

Operations Research

Vorlesungsskript

Thorsten Dickhaus
Universität Bremen
Wintersemester 2016 / 2017
Version: 25. März 2017

Vorbemerkungen

Das Material zu diesem Skript habe ich im Wesentlichen aus den Büchern von Nickel et al. (2014), Kathöfer and Müller-Funk (2008), Kohlas (1977), Heyman and Sobel (1982, 1984), Boyd and Vandenberghe (2004), Eiselt and Sandblom (2012), Ellinger et al. (2003) sowie einem Vorlesungsskript von Gerhard Dikta entnommen. Sollten sich in den übernommenen Teilen Fehler finden, so bin dafür natürlich ich verantwortlich. Lob und positive Kritik gebührt indes den Original-Autoren.

Für die Manuskripterstellung danke ich Natalia Sirotko-Sibirskaya.

Übungsaufgaben und MATLAB-Programme zu diesem Kurs stelle ich auf Anfrage gerne zur Verfügung. Einige Referenzen dazu finden sich im Text an den zugehörigen Stellen.

Inhaltsverzeichnis

0	Einführung und Übersicht	1
1	Lineare Optimierung	4
2	Ganzzahlige lineare Optimierung	18
3	Nichtlineare Optimierung	24
3.1	Lokale Extrema ohne Nebenbedingungen	24
3.2	Extrema konvexer und konkaver Funktionen (ohne Nebenbedingungen)	29
3.3	Nicht-lineare Optimierung unter Nebenbedingungen	37
4	Dynamische und stochastische Optimierung	46
4.1	Dynamische Programmierung	46
4.2	Stochastische dynamische Optimierung	55
	Literaturverzeichnis	68

Kapitel 0

Einführung und Übersicht

Operations Research (OR) befasst sich mit der Lösung von Optimierungsproblemen. Einfache Optimierungsprobleme sind aus der Analysis bekannt (Bestimmung von lokalen und globalen Extrema einer gegebenen differenzierbaren Funktion f einer oder mehrerer Veränderlicher).

Wir werden in dieser Vorlesung kompliziertere Optimierungsprobleme angehen. Die folgenden Komplikationen werden auftreten:

1. Die Zielfunktion f soll unter Nebenbedingungen optimiert werden.
2. Nur ganzzahlige Lösungen kommen in Frage.
3. Die Funktion f ist nicht in expliziter Form gegeben.
4. Es herrschen Unsicherheiten über die Gegebenheiten des Problems (Zufallseinflüsse treten auf).

Alle diese Komplikationen haben große Praxisrelevanz, wie die folgenden Beispiele illustrieren mögen.

Beispiel 0.1 (Nebenbedingungen)

Ein Farbhersteller produziert zwei unterschiedliche Farben, sagen wir Außenfarbe und Innenfarbe. Zur Herstellung der Farben werden zwei Basisrohstoffe A und B benötigt. Pro Tag stehen dabei 6 Tonnen des Rohstoffs A und 8 Tonnen des Rohstoffs B zu Verfügung.

Zur Produktion einer Tonne Außenfarbe werden eine Tonne von A und zwei Tonnen von B benötigt. Bei der Innenfarbe ist dieses Verhältnis genau umgekehrt. Der Verkaufspreis pro Tonne Außenfarbe beträgt 3.000 EUR, der pro Tonne Innenfarbe beträgt 2.000 EUR.

Schließlich ist aus einer Marktstudie bekannt, dass höchstens zwei Tonnen Innenfarbe pro Tag auf dem relevanten Markt nachgefragt werden und dass höchstens eine Tonne Innenfarbe mehr als Außenfarbe nachgefragt wird.

Frage: Was sind die optimalen täglichen Produktionsmengen an Außen- und Innenfarbe, um den Erlös unter den genannten Bedingungen zu maximieren?

Beispiel 0.2 (Ganzzahligkeit)

Eine Fluggesellschaft beabsichtigt, neue Maschinen zu erwerben. Es kommen zwei unterschiedliche Flugzeugtypen A und B in Frage. Insgesamt stehen 24 Mio. USD für die Anschaffungen zur Verfügung. Ferner kann die Fluggesellschaft maximal 1.300 Arbeitsstunden pro Jahr für die Wartung der Maschinen zur Verfügung stellen. Die relevanten Daten für die Kaufentscheidung seien wie folgt.

	Typ A	Typ B
Profit pro Jahr [1.000 USD]	200	100
Preis pro Maschine [1 Mio. USD]	5	4
Wartungsstunden pro Jahr	200	500

Frage: Wie viele Maschinen der beiden Typen A und B sollen angeschafft werden, um den jährlichen Profit unter den genannten Bedingungen zu maximieren?

Beachte: Die Lösung (x_A, x_B) muss offenbar ganzzahlig sein!

Natürlich kann man alle Möglichkeiten (x_A, x_B) durchprobieren, die jeweiligen sich ergebenden Profite auf tabellieren und das Maximum ablesen. Für größere Probleme führt dies indes zu kombinatorischer Explosion, weswegen effiziente Suchstrategien zur Findung des Optimums notwendig sind.

Beispiel 0.3 (Zielfunktion nicht explizit gegeben)

Angenommen, $N = 10$ Artikel unterschiedlicher Größe sollen in Holzschachteln verpackt werden, wobei die i -te Holzschachtel c_i Geldeinheiten (GE) kostet. Die Artikel seien der Größe nach absteigend geordnet, so dass $c_1 > c_2 > \dots > c_N = c_{10}$ gilt.

Da kleinere Artikel natürlich auch in größere Schachteln passen, hat man entschieden, nur fünf unterschiedliche Schachtelgrößen herzustellen.

Ziel: Finde fünf Größen so, dass alle $N = 10$ Artikel verpackt werden können und die Gesamtverpackungskosten minimal werden! Hier hängen die fünf Argumente, die in die Zielfunktion (Kostenfunktion) eingesetzt werden, selbst von der Optimierungsstrategie ab!

Beispiel 0.4 (Zufallseinflüsse)

Der Zustand einer Maschine lasse sich durch eine der Zahlen $0, 1, \dots, M$ mit $M \in \mathbb{N}$ beschreiben. Eine neue Maschine befinde sich in Zustand 0. Die Maschine unterliege einem Verschleiß, der eine Verschlechterung (Anwachsen) des Zustands bewirken kann. Allerdings kann nicht mit

100 %iger Sicherheit vorhergesagt werden, wie intensiv die Maschine genutzt wird und in welcher Weise sich die Nutzung auf den Zustand auswirkt. Die Maschine werde zu diskreten Zeitpunkten $n = 0, \dots, N$ inspiziert. In Abhängigkeit vom beobachteten Zustand $s_n \in \{0, \dots, M\}$ kann eine der beiden Entscheidungen „Maschine ersetzen“ ($a_n = 1$) oder „Maschine nicht ersetzen“ ($a_n = 0$) getroffen werden, wobei eine Ersatzmaschine unmittelbar verfügbar sei. Zu berücksichtigen seien Betriebskosten der Höhe $c(z_n)$ GE in Abhängigkeit vom Zustand $z_n = (1 - a_n)s_n$ der Maschine unmittelbar nach der Entscheidung zum Zeitpunkt n , sowie im Falle $a_n = 1$ zusätzlich Ersetzungskosten der Höhe k GE. Eine Verschlechterung der Zustands der Maschine zwischen den Zeitpunkten n und $n + 1$ von z_n nach s_{n+1} ergebe sich mit Wahrscheinlichkeit $q(z_n, s_{n+1})$.

Ziel: Ersetzungsstrategie, die erwarteten Gesamtkosten minimiert (über den Beobachtungshorizont $n = 0, \dots, N$).

Während alle diese vier Beispiele aus dem Wirtschaftsumfeld stammen und die Wirtschaftswissenschaften ein wesentliches Anwendungsgebiet des OR darstellen, sind die Lösungsverfahren in der Regel mathematisch begründet (Optimalität der Lösung kann bewiesen werden) und auch anderweitig einsetzbar. So werden etwa in der mathematischen Statistik Likelihood-basierte Schätz- und Testverfahren vermittels nichtlinearer Optimierung (unter Nebenbedingungen) formalisiert und durchgeführt.

Übersicht über die Inhalte dieser Lehrveranstaltung:

Kapitel 1: Lineare Optimierung

Kapitel 2: Ganzzahlige lineare Optimierung

Kapitel 3: Nichtlineare Optimierung

Kapitel 4: Dynamische und stochastische Optimierung

Kapitel 1

Lineare Optimierung

Beispiel 1.1 (Beispiel 0.1 fortgesetzt:)

Wir greifen Beispiel 0.1 (Farbenproduktion) noch einmal auf, formulieren das Optimierungsproblem mathematisch und ermitteln die optimalen Tagesproduktionsmengen argumentativ. Die relevanten Daten finden sich in der nachfolgenden Tabelle.

	Verbrauch in Tonnen pro Tonne Farbe		Maximum
	außen	innen	
Rohmaterial A	1	2	6
Rohmaterial B	2	1	8
Verkaufspreis [EUR]	3000	2000	
Nachfrage [Tonnen]	$\geq \text{innen} - 1$	≤ 2	

Wir bezeichnen die Tagesproduktionsmengen (in Tonnen) an Außen- und Innenfarbe mit x_a bzw. x_i . Als Geldeinheit wählen wir 1000 EUR. Somit ergibt sich die zu maximierende Zielfunktion (Erlösfunktion) als

$$f(x_a, x_i) = 3x_a + 2x_i.$$

Wir müssen f maximieren unter den folgenden linearen (in x_a und x_i) Nebenbedingungen (Constraints).

