_i(x) d \leq 0,\ i \in I(x)\} = \{d \in \mathbb{R}^n : a_i d \leq 0,\ i \in I(x)\}$$

ist, kann man auch die umgekehrte Inklusion $G'_S(x) \subset T_S(x)$ leicht sehen: Sei $d \in G'_S(x)$, also $a_i d \leq 0$, $i \in I(x)$. Für $i \notin I(x)$ hat man $a_j x < b_j$, und somit gibt es ein $\delta > 0$, sodass

$$a_j(x + \lambda d) < b_j, \quad j \notin I(x),\ 0 < \lambda < \delta$$

und

$$a_i(x + \lambda d) = a_i x + \lambda \underbrace{a_i d}_{\leq 0} \leq a_i x = b_i, \quad i \in I(x),\ \lambda > 0.$$

Damit folgt aber $x + \lambda d \in S$ für $0 < \lambda < \delta$, also $d \in D_S(x) \subset T_S(x)$.

Beispiel 5.3.3. Wir betrachten ein *quadratisches Optimierungsproblem* mit linearen Restriktionen: Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) := \frac{1}{2} x^T\, Q\, x + c^T\, x$$

soll minimiert werden unter den Restriktionen

$$Ax = b,$$

wobei Q eine symmetrische, positiv definite $(n \times n)$-Matrix ist,

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

eine reelle $(m \times n)$-Matrix mit Zeilen $a_1, ..., a_m$ und Rang m, $b \in \mathbb{R}^m$ und $c \in \mathbb{R}^n$.

Offensichtlich ist der Gradient $\nabla f(x) = x^T Q + c^T$, und die Hesse-Matrix ist $H_f(x) = Q$, sodass die positive Definitheit von Q die Konvexität nach sich zieht. Da alle Restriktionen linear sind, ist es nach der vorangegangenen Diskussion und den Sätzen 5.1.27 und 5.2.10 notwendig und hinreichend für das Vorliegen eines globalen Minimums in einem Punkt x mit $Ax = b$, dass

$$x^T\, Q + c^T + \sum_{i=1}^{m} \lambda_i a_i = x^T\, Q + c^T + \lambda^T A = 0$$

erfüllbar ist für geeignete $\lambda_i \in \mathbb{R}$, hier schreiben wir $\lambda = (\lambda_1, ..., \lambda_m)$ (KKT-Bedingung). Das gilt genau dann, wenn

$$Qx + A^T \lambda = -c \quad \text{und} \quad Ax = b$$

gilt bzw. in Matrix-Form,

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

Wir behaupten nun, dass die linke Matrix invertierbar ist: Weil

$$\begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} = \begin{pmatrix} Q & 0 \\ A & E_m \end{pmatrix} \begin{pmatrix} E_m & Q^{-1}A^T \\ 0 & -AQ^{-1}A^T \end{pmatrix}$$

ist, folgt, dass

$$\det \begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix} = \underbrace{\det(Q)}_{\neq 0} \cdot \det(-AQ^{-1}A^T)$$

gilt, und somit genügt es zu zeigen, dass $AQ^{-1}A^T$ injektiv (und somit invertierbar auf ihrem Bild) ist. Da Q positiv definit ist, ist auch Q^{-1} positiv definit, und es gibt eine symmetrische und positiv definite $(n\times n)$-Matrix S mit $Q^{-1} = SS^T$. Sei nun $AQ^{-1}A^Ty = 0$ für ein $y \in \mathbb{R}^m$. Dann hat man

$$\begin{aligned} 0 = \left\langle AQ^{-1}A^Ty, y\right\rangle = \left\langle ASS^TA^Ty, y\right\rangle = \left\langle (AS)(AS)^Ty, y\right\rangle \\ = \left\langle (AS)^Ty, (AS)^Ty\right\rangle = \left\|(AS)^Ty\right\|^2, \end{aligned}$$

also $(AS)^Ty = 0$. Nun ist aber $(AS)^T$ eine reelle $(n \times m)$-Matrix mit Rang m, also injektiv, und somit muss $y = 0$ sein. Das bedeutet, $AQ^{-1}A^Ty$ ist injektiv, und die obige Behauptung ist richtig. Wir erhalten

$$\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} Q & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} -c \\ b \end{pmatrix},$$

und das sagt uns, dass in x die KKT-Bedingung mit λ gilt, dass somit x nach Satz 5.2.10 eine Minimalstelle für f auf dem zulässigen Bereich ist, also eine Lösung des vorliegenden Optimierungsproblems. Zudem lässt sich x aus den gegebenen Daten A, b, c und Q leicht durch eine lineare Gleichung berechnen.