(C1) $x_a + 2x_i \leq 6$

(C2) $2x_a + x_i \leq 8$

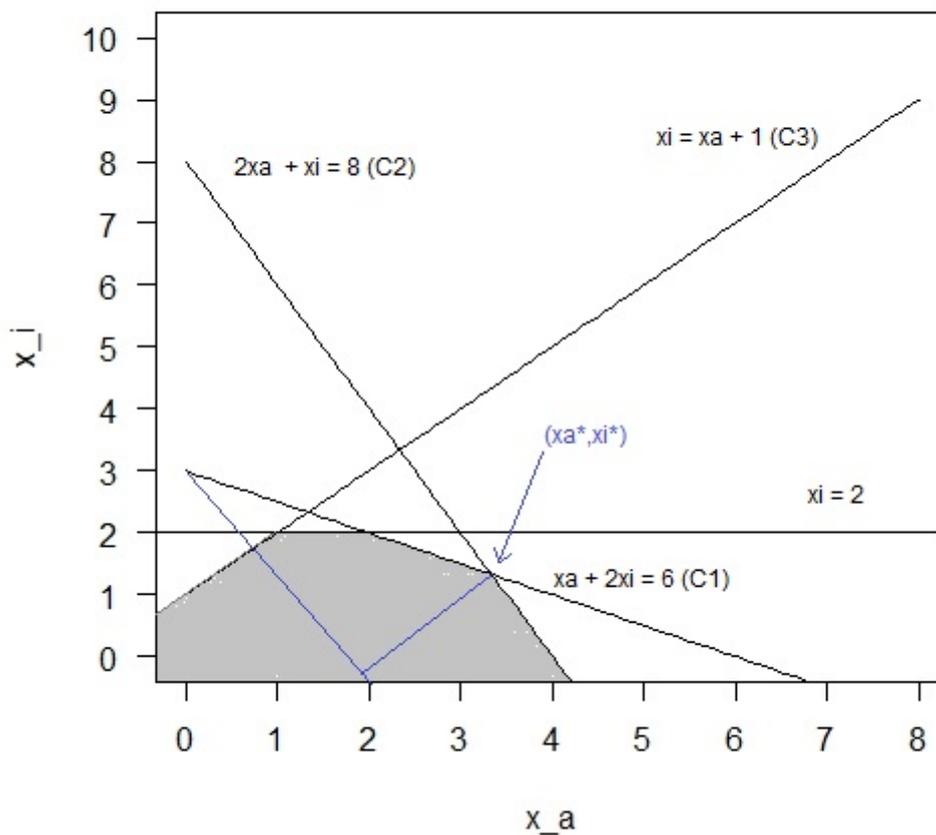
(C3) $x_a \geq x_i - 1$

(C4) $x_i \leq 2$

(C5) $x_i \geq 0$

(C6) $x_a \geq 0$

Zur weiteren Argumentation ist es hilfreich, sich den zulässigen Bereich für (x_a, x_i) , d.h., die Menge aller Wertepaare (x_a, x_i) , die alle sechs Constraints erfüllen, in der (x_a, x_i) -Ebene grafisch zu veranschaulichen.



Der zulässige Bereich ist schraffiert. Setzen wir nun die Zielfunktion gleich $K \geq 0$, so erhalten wir die Isoerlösgerade

$$x_i = \frac{K - 3x_a}{2}.$$

Für $K = 6$ ist die Isoerlösgerade farbig eingezeichnet.

Eine Erhöhung des Erlöses K bewirkt eine Parallelverschiebung der Isoerlösgeraden nach rechts oben. Je weiter die Isoerlösgeraden also nach rechts oben parallel verschoben werden kann, ohne den zulässigen Bereich zu verlassen, desto größer wird der Erlös unter den gemachten Nebenbedingungen.

Führt man diese Parallelverschiebung geometrisch durch, so sieht man, dass der Eckpunkt durch Schnitt der zu (C1) und (C2) gehörigen Geraden zustande kommt, zum maximalen Erlös führt.

Wir lösen auf:

$$x_a + 2x_i = 6 \quad (I)$$

$$2x_a + x_i = 8 \quad (II)$$

$$(I) + (II) \Rightarrow 3x_a + 3x_i = 14 \iff x_a = \frac{14-3x_i}{3} = \frac{14}{3} - x_i \quad (III)$$

$$(III) \text{ in } (I) : \frac{14}{3} + x_i = 6 = \frac{18}{3} \iff x_i = \frac{4}{3} \quad (IV)$$

$$(IV) \text{ in } (III) : x_a = \frac{10}{3}$$

\Rightarrow Die Lösung des Optimierungsproblems lautet $(x_a^*, x_i^*) = (\frac{10}{3}, \frac{4}{3}) = (3\frac{1}{3}, 1\frac{1}{3})$.

Ein wesentliches mathematisches Ergebnis der Theorie der linearen Optimierung ist, dass das Optimum einer linearen Zielfunktion über einen zulässigen Bereich, der durch lineare Constraints beschrieben wird, stets an einem Eckpunkt des zulässigen Bereichs angenommen wird (falls es denn existiert). Das Simplexverfahren bietet darüber hinaus eine effiziente Möglichkeit, die Eckpunkte systematisch abzusuchen.

Definition 1.2

Unter einem linearen Optimierungsproblem bzw. einem linearen Programm (LP) versteht man die Aufgabe, die lineare Zielfunktion f , gegeben durch

$$f(x) = c^\top x, \quad c \in \mathbb{R}^n, \quad x \in \mathbb{R}^n,$$

zu maximieren, wobei c ein fest vorgegebener Vektor ist. Das Argument x muss dabei aus dem zulässigen Bereich \mathcal{M} sein, der durch eine vorgegebene Matrix $A \in \mathbb{R}^{m \times n}$ und einen vorgegebenen Vektor $b \in \mathbb{R}^m$ beschrieben ist. Die Matrix A und der Vektor b stellen m lineare Nebenbedingungen an x .

Kanonische Form des LP:

$$\mathcal{M} = \{x \geq 0 : Ax \leq b\},$$

wobei die Relationszeichen komponentenweise zu verstehen sind.

Standardform des LP:

$$\mathcal{M} = \{x \geq 0 : Ax = b\}$$

Wir schreiben kurz: $LP_{\max}(A, b, c)$

Bemerkung 1.3

Offenbar sind lineare Minimierungs- und Maximierungsprobleme wegen der Beziehung

$$\min c^\top x = -\max\{-c^\top x\}$$

äquivalent. Es genügt also, sich auf Maximierungsprobleme zu beschränken. Mit anderen Worten ist $LP_{\min}(A, b, c)$ äquivalent zu $LP_{\max}(A, b, -c)$.

Lemma 1.4

Jedes LP_{\max} in kanonischer Form lässt sich in ein äquivalentes LP_{\max} in Standardform durch Einführung zusätzlicher Variablen, sogenannter "Schlupfvariablen", umschreiben. Es genügt also, LPs in Standardform zu betrachten.

Beweis:

Angenommen, das LP ist in kanonischer Form gegeben. Das heißt, es existieren m Nebenbedingungen der Form

$$(*) \quad a_{i,1}x_1 + \dots + a_{i,n}x_n \leq b_i, \quad 1 \leq i \leq m.$$

Definiere nun für die i -te Nebenbedingung eine neue Variable x_{n+i} mittels

$$x_{n+i} := b_i - a_{i,1}x_1 - \dots - a_{i,n}x_n.$$

Dann ist (*) äquivalent zu $x_{n+i} \geq 0$. Somit kann x_{n+i} als neue Variable eingeführt werden, die die i -te Nebenbedingung kodiert, $1 \leq i \leq m$. Also lässt sich ein kanonisches $LP_{\max}(A, b, c)$ in ein äquivalentes $LP_{\max}(\tilde{A}, b, \tilde{c})$ umschreiben, wobei

$$\tilde{c} := (c_1, \dots, c_n, 0, \dots, 0)^\top \in \mathbb{R}^{n+m} \quad \text{und}$$

$$\tilde{A} = \left[\begin{array}{ccc|ccc} a_{1,1} & \dots & a_{1,n} & 1 & 0 & \dots & 0 \\ a_{2,1} & \dots & a_{2,n} & 0 & 1 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} & 0 & 0 & \dots & 1 \end{array} \right] = (A|I_m) \in \mathbb{R}^{m \times (n+m)}.$$

Der zulässige Bereich $\mathcal{M}_{\text{kanonisch}} = \{x \in \mathbb{R}^n | x \geq 0, Ax \leq b\}$ der kanonischer Form wird damit zum zulässigen Bereich $\mathcal{M}_{\text{Standard}} = \{\tilde{x} \in \mathbb{R}^{n+m} | \tilde{x} \geq 0, \tilde{A}\tilde{x} = b\}$ der Standardform. Hierbei ist also

$$\tilde{x} = (x_1, \dots, x_n, x_{n+1}, \dots, x_{n+m})^\top = \begin{pmatrix} x \\ b - Ax \end{pmatrix}.$$

Man überprüft leicht, dass das $\text{LPmax}(A, b, c)$ in kanonischer Form äquivalent zum $\text{LPmax}(\tilde{A}, b, \tilde{c})$ in Standardform ist. ■

Definition 1.5

Der Durchschnitt endlich vieler Halbräume der Form $\{x \in \mathbb{R}^n : a_j^\top x \leq b_j\}$ mit fest vorgegebenen Vektoren a_j ($j = 1, 2, \dots$) und fest vorgegebenen reellen Zahlen b_j ($j = 1, 2, \dots$) heißt ein Polyeder.

Beachte:

$$\{x \in \mathbb{R}^n : a_j^\top x = b_j\} = \{x \in \mathbb{R}^n : a_j^\top x \leq b_j\} \cap \{x \in \mathbb{R}^n : -a_j^\top x \leq -b_j\}.$$

Somit ist der zulässige Bereich \mathcal{M} eines LPmax in Standardform also ein Polyeder.

Definition 1.6

Eine Menge $\mathcal{M} \subseteq \mathbb{R}^n$ heißt *konvex*, falls für alle $x, y \in \mathcal{M}$ und $\lambda \in [0, 1]$ stets folgt, dass $\lambda x + (1 - \lambda)y \in \mathcal{M}$. Eine Funktion $f : \mathcal{M} \rightarrow \mathbb{R}$ heißt *konvex*, falls \mathcal{M} nicht leer und konvex ist und falls $\forall x, y \in \mathcal{M} : \forall \lambda \in [0, 1] : f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$.

Bemerkung 1.7

(a) Halbbäume sind konvexe Mengen.

(b) Lineare Funktionen sind konvex.

(c) Falls eine Menge $\mathcal{M} \subseteq \mathbb{R}^n$ konvex ist, so folgt aus $x_i \in \mathcal{M}$, $\lambda_i \geq 0$ ($1 \leq i \leq k \in \mathbb{N}$) und $\sum_{i=1}^k \lambda_i = 1$ stets, dass $\sum_{i=1}^k \lambda_i x_i \in \mathcal{M}$.

Lemma 1.8

Seien \mathcal{M}_i konvexe Mengen für alle i aus einer beliebigen Indexmenge I . Dann ist auch $\mathcal{M} := \bigcap_{i \in I} \mathcal{M}_i$ konvex.

Beweis: Übungsaufgabe.

Korollar 1.9

Jedes Polyeder ist konvex und abgeschlossen, aber nicht notwendigerweise beschränkt.

Satz 1.10 (Extremalsatz von Weierstraß)

Jede reellwertige Funktion, die auf einer abgeschlossenen und beschränkten Teilmenge \mathcal{M} des \mathbb{R}^n stetig ist, nimmt auf \mathcal{M} ihr Maximum und ihr Minimum an.

Beweis: Siehe Satz 36.3 in Heuser (2009).

Korollar 1.11

Falls der zulässige Bereich \mathcal{M} eines LP nicht-leer und beschränkt ist, so existiert eine Lösung des LPs, da die Zielfunktion f linear und damit stetig ist.

Definition 1.12

Sei \mathcal{M} eine konvexe Menge und $a \in \mathcal{M}$. Dann heißt a ein Extremalpunkt von \mathcal{M} , falls aus $x, y \in \mathcal{M}$, $\lambda \in (0, 1)$ und $a = \lambda x + (1 - \lambda)y$ stets $a = x = y$ folgt. Die Extremalpunkte eines Polyeder $\mathcal{M} \in \mathbb{R}^n$ heißen die Ecken von \mathcal{M} . Eine konvexe Teilmenge $\mathcal{E} \subset \mathcal{M}$ einer konvexen Menge \mathcal{M} heißt Extremalmenge von \mathcal{M} , falls aus $a \in \mathcal{E}$, $x, y \in \mathcal{M}$, $\lambda \in (0, 1)$ und $a = \lambda x + (1 - \lambda)y$ stets folgt, dass $x, y \in \mathcal{E}$.

Lemma 1.13

Jedes Polyeder \mathcal{M} der Form $\mathcal{M} = \{x \geq 0 : Ax = b\}$ besitzt mindestens einen Eckpunkt. Ferner ist die Zahl der Eckpunkte von \mathcal{M} endlich.

Beweisskizze:

Die Endlichkeit der Menge der Eckpunkte folgt aus der Tatsache, dass jeder Eckpunkt durch Schnitt von Hyperebenen zu Stande kommt, die durch die Zeilen von A definiert sind, vgl. Satz 3.1.5 in Jungnickel (2015).

Die Existenz eines Eckpunktes zeigt man wie folgt: Sei $a \in \mathcal{M}$. Falls a keine Ecke ist, so gibt es eine Richtung $h \neq 0$, so dass $a \pm h \in \mathcal{M}$. Man laufe in Richtung $a + \lambda h$ so lange, bis man für $\lambda > 0$ oder $\lambda < 0$ an den Rand von \mathcal{M} stößt. Der Randpunkt sei a' . Dann ist entweder a' eine Ecke oder obiger Vorgang lässt sich mit a' als neuen Startpunkt und einer von h linear unabhängigen Richtung h' wiederholen. Nach n Wiederholungen gibt es keine linear unabhängigen Richtungen mehr, d. h., man muss spätestens dann eine Ecke gefunden haben (nach Übungsaufgabe 3.9.4 in Jarre and Stoer (2004)). ■

Satz 1.14

Falls ein LP Lösungen besitzt, so ist die Lösungsmenge \mathcal{E} eine Extremalmenge (also Seitenfläche) des Polyeders \mathcal{M} , der den zulässigen Bereich beschreibt.

Beweis: Angenommen, das interessierende LPmax besitzt Lösungen. Dann suchen wir

$$f^* := \max\{c^\top x : x \in \mathcal{M}\} \in \mathbb{R}.$$

Die Lösungsmenge $\mathcal{E} = \{x \in \mathcal{M} : c^\top x = f^*\} \subseteq \mathcal{M}$ ist ein Polyeder in \mathcal{M} , also insbesondere eine konvexe Teilmenge von \mathcal{M} .

Annahme: \mathcal{E} ist keine Extremalmenge von \mathcal{M} .

Sei nun $a \in \mathcal{E}$ und besitze die Darstellung $a = \lambda x + (1 - \lambda)y$ mit $0 < \lambda < 1$, $x, y \in \mathcal{M}$ und o.B.d.A. $x \notin \mathcal{E}$. Damit ist $c^\top x < f^*$, und natürlich ist überdies $c^\top y \leq f^*$. Wir erhalten den

Widerspruch

$$f^* = c^\top a = \lambda c^\top x + (1 - \lambda)c^\top y < \lambda f^* + (1 - \lambda)f^* = f^*.$$

Also kann die gemachte Annahme nicht richtig sein; \mathcal{E} ist also eine Extremalmenge von \mathcal{M} . ■

Korollar 1.15

- (a) Falls ein LP mit zulässigem Bereich \mathcal{M} in Standardform Lösungen besitzt, so gibt es unter den Lösungen auch Eckpunkte von \mathcal{M} .
- (b) Falls unter den Gegebenheiten von Teil (a) der zulässige Bereich \mathcal{M} nicht-leer und beschränkt ist, so existieren Lösungen des LP, die Eckpunkte von \mathcal{M} sind.

Bemerkung 1.16

Korollar 1.15 macht keine Aussagen zur Eindeutigkeit der Lösung. Falls man nur an f^* (dem optimalen Wert der Zielfunktion) selbst interessiert ist, so stellt sich die Frage nach dieser Eindeutigkeit nicht.

Satz 1.17 (Charakterisierung von Ecken)

Gegeben sei ein LPmax in Standardform, also

$$\mathcal{M} = \{x \in \mathbb{R}^{n+m} : x \geq 0, Ax = b\}$$

für $A \in \mathbb{R}^{m \times n+m}$, $x = (x_1, \dots, x_{n+m})^\top$, $b \in \mathbb{R}^m$. Wir nehmen an, dass $\text{rg}(A) = m$ ist, da ansonsten redundante Nebenbedingungen existieren. Dann gilt:

- (i) Sei x ein Eckpunkt von \mathcal{M} und bezeichne $\Delta := \{j : x_j > 0\}$. Dann besteht die Menge $\{A^{(j)} : j \in \Delta\}$ aus linear unabhängigen Spaltenvektoren der Matrix A .
- (ii) Sei $\Delta \subset \{1, \dots, n+m\}$ mit $|\Delta| = m$. Falls dann $\{A^{(j)} : j \in \Delta\}$ aus m linear unabhängigen Spaltenvektoren von A besteht und x^* eine Lösung von $Bx = b$ mit $x^* \geq 0$ ist, wobei $B = (A^{(j)} : j \in \Delta) \in \mathbb{R}^{m \times m}$ die $(m \times m)$ -Matrix bezeichnet, deren Spalten gerade die $A^{(j)}$, $j \in \Delta$, sind, dann ist x mit

$$x_j := \begin{cases} x_j^*, & \text{falls } j \in \Delta, \\ 0, & \text{sonst} \end{cases}$$

ein Eckpunkt von \mathcal{M} .

Beweis:

zu (i): Sei x ein Eckpunkt von \mathcal{M} . Zur Vereinfachung der Notation nehmen wir an, dass $x = (x_1, \dots, x_k, 0, \dots, 0)^\top$ gilt, wobei $0 \leq k \leq n+m$ und $x_i > 0$ für alle $1 \leq i \leq k$ ist. Aus $Ax = b$

folgt, dass $b = \sum_{i=1}^k x_i A^{(i)}$.