Kapitel 6

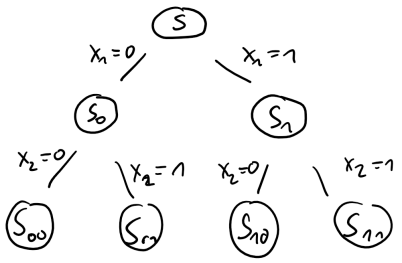
Ganzzahlige Optimierung

6.1 Teile und Herrsche

Für $S \subset \mathbb{R}^n$, $c \in \mathbb{R}^n$ betrachten wir das Maximierungsproblem $\max c^T x, x \in S$. Wir wollen das Problem derart in Teile zerlegen, dass wir das Ausgangsproblem lösen knnen.

Lemma 6.1.1. *Für $k \in \{1, \ldots, m\}$ seien $S_k \subset \mathbb{R}^n$, $z^k = \max\{c^T x \,|\, x \in S_k\}$ und $S = \bigcup_{k=1}^m S_k$. Dann gilt* $\max\{c^T x \,|\, x \in S\} = \max\{z^k \,|\, k \in \{1, \ldots, m\}\}$.

Beispiel 6.1.2. $S \subset \{0,1\}^2$



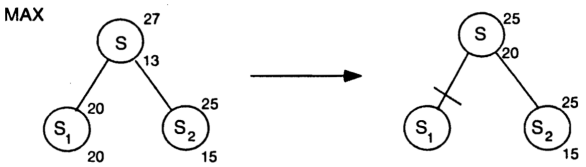
Obere und untere Schranken für Teilprobleme liefern uns Schranken fr das Ausgangsproblem, denn es gilt das

Lemma 6.1.3. *Für $k \in \{1, \ldots, m\}$ seien $S_k \subset \mathbb{R}^n$, $z^k = \max\{c^T x \,|\, x \in S_k\}$, sowie $\bar{z}^k \geq z^k \geq \underline{z}^k$ für gewisse $\bar{z}^k, \underline{z}_k \in \mathbb{R}$ und $S = \bigcup_{k=1}^m S_k$. Sei $z_0 \in S$. Dann gilt*

$$\bar{z} := \max_k \bar{z}^k \geq c^T z_0 \geq \max_k \underline{z}^k =: \underline{z}.$$

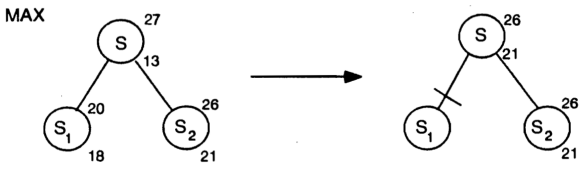
Beispiel 6.1.4. In folgenden Beispielen seien die Zahlen an den Knoten des Baumes jeweils obere und untere Schranken des (Teil-)Problems.

1. Hier haben wir $\bar{z} = \max\{20, 25\} = 25$ und $\underline{z} = \max\{20, 15\} = 20$.



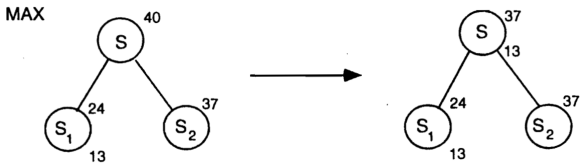
   Also ist S_1 gelöst und wir müssen dieses Teilproblem nicht weiter betrachten.

2. Hier haben wir $\bar{z} = \max\{20, 26\} = 26$ und $\underline{z} = \max\{18, 21\} = 21$.



Wegen $\bar{z}^1 < \underline{z}^2$ mssen wir S_1 nicht weiter betrachten.

3. Hier haben wir $\bar{z} = \max\{24, 37\} = 24$ und $\underline{z} = \max\{13, -\} = 13$.



Daher mssen beide Teilprobleme weiter betrachtet werden.

Aus diesem Beispiel leiten wir eine Vorgehensweise zum Abarbeiten des Baumes ab. Wir können Blätter des Baumes „wegstreichen“, wegen

1. Optimaliät: $z_t = \max\{cx^T \mid x \in S_t\}$ ist glöst
2. Beschränktheit: $\bar{z}_t < \underline{z}$
3. Zulässigkeit: $S_t = \emptyset$

Es ergeben sich zentrale Fragen zur Vorgehensweise. Zum Beispiel: Wie teilt man das Ausgangspromblem auf? Wie ermittelt man Schranken?

6.2 Branch and Bound