Zu zeigen: $(A^{(1)}, \dots, A^{(k)})$ sind linear unabhängig.

Wir führen einen Widerspruchsbeweis. Angenommen, es gäbe $\lambda_1, \dots, \lambda_k \in \mathbb{R}$, die nicht alle gleich Null sind mit $\sum_{i=1}^k \lambda_i A^{(i)} = 0$. Wir definieren dann

$$\alpha := \min(\{x_i/\lambda_i : \lambda_i > 0\} \cup \{-x_i/\lambda_i : \lambda_i < 0\})$$

sowie zwei Vektoren u und v wie folgt:

$$u = (x_1 + \alpha\lambda_1, \dots, x_k + \alpha\lambda_k, 0, \dots, 0)^\top,$$

$$v = (x_1 - \alpha\lambda_1, \dots, x_k - \alpha\lambda_k, 0, \dots, 0)^\top.$$

Offenbar gilt $u \geq 0$ und $v \geq 0$. Ferner liegen sowohl u als auch v im zulässigen Bereich, denn

$$\begin{aligned} A_u &= \sum_{i=1}^k (x_i + \alpha\lambda_i) A^{(i)} \\ &= \sum_{i=1}^k x_i A^{(i)} + \alpha \sum_{i=1}^k \lambda_i A^{(i)} \\ &= b + \alpha \vec{0} = b, \end{aligned}$$

und analog zeigt man, dass $Av = b$ ist.

Allerdings ist $x = \frac{1}{2}u + \frac{1}{2}v$ eine echte Konvexkombination aus $u, v \in \mathcal{M}$ und somit kein Eckpunkt von \mathcal{M} .

zu (ii): Sei $\Delta \subset \{1, \dots, n+m\}$ mit $|\Delta| = m$ so vorgegeben, dass $\{A^{(j)} : j \in \Delta\}$ aus m linear unabhängigen Spaltenvektoren von A besteht. Somit ist $B = (A^{(j)} : j \in \Delta) \in \mathbb{R}^{m \times m}$ eine invertierbare Matrix.

Sei $x^* \geq 0$ eine Lösung von $Bx = b$. Wir können die Existenz von x^* voraussetzen, weil sonst nichts zu zeigen ist. Zur Vereinfachung der Notation nehmen wir $\Delta = \{1, \dots, m\}$ an. Definiere $x = (x_1^*, \dots, x_k^*, 0, \dots, 0)^\top$, wobei $k \leq m$ die Anzahl der echt positiven Einträge von $x^* \in \mathbb{R}^m$ bezeichnet.

Zu zeigen: x ist ein Eckpunkt von \mathcal{M} .

Wir führen erneut einen Widerspruchsbeweis. Angenommen, es existieren $y, z \in \mathcal{M}$ mit $y \neq z$ und $x = \lambda y + (1 - \lambda)z$ für $0 < \lambda < 1$. Da $y \geq 0$ und $z \geq 0$ gilt, folgt aus der Darstellung von x ,

dass $y = (y_1, \dots, y_k, 0, \dots, 0)^\top$ sowie $z = (z_1, \dots, z_k, 0, \dots, 0)^\top$ sein muss. Wir folgern, dass

$$\sum_{i=1}^k y_i A^{(i)} = b = \sum_{i=1}^k z_i A^{(i)}.$$

Daraus folgt aber wiederum, dass $\sum_{i=1}^k (y_i - z_i) A^{(i)} = 0$. Da die Spaltenvektoren $A^{(1)}, \dots, A^{(k)}$ aber als linear unabhängig vorausgesetzt waren, muss $y_i = z_i$ für alle $1 \leq i \leq k$ und damit $y = z$ gelten, was einen Widerspruch ergibt. Also ist x ein Eckpunkt von \mathcal{M} . ■

Beispiel 1.18

Sei $n = 3$, $c = (7, -2, 5)^\top$ und betrachte $LPmin(A, b, c)$ mit den Nebenbedingungen

(C1) $x_1 - 2x_3 \leq 3$

(C2) $-x_1 + x_2 + x_3 = 2$

(C3) $x_1 - 3x_3 \geq 1$

(C4) $x = (x_1, x_2, x_3)^\top \geq 0$.

Wir benötigen zwei Schlupfvariablen x_4 und x_5 , um das $LPmin$ in Standardform zu bringen. Definiere dazu

$$A := \begin{pmatrix} 1 & 0 & -2 & 1 & 0 \\ -1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 3 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 5},$$

sowie $b := (3, 2, -1)^\top$ und $\mathcal{M} := \{x \in \mathbb{R}^5 : x \geq 0, Ax = b\}$.

Wir erhalten das zu lösende Gleichungssystem

$$\left(\begin{array}{ccccc|c} 1 & 0 & -2 & 1 & 0 & 3 \\ -1 & 1 & 1 & 0 & 0 & 2 \\ -1 & 0 & 3 & 0 & 1 & -1 \end{array} \right) \begin{array}{l} (I) \\ (II) \\ (III) \end{array}$$

für Punkte in \mathcal{M} .

(II) \rightarrow (I) + (II) und (III) \rightarrow (I) + (III) \Rightarrow

$$\left(\begin{array}{ccccc|c} 1 & 0 & -2 & 1 & 0 & 3 \\ 0 & 1 & -1 & 1 & 0 & 5 \\ 0 & 0 & 1 & 1 & 1 & 2 \end{array} \right) \begin{array}{l} (I) \\ (IV) \\ (V) \end{array}$$

(I) \rightarrow (I) + 2 · (V) und (IV) \rightarrow (IV) + (V) \Rightarrow

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & 3 & 2 & 7 \\ 0 & 1 & 0 & 2 & 1 & 7 \\ 0 & 0 & 1 & 1 & 1 & 2 \end{array} \right) \quad (1.1)$$

Hier lässt sich eine Lösung von $Ax = b$ sofort ablesen, nämlich

$$x = (x_1, x_2, x_3, x_4, x_5)^\top = (7, 7, 2, 0, 0)^\top.$$

Nach Satz 1.17.(ii) ist dieses x zudem ein Eckpunkt der zulässigen Bereiches, denn die ersten drei Spalten der Matrix in (1.1) sind offenbar linear unabhängig und $x_4 = x_5 = 0$.

Wir erkennen also, dass sich eine zulässige Ecke dadurch finden lässt, dass wir A zunächst in die sogenannte Basisform bringen.

Definition 1.19

(i) Eine k -zeilige Matrix $A \in \mathbb{R}^{k \times \ell}$ mit $\ell \geq k$ besitzt Basisform, falls in A alle k Einheitsspalten der Form

$$\begin{aligned} &(1, 0, 0, \dots, 0)^\top, \\ &(0, 1, 0, \dots, 0)^\top, \dots, \\ &(0, 0, \dots, 1, 0)^\top, \\ &(0, 0, \dots, 0, 1)^\top \in \mathbb{R}^k \end{aligned}$$

vorkommen.

(ii) Angenommen, die Matrix $A \in \mathbb{R}^{m \times n+m}$ mit $\text{rg}(A) = m$ eines LPs in Standardform liegt in Basisform vor. Dann heißt eine Lösung x von $Ax = b$ eine Basislösung, falls

$$x_i = \begin{cases} b_j, & A^{(i)} \text{ ist Einheitsspalte mit der Eins in Koordinate } j, \\ 0, & \text{sonst.} \end{cases}$$

Nach Satz 1.17 ist ein solches x eine Ecke des zulässigen Bereichs. Die x_i mit $x_i > 0$ heißen Basisvariablen.

Ziel des Simplexverfahrens ist es nun, durch wiederholten Basiswechsel die Ecken des zulässigen Bereichs effizient nach dem Optimum der Zielfunktion abzusuchen.

Wir liefern hier keine vollständige Beschreibung des Simplexverfahrens, sondern veranschaulichen es lediglich an dem bereits bekannten Beispiel 1.1. Gängige Computersysteme wie MATLAB enthalten zuverlässige Implementierungen des Simplexverfahrens.

Beispiel 1.20 (Beispiel 1.1 fortgesetzt)

Das Farbenproduktionsbeispiel ist gegeben durch

$$f(x_a, x_i) = 3x_a + 2x_i = (3, 2) \begin{pmatrix} x_a \\ x_i \end{pmatrix} \stackrel{!}{=} \max$$

unter den Nebenbedingungen $x_a, x_i \geq 0$ sowie

$$x_a + 2x_i \leq 6,$$

$$2x_a + x_i \leq 8,$$

$$-x_a + x_i \leq 1,$$

$$x_i \leq 2.$$

Wir benötigen also vier Schlupfvariablen y_1, y_2, y_3, y_4 . In Standardform ist der zulässige Bereich daher gegeben durch

$$A \cdot (x_a, x_i, y_1, y_2, y_3, y_4)^\top = (6, 8, 1, 2)^\top,$$

wobei $x_a, x_i, y_1, y_2, y_3, y_4 \geq 0$ und

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

Wir erkennen, dass A in Basisform vorliegt, wobei y_1, y_2, y_3 und y_4 die Basisvariablen sind, wenn als naheliegende Startecke (bzw. Basislösung) $e^{(0)} := (0, 0, 6, 8, 1, 2)^\top$ gewählt wird. Wir tragen diese Informationen in das sogenannte Simplextableau ein.

Basis	f	x_a (Pivot)	x_i	y_1	y_2	y_3	y_4	Lösung
f	1	-3	-2	0	0	0	0	0
y_1	0	1	2	1	0	0	0	6
y_2	0	2	1	0	1	0	0	8
y_3	0	-1	1	0	0	1	0	1
y_4	0	0	1	0	0	0	1	2

In der ersten Spalte notieren wir, welche vier Variablen als Basisvariablen der aktuellen Ecke dienen. Die mit „ f “ gekennzeichnete Zeile bedeutet

$$f(x) - 3x_a - 2x_i = 0 \iff f(x) = 3x_a + 2x_i,$$

wobei $x = (x_a, x_i, y_1, y_2, y_3, y_4)^\top$. Die mit „ f “ gekennzeichnete Spalte gibt die Beiträge der (aktuellen) Basisvariablen zur Zielfunktion an. Die restlichen Zeilen kodieren den zulässigen Bereich,

enthalten also die Matrix A und die rechte Seite b . Mit der Startecke $e^{(0)}$ erzielen wir einen Erlös von 0 GE. Wir sollten also eine „bessere Ecke“ finden können. Dies lässt sich daran ablesen, dass sich in der „ f “-Zeile negative Einträge für Nicht-Basisvariablen finden. Der Erlös wird sich also steigern, wenn diese Variablen zu Basisvariablen werden. Daraus ergibt sich das

Abbruchkriterium:

Haben im Falle der Maximierung (bzw. Minimierung) alle Nicht-Basisvariablen nur noch nicht-negative (bzw. nicht-positive) Einträge in der „ f “-Zeile, so ist die aktuelle Basislösung eine optimale Ecke.

Nun wird iterativ, solange das Abbruchkriterium noch nicht erreicht ist, eine Basisvariable ausgetauscht. Dabei sucht man jeweils nach der optimalen Verbesserung der Zielfunktion. Wir nehmen also x_a als neue Basisvariable auf, da hier in der „ f “-Zeile der betragsmäßig größere Eintrag steht. Die zu x_a gehörige Spalte nennt man Pivotspalte.

Nun muss noch die zu entfernende Basisvariable bestimmt werden, das heißt, die Pivotzeile. Dazu berechnet man für alle Basisvariablen mit positivem Eintrag in der Pivotspalte deren relativen Beitrag zur Zielfunktion, also den Quotienten aus ihrem Eintrag in der Lösungsspalte und ihrem Eintrag in der Pivotspalte. Der minimale Wert dieser Quotienten liefert die Pivotzeile.

Hier ist:

$$\text{quotient}(y_1) = 6/1 = 6,$$

$$\text{quotient}(y_2) = 8/2 = 4.$$

Also wird y_2 als Basisvariable entfernt und durch x_a ersetzt. Nun erfolgt die Aktualisierung des Tableaus wie folgt:

1. Die Pivotzeile wird durch das Pivotelement (also das Element der Pivotzeile, das zur Pivotspalte gehört) dividiert, und es wird in der „Basis“-Spalte die neue zugehörige Basisvariable notiert. Wie nennen diese aktualisierte Zeile die neue Pivotzeile (NPZ).
2. Alle anderen Zeilen (auch die „ f “-Zeile) werden gemäß der Formel

$$NZ = AZ - PK \times NPZ$$

aktualisiert, wobei

NZ: neue Zeile,

AZ: alte Zeile,

PK: Koeffizient der alten Zeile in der Pivotspalte.

In unserem Beispiel erhalten wir das folgende aktualisierte Simplextableau.

Basis	f	x_a	x_i (Pivot)	y_1	y_2	y_3	y_4	Lösung
f	1	0	-1/2	0	3/2	0	0	12
y_1	0	0	3/2	1	-1/2	0	0	2
(NPZ) x_a	0	1	1/2	0	1/2	0	0	4
y_3	0	0	3/2	0	1/2	1	0	5
y_4	0	0	1	0	0	0	1	2

Wir erkennen, dass sich der Wert der Zielfunktion von $f(e^{(0)}) = 0$ auf $f(e^{(1)}) = 12$ verbessert hat, wobei die neue Basislösung (nach der ersten Simplexiteration) gegeben ist durch

$$e^{(1)} = (4, 0, 2, 0, 5, 2)^\top$$

mit zugehörigen Basisvariablen x_a, y_1, y_3, y_4 .

Wir benötigen indes noch (mindestens) eine zweite Iteration, denn in den neuen „f“-Zeile hat die Nicht-Basisvariable x_i einen negativen Eintrag, verspricht also eine Verbesserung der Zielfunktion. Die neue Pivotspalte ist also die, die zu x_i gehört. Wir ermitteln die Pivotzeile:

$$\text{quotient}(y_1) = 2 \cdot 2/3 = 4/3,$$

$$\text{quotient}(x_a) = 4 \cdot 2 = 8,$$

$$\text{quotient}(y_3) = 5 \cdot 2/3 = 10/3,$$

$$\text{quotient}(y_4) = 2/1 = 2.$$

Also wird y_1 ausgetauscht und durch x_i ersetzt. Das nochmals aktualisierte Tableau ergibt sich nach dem zweiten Iterationsschritt wie folgt.

Basis	f	x_a	x_i	y_1	y_2	y_3	y_4	Lösung
f	1	0	0	1/3	4/3	0	0	$12 \frac{2}{3}$
(NPZ) x_i	0	0	1	2/3	-1/3	0	0	4/3
x_a	0	1	0	-1/3	2/3	0	0	10/3
y_3	0	0	0	-1	1	1	0	3
y_4	0	0	0	-2/3	1/3	0	1	2/3

Hier ist die Iteration beendet, denn die neue „f“-Zeile hat nur noch nicht-negative Einträge. Die optimale Ecke ist somit gegeben durch

$$e^{(2)} = (10/3, 4/3, 0, 0, 3, 2/3)^\top$$

mit optimalen Wert $f^* = f(e^{(2)}) = 12 \frac{2}{3}$ der Zielfunktion.

Probe: $f(e^{(2)}) = 3 \cdot 10/3 + 2 \cdot 4/3 = 10 + 8/3 = 10 + 2\frac{2}{3} = 12\frac{2}{3}$,

$$A \cdot e^{(2)} = \begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 10/3 \\ 4/3 \\ 0 \\ 0 \\ 3 \\ 2/3 \end{pmatrix} = \begin{pmatrix} 18/3 = 6 \\ 24/3 = 8 \\ 3/3 = 1 \\ 6/3 = 2 \end{pmatrix}$$

wie gewünscht. Die optimale Lösung $(x_a^*, x_i^*) = (10/3, 4/3)$ ist die selbe wie die unter Beispiel 1.1 mit Hilfe des grafischen Verfahrens ermittelte.

Bemerkung 1.21

- (a) Eine detaillierte Begründung für die Update-Schritte 1) und 2) haben wir hier nicht gegeben. Allerdings haben wir beobachtet, dass diese Update-Schritte genau dem Basiswechsel entsprechen, d.h., dass die aktualisierten Spalten der (jeweils neuen) Basisvariablen gerade die benötigten Einheitsspalten sind. Genau diese Transformation liegt den genannten Update-Schritten zu Grunde (Formel für den Basiswechsel, siehe Lineare Algebra).
- (b) Es existieren eine Reihe von Verfeinerungen und Varianten des Simplexverfahrens, auf die wir indes hier nicht näher eingehen wollen.
- (c) Das Simplexverfahren ist ein Algorithmus, der problemlos in Computersoftware implementiert werden kann.

Kapitel 2

Ganzzahlige lineare Optimierung

Beispiel 2.1 (Beispiel 0.2 fortgesetzt)

Zur Motivation greifen wir Beispiel 0.2 (Anschaffung von Maschinen einer Fluggesellschaft) noch einmal auf. Die relevanten Daten waren wie folgt gegeben.

	Typ A	Typ B	Maximum
Profit pro Jahr [1.000 USD]	200	100	?
Preis pro Maschine [1 Mio. USD]	5	4	24
Wartungsstunden pro Jahr	200	500	1300

Fassen wir dieses Maximierungsproblem als ein LP auf, so können wir es wie folgt formalisieren.

$$f(x_A, x_B) = 200x_A + 100x_B \rightarrow \max!$$

$$(C1) \quad 5x_A + 4x_B \leq 24$$

$$(C2) \quad 200x_A + 500x_B \leq 1300$$

$$(C3) \quad x_A, x_B \in \mathbb{N}_0$$

(x_γ : Anzahl anzuschaffender Maschinen vom Typ $\gamma \in \{A, B\}$).

Ersetzen wir die Nebenbedingung (C3) durch

$$(C3') \quad x_A, x_B \geq 0 \text{ (Lockerung, Relaxation)},$$

so liefert das Simplexverfahren die folgende Lösung:

$$(x_A^*, x_B^*) = (4.8, 0) \quad \text{mit} \quad f^* = f(x_A^*, x_B^*) = 960.$$

Aufrunden zu $(\tilde{x}_A, \tilde{x}_B) = (5, 0)$ verletzt die erste Nebenbedingung (C1), denn die Anschaffung von 5 Maschinen des Typs A würde 25 Mio USD kosten, also das zur Verfügung stehende Budget überschreiten.

Abrunden zu $(\tilde{x}_A, \tilde{x}_B) = (4, 0)$ hält zwar alle Nebenbedingungen ein, reduziert den Wert der Zielfunktion indes merklich auf $800 < 960$.

Durch Ausprobieren erkennt man leicht, dass die (zulässige) Lösung $(x_A^{**}, x_B^{**}) = (4, 1)$ besser ist als $(\tilde{x}_A, \tilde{x}_B) = (4, 0)$, denn

$$f(x_A^{**}, x_B^{**}) = 900 > 800.$$

Dieses Beispiel zeigt, dass ein Runden der durch das Simplexverfahren gefundenen Lösung des relaxierten LPs im Allgemeinen nicht zur optimalen Lösung des ILPs (integer-valued linear program) führt. Stattdessen muss eine Strategie verfolgt werden, die die in Frage kommenden ganzzahligen Lösungen systematisch absucht. Eine solche Strategie liefert der Branch and Bound-Algorithmus.

Definition 2.2

Unter einem ganzzahligen linearen Optimierungsproblem (ILP) versteht man die Aufgabe, die lineare Zielfunktion $f(x) = c^\top x$ unter den Nebenbedingungen $x \in \mathbb{N}_0^n$ und $Ax \leq b$ zu maximieren. Dabei ist $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ und $A \in \mathbb{R}^{m \times n}$. Man spricht auch von einem kombinatorischen Optimierungsproblem. Wir bezeichnen dieses Ausgangsproblem im Weiteren mit P_0 .

Definition 2.3 (Branching)

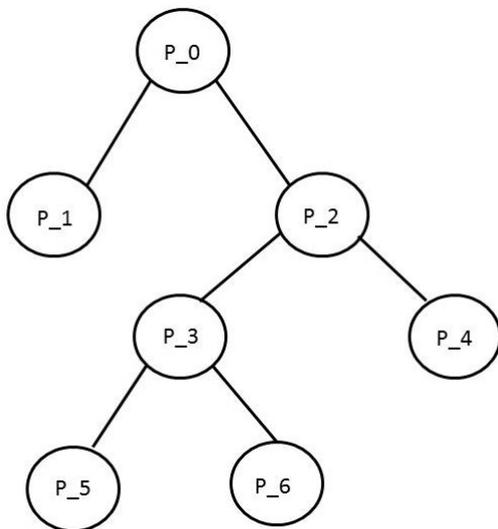
Der Vorgang des Branching beschreibt das Aufteilen eines Problems P_0 in mehrere Teilprobleme $P_1, \dots, P_k, k \in \mathbb{N}$. Man sagt auch, dass P_0 in die Teilprobleme P_1, \dots, P_k verzweigt wird. Dabei sollen die zulässigen Bereiche $\mathcal{M}(P_1), \dots, \mathcal{M}(P_k)$ der Teilprobleme eine Partition des zulässigen Bereichs $\mathcal{M}(P_0)$ des Ausgangsproblems bilden, d. h.

$$\mathcal{M}(P_0) = \bigcup_{i=1}^k \mathcal{M}(P_i) \quad \text{und} \quad \mathcal{M}(P_i) \cap \mathcal{M}(P_j) = \emptyset,$$

für alle $1 \leq i \neq j \leq k$.

In gleicher Weise lässt sich im Weiteren jedes der Teilprobleme seinerseits verzweigen, so dass ein Lösungsbaum entsteht, der alle Teilprobleme enthält; siehe die folgende schematische Darstellung mit $k = 6$.

Schema 2.4 (Lösungsbaum)



Bemerkung 2.5

Eine naheliegende Branching-Regel für ein Problem P_i lautet: Wähle eine ganzzahlige Variable x_j , deren Wert im optimalen Punkt der Relaxierung P_i' von P_i nicht ganzzahlig ist, und bilde die beiden Teilprobleme P_i^1 und P_i^2 , die aus P_i durch Hinzufügen der Nebenbedingung $x_j \leq \lfloor x_j^* \rfloor$ bzw. $x_j \geq \lfloor x_j^* \rfloor + 1$ entstehen. Dabei ist x_j^* der optimale Wert für x_j in P_i' .

Definition 2.6 (Bounding)

Beim Abarbeiten des Lösungsbaums wird stets eine für alle Teilprobleme gültige (globale) untere Schranke \underline{f} für den optimalen Zielfunktionswert f^* und eine für ein spezielles Teilproblem P_i gültige (lokale) obere Schranke \overline{f}_i mitgeführt. Gilt $\overline{f}_i \leq \underline{f}$, so kann \underline{f} durch Auswertung der Zielfunktion an Punkten aus $\mathcal{M}(P_i)$ nicht verbessert werden, so dass P_i nicht weiter verzweigt zu werden braucht. Findet man während des Verfahrens einen zulässigen Punkt mit größerem Zielfunktionswert als \underline{f} , so aktualisiert man \underline{f} auf diesen größeren Wert.

Beispiel 2.7

Wir maximieren f , gegeben durch $f(x, y) = 2x + 3y$, unter den Nebenbedingungen

(C1) $5x + 7y \leq 35$

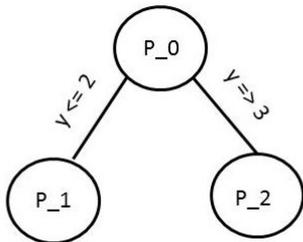
(C2) $4x + 9y \leq 36$

(C3) $x, y \in \mathbb{N}_0$

Im ersten Schritt lösen wir die relaxierte Version P'_0 dieses Ausgangsproblems P_0 , in dem wir (C3) ersetzen durch

(C3') $x, y \geq 0$.

Wir erhalten als optimale Lösung von P'_0 die Maximalstelle $(x^*, y^*) = (3.706, 2.353)$ mit zugehörigem Zielfunktionswert $f^* = 14.471$. Da dieser Wert besser ist als $\underline{f} = 0$ ($(x, y) = (0, 0)$), verzweigen wir P_0 . Wählen wir die Variable y als Verzweigungsvariable, so erhalten wir den folgenden Teilbaum.



Das relaxierte Teilproblem P'_1 ist gegeben durch

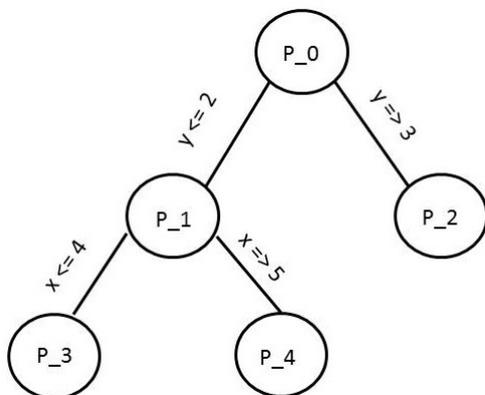
(C1) $5x + 7y \leq 35$

(C2) $4x + 9y \leq 36$

(C3) $y \leq 2$

(C4) $x \geq 0, y \geq 0$

Wir erhalten als optimale Lösung von P'_1 die Maximalstelle $(x^*, y^*) = (4.2, 2)$ mit optimalem Zielfunktionswert $f^* = 14.4$. Also muss P_1 weiter verzweigt werden, denn der Punkt (x^*, y^*) ist nicht zulässig für P_0 . Der erweiterte Lösungsbaum sieht nun wie folgt aus.



Als Lösung des relaxierten Problems $P_3^!$ erhalten wir $(x^*, y^*) = (4, 2)$ mit $f^* = 14$. Da $(x^*, y^*) = (4, 2)$ zulässig für P_0 ist, aktualisieren wir die untere Schranke für den Zielfunktionswert auf $\underline{f} = 14$.

Als Lösung von $P_4^!$ ergibt sich $(x^*, y^*) = (5, 1.43)$ mit $f^* = 14.29$. Wir prüfen also die Punkte $f(5, 1) = 13 < \underline{f}$ und $f(5, 0) = 10 < \underline{f}$. Damit ist der Ast $y \leq 2$ komplett abgearbeitet mit der (lokal) optimalen Lösung (x^*, y^*) und $f^* = \underline{f} = 14$.

Bleibt, Teilproblem P_2 bzw. seine relaxierte Version $P_2^!$ zu lösen. Wir erhalten $(x^*, y^*) = (2.25, 3)$ mit $f^* = 13.5 < \underline{f}$. Somit braucht P_2 nicht weiter verzweigt zu werden und das Verfahren stoppt mit der (global) optimalen Lösung $(x^*, y^*) = (4, 2)$ und $f^* = 14$ von P_0 .

Bemerkung 2.8

- (a) Mit dem Branch and Bound-Algorithmus können auch gemischt-ganzzahlige LPs gelöst werden, in denen nur einige der Variablen Ganzzahligkeits-Nebenbedingungen unterliegen. Hierzu werden jene Variablen, für die solche Nebenbedingungen nicht gelten, beim Branching unberücksichtigt gelassen.
- (b) Der Branch and Bound-Algorithmus kann auch für Probleme verwendet werden, in denen die Variablen binäre Indikatoren sind, siehe Beispiel 2.9.
- (c) Der Branch and Bound-Algorithmus ist in der MATLAB-Funktion `intlinprog` implementiert.
- (d) Die generelle Idee des Branch and Bound lässt sich auch auf viele andere (nicht-lineare) kombinatorische Optimierungsprobleme anwenden.

Beispiel 2.9 (binäres LP)

Eine Fertigungsgesellschaft hat sich entschieden, eine neue Fabrik entweder in Berlin oder in München zu bauen. Ebenso wird erwogen, ein neues Lagerhaus in derjenigen Stadt zu bauen, in der auch die Fabrik gebaut wird. Die zur Entscheidungsfindung relevanten Daten sind in der folgenden Tabelle zusammengefasst.

Bauobjekt	Entscheidungsvariable	Kapitalwert [Mio EUR]	notwendiges Kapital [Mio EUR]
Fabrik in Berlin	x_1	7	20
Fabrik in München	x_2	5	15
Lager in Berlin	x_3	4	12
Lager in München	x_4	3	10

Das maximal verfügbare Kapital sei 25 Mio. EUR. Die Entscheidungsvariablen sind hier binär, wobei

$$x_j = \begin{cases} 1, & \text{Entscheidung ist „ja“,} \\ 0, & \text{Entscheidung ist „nein“,} \end{cases}$$

$1 \leq j \leq 4$. Das Zielkriterium ist die Maximierung des Kapitalwertes (abgezinste Überschüsse abzüglich des Kapitaleinsatzes), d. h.,

$$f(x_1, x_2, x_3, x_4) = 7x_1 + 5x_2 + 4x_3 + 3x_4.$$

Neben den offensichtlichen Nebenbedingungen wie z. B.

$$20x_1 + 15x_2 + 12x_3 + 10x_4 \leq 25$$

(Kapitalgrenze) müssen hier auch die Abhängigkeiten zwischen den Entscheidungsvariablen (manche Entscheidungen bedingen einander bzw. schließen sich gegenseitig aus) durch geeignete Nebenbedingungen kodiert werden. Es ergeben sich daraus die folgenden Constraints:

$$x_1 + x_2 = 1 \quad (\text{Entweder Berlin oder München})$$

$$x_3 + x_4 \leq 1 \quad (\text{Höchstens ein Lagerhaus})$$

$$-x_1 + x_3 \leq 0 \quad (\text{Lager in Berlin nur, falls auch Fabrik in Berlin})$$

$$-x_2 + x_4 \leq 0 \quad (\text{Lager in München nur, falls auch Fabrik in München})$$

$$x_j \leq 1 \quad \forall 1 \leq j \leq 4$$

$$x = (x_1, x_2, x_3, x_4)^\top \in \mathbb{N}_0^4.$$

Damit ist das Problem in ein ILP überführt werden.

Kapitel 3

Nichtlineare Optimierung

3.1 Lokale Extrema ohne Nebenbedingungen

Die Aufgabe dieses Abschnittes ist es, notwendige und hinreichende Bedingungen dafür zu finden, dass $x^* \in \mathcal{M}$ die Funktion

$$f : \mathbb{R}^n \supseteq \mathcal{M} \rightarrow \mathbb{R}$$
$$x \in \mathcal{M} \mapsto f(x) \in \mathbb{R}$$

lokal maximiert bzw. minimiert. D. h., dass in einer Umgebung von x^* die Funktionswerte $f(x)$ stets nicht größer bzw. nicht kleiner als $f(x^*)$ sind.

Definition 3.1

Sei $n \in \mathbb{N}$. Wir bezeichnen mit $\mathcal{O}(\mathbb{R}^n)$ das System der offenen Mengen des \mathbb{R}^n (bezüglich der üblichen Topologie). In der Folge wird stets angenommen, dass $\mathcal{M} \in \mathcal{O}(\mathbb{R}^n)$ ist. Ferner benutzen wir die Bezeichnungen

$$\mathcal{C}_n^0(\mathcal{M}) : = \{f : \mathcal{M} \rightarrow \mathbb{R} : f \text{ stetig}\},$$
$$\mathcal{C}_n^1(\mathcal{M}) : = \{f : \mathcal{M} \rightarrow \mathbb{R} : \forall x \in \mathcal{M} \text{ und } \forall 1 \leq i \leq n \exists D_i f(x) \text{ und ist stetig}\},$$

wobei $D_i f(x_0)$ die i -te partielle Ableitung von f an der Stelle $x_0 \in \mathcal{M}$ bezeichnet, d.h.,

$$D_i f(x_0) = \left. \frac{\partial f(x)}{\partial x_i} \right|_{x=x_0}, \quad 1 \leq i \leq n, \quad x = (x_1, \dots, x_n)^\top.$$

Können wir für f an der Stelle x_0 alle n partiellen Ableitungen bilden, so bezeichnen wir mit

$$\nabla f(x_0) := (D_1 f(x_0), \dots, D_n f(x_0))^\top$$

den Gradienten von f an der Stelle x_0 . Ist $\nabla f(x_0)$ für alle Punkte $x_0 \in \mathcal{M}$ stetig, so ist $f \in \mathcal{C}_n^1(\mathcal{M})$ und wir schreiben auch $D^1 f(x_0)$ statt $\nabla f(x_0)$, $x_0 \in \mathcal{M}$. Falls für alle $1 \leq i \leq n$ die Abbildung

$$D_i f : \mathcal{M} \rightarrow \mathbb{R} \\ x \mapsto D_i f(x) \in \mathbb{R}$$

ebenfalls wieder partielle Ableitungen besitzt, so bezeichnen wir die partielle Ableitung von $D_i f$ nach der j -ten Komponente an der Stelle $x_0 \in \mathcal{M}$ mit $D_{i,j} f(x_0)$ und nennen

$$Hf(x_0) = \begin{pmatrix} D_{1,1}f(x_0) & \cdots & D_{1,n}f(x_0) \\ \vdots & \ddots & \vdots \\ D_{n,1}f(x_0) & \cdots & D_{n,n}f(x_0) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

die Hesse-Matrix von f an der Stelle $x_0 \in \mathcal{M}$. Falls für alle $1 \leq i, j \leq n$ die zweiten partiellen Ableitungen an der Stelle x_0 stetig sind, so nennen wir f an der Stelle x_0 zweimal stetig differenzierbar und schreiben auch $D^2 f(x_0)$ statt $Hf(x_0)$. Die Menge aller auf \mathcal{M} zweimal stetig differenzierbare Funktionen wird bezeichnet mit

$$\mathcal{C}_n^2(\mathcal{M}) := \{f : \mathcal{M} \rightarrow \mathbb{R} : \forall x \in \mathcal{M} \text{ und } \forall 1 \leq i, j \leq n \exists D_{i,j} f(x) \text{ und ist stetig}\}.$$

Analog verfahren wir für höhere Ableitungen und schreiben

$$\mathcal{C}_n^k(\mathcal{M}) := \{f : \mathcal{M} \rightarrow \mathbb{R} : \forall x \in \mathcal{M} \text{ und } \forall 1 \leq i_1, \dots, i_k \leq n \exists D_{i_1, \dots, i_k} f(x) \text{ und ist stetig}\},$$

$k \geq 1$. Abschließend definieren wir

$$\mathcal{C}_n^\infty(\mathcal{M}) := \bigcap_{k \geq 1} \mathcal{C}_n^k(\mathcal{M})$$

als die Menge aller beliebig oft stetig differenzierbaren (reellwertigen) Funktionen auf $\mathcal{M} \in \mathcal{O}(\mathbb{R}^n)$.

Satz 3.2 (Mehrdimensionale Taylor-Formel, Analysis II)

Sei $K \in \mathbb{N}$, $f \in \mathcal{C}_n^{K+1}(\mathcal{M})$ und seien $x, y \in \mathcal{M}$ derart, dass die Strecke zwischen x und y in \mathcal{M} erhalten ist, d. h.,

$$\forall 0 \leq \alpha \leq 1 : x + \alpha(y - x) \in \mathcal{M}.$$

Dann gibt es ein $0 < \gamma < 1$ derart, dass

$$f(y) = \sum_{k=0}^K \frac{1}{k!} (D^1 f(x)(y-x))^k + \frac{1}{(K+1)!} (D^1 f(x + \gamma(y-x))(y-x))^{K+1}. \quad (3.1)$$

In (3.1) bezeichnet

$$(D^1 f(x)(y-x))^0 := f(x) \text{ und}$$

$$(D^1 f(x)(y-x))^k := \sum_{1 \leq i_1, \dots, i_k \leq n} D_{i_1, \dots, i_k} f(x) (y_{i_1} - x_{i_1}) \cdots (y_{i_k} - x_{i_k}), \quad 1 \leq k \leq K+1.$$

Definition 3.3

Sei $A = (a_{ij})_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}$ eine $(n \times n)$ -Matrix. Dann heißt A positiv (semi-)definit, falls für alle $x \in \mathbb{R}^n \setminus \{0\}$ die quadratische Form

$$x^\top Ax = \sum_{1 \leq i, j \leq n} a_{ij} x_i x_j > (\geq) 0 \quad (3.2)$$

ist. Die Matrix A heißt negativ (semi-) definit, falls in (3.2) statt $> (\geq)$ die Relation $< (\leq)$ gilt.

Satz 3.4 (hinreichende Kriterien für lokale Extremstellen)

Sei $f \in \mathcal{C}_n^2(\mathcal{M})$ und $x^* \in \mathcal{M}$ ein Punkt mit den folgenden beiden Eigenschaften.

- (i) $\nabla f(x^*) = 0$,
- (ii) $D^2 f(x^*)$ ist positiv (bzw. negativ) definit.

Dann besitzt f im Punkte x^* ein striktes lokales Minimum (bzw. Maximum).

Beweis: Angenommen, $D^2 f(x^*)$ ist positiv definit. Da $f \in \mathcal{C}_n^2(\mathcal{M})$ und somit zweimal stetig differenzierbar ist, ist auch $D^2 f(x)$ positiv definit für alle x aus einer Umgebung von x^* . Da $\mathcal{M} \in \mathcal{O}(\mathbb{R}^n)$ ist, gibt es eine ε -Kugel $K_\varepsilon(x^*)$ um den Punkt x^* , die vollständig in \mathcal{M} enthalten ist. Für alle $x \in K_\varepsilon(x^*)$ liegt die Verbindungsstrecke zwischen x und x^* in $K_\varepsilon(x^*)$, und somit erst recht in \mathcal{M} . Also existiert eine offene Umgebung $\mathcal{U} \subseteq \mathcal{M}$ von x^* mit den Eigenschaften

- (a) $\forall x \in \mathcal{U}$ ist $D^2 f(x)$ positiv definit,
- (b) $\forall x \in \mathcal{U}$ liegt die Verbindungsstrecke von x und x^* vollständig in \mathcal{U} .

Aus der Taylor-Formel (3.1) erhalten wir somit für $x \in \mathcal{U}$, dass

$$\begin{aligned} f(x) &= f(x^*) + (D^1 f(x^*)(x - x^*))^1 + \frac{1}{2} (D^1 f(x^* + \gamma(x - x^*))(x - x^*))^2 \\ &= f(x^*) + \sum_{i=1}^n D_i f(x^*)(x_i - x_i^*) + \frac{1}{2} \sum_{1 \leq i, j \leq n} D_{i,j} f(x^* + \gamma(x - x^*))(x_i - x_i^*)(x_j - x_j^*) \end{aligned}$$

für ein geeignetes $0 < \gamma < 1$. Nach Voraussetzung (i) ist die Einfachsumme identisch gleich Null. Ferner ist wegen (a) und (b) jeder Summand der Doppelsumme echt größer als Null, falls $x \neq x^*$ ist. Also ist $\forall x \in \mathcal{U}$ mit $x \neq x^* : f(x) > f(x^*)$. Das heißt aber gerade, dass f an der Stelle x^* ein striktes lokales Minimum besitzt. Der Beweis für den Fall eines lokales Maximums verläuft analog. ■

Definition 3.5 (Richtungsableitung)

Sei $f \in \mathcal{C}_n^1(\mathcal{M})$, $x_0 \in \mathcal{M}$ und $d \in \mathbb{R}^n$ mit $\|d\|_2 = 1$. Dann heißt

$$\begin{aligned} \psi_{d, x_0} : \mathbb{R}_{\geq 0} &\rightarrow \mathbb{R} \\ t &\mapsto f(x_0 + td) \end{aligned}$$

eindimensionale Einschränkung von f und $\psi'_{d,x_0} = \langle \nabla f(x_0), d \rangle$ die Richtungsableitung von f in Richtung d an der Stelle x_0 .

Lemma 3.6

Sei $f \in C_n^1(\mathbb{R}^n)$ und $x_0 \in \mathbb{R}^n$ ein Punkt mit $\nabla f(x_0) \neq 0$. Dann zeigt $\nabla f(x_0)$ in die Richtung des steilsten Anstiegs von f (vom Punkte x_0 aus gesehen), und $-\nabla f(x_0)$ in die Richtung des steilsten Abstiegs bzw. Abfalls.

Beweis: Nach der Cauchy-Schwarz'schen Ungleichung gilt für jede Richtung $d \in \mathbb{R}^n, \|d\|_2 = 1$, dass

$$-\|\nabla f(x_0)\|_2 \leq \langle \nabla f(x_0), d \rangle \leq \|\nabla f(x_0)\|_2 \tag{3.3}$$

ist. Offenbar wird die untere Schranke in (3.3) für $\underline{d} = -\frac{\nabla f(x_0)}{\|\nabla f(x_0)\|_2}$ angenommen, und die obere Schranke für $\bar{d} = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|_2}$. Also maximiert (minimiert) die Richtung $(-)\frac{\nabla f(x_0)}{\|\nabla f(x_0)\|_2}$ die Richtungsableitung (also die Steigung im Punkte x_0 in Richtung d) von f über $d \in \mathbb{R}^n, \|d\|_2 = 1$. Die maximale (minimale) Steigung von f im Punkte x_0 ist gegeben durch $(-)\|\nabla f(x_0)\|_2$. ■

Korollar 3.7 (Fermat'sche Regel, notwendiges Kriterium für lokale Extremstellen)

Sei $f \in C_n^1(\mathbb{R}^n)$ und sei $x^* \in \mathbb{R}^n$ eine lokale Extremstelle von $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Dann ist x^* ein kritischer Punkt von f , d. h., es gilt $\nabla f(x^*) = 0$.

Beweis: Sei $x^* \in \mathbb{R}^n$ mit $\nabla f(x) \neq 0$. Dann gibt es nach Lemma 3.6 Richtungen \bar{d} und \underline{d} , in die man $(x, f(x))$ mit echt wachsenden bzw. echt fallenden Funktionswerten verlassen kann. Also kommen solche Punkte x als lokale Extremstellen nicht in Frage. ■

Bemerkung 3.8

Mit Hilfe von Satz 3.4 und Korollar 3.7 ergibt sich für $f \in C_n^2(\mathcal{M})$ die folgende Strategie zum Ermitteln lokaler Extremstellen.

1. Löse das Gleichungssystem $\nabla f(x) = 0$ nach x , um die kritischen Punkte von f zu finden.
2. Für jeden kritischen Punkt $x_0 \in \mathcal{M}$ von f prüfe $Hf(x_0)$ auf Definitheit.

Zur praktischen Umsetzung des zweiten Schrittes ist das folgende Resultat aus der linearen Algebra hilfreich.

Lemma 3.9

Sei A eine $(n \times n)$ -Matrix. Dann sind die folgenden Aussagen äquivalent:

- (i) A ist positiv definit,

(ii) Für alle $1 \leq \ell \leq n$ ist die Determinante von $\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,\ell} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\ell,1} & a_{\ell,2} & \cdots & a_{\ell,\ell} \end{pmatrix}$ positiv.

Bemerkung 3.10

- (a) A ist genau dann negativ definit, wenn $-A$ positiv definit ist.
- (b) Im Falle $n = 2$ erhalten wir aus Lemma 3.9: Die Matrix $A = (a_{ij})_{1 \leq i, j \leq 2}$ ist genau dann positiv definit, wenn $a_{11} > 0$ und $a_{11}a_{22} - a_{12}a_{21} > 0$ ist. Sie ist genau dann negativ definit, wenn $a_{11} < 0$ und $a_{11}a_{22} - a_{12}a_{21} > 0$ ist.
- (c) Verschwindet für $f \in \mathcal{C}_2^2(\mathcal{M})$ in einem Punkte $x_0 \in \mathcal{M}$ der Gradient und gilt darüber hinaus $D_{1,1}f(x_0)D_{2,2}f(x_0) - D_{2,1}f(x_0)D_{1,2}f(x_0) < 0$, so liegt in x_0 kein lokales Extremum vor.

Beispiel 3.11

Betrachte

$$\begin{aligned} f : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto f(x, y) = (x + y)^3 - 12xy. \end{aligned}$$

Man erkennt sofort, dass $f \in \mathcal{C}_2^2(\mathbb{R}^2)$ ist. Der Gradient von f ist gegeben durch

$$\begin{aligned} \nabla f(x, y) &= (D_1f(x, y), D_2f(x, y))^\top \\ &= (3(x + y)^2 - 12y, 3(x + y)^2 - 12x)^\top. \end{aligned}$$

Aus $D_1f(x, y) = D_2f(x, y) = 0$ ergibt sich für kritische Punkte von f sofort die Bedingung $x = y$. Lösen wir

$$12x^2 - 12x = 0 \Leftrightarrow x(x - 1) = 0,$$

so erkennen wir, dass es nur die beiden kritischen Punkte $P = (0, 0)^\top$ und $Q = (1, 1)^\top$ von f gibt. Ferner gilt für die Hesse-Matrix von f , dass

$$D^2f(x, y) = \begin{pmatrix} 6(x + y) & 6(x + y) - 12 \\ 6(x + y) - 12 & 6(x + y) \end{pmatrix}$$

ist. Somit ist $\det(D^2f(0, 0)) = -144$ und P nach Teil (c) von Bemerkung 3.10 also keine lokale Extremstelle von f . Allerdings ist $\det(D^2f(1, 1)) = 144 - 0 = 144 > 0$ und $D_{1,1}f(1, 1) = 12 > 0$. Also besitzt f an der Stelle Q ein lokales Minimum, nach Teil (b) von Bemerkung 3.10.

3.2 Extrema konvexer und konkaver Funktionen (ohne Nebenbedingungen)

Definition 3.12

Eine Funktion

$$f : \mathcal{M} \rightarrow \mathbb{R}$$

$$x \mapsto f(x) \in \mathbb{R}$$

mit $\mathcal{M} \subseteq \mathbb{R}^n$ besitzt im Punkt $x^* \in \mathcal{M}$ eine globale Maximalstelle bzw. Minimalstelle, falls $\forall x \in \mathcal{M}$ gilt: $f(x) \leq f(x^*)$ bzw. $f(x) \geq f(x^*)$. Eine globale Maximalstelle bzw. Minimalstelle x^* von f heißt strikt, falls $\forall x \in \mathcal{M}$ mit $x \neq x^*$ gilt: $f(x) < f(x^*)$ bzw. $f(x) > f(x^*)$.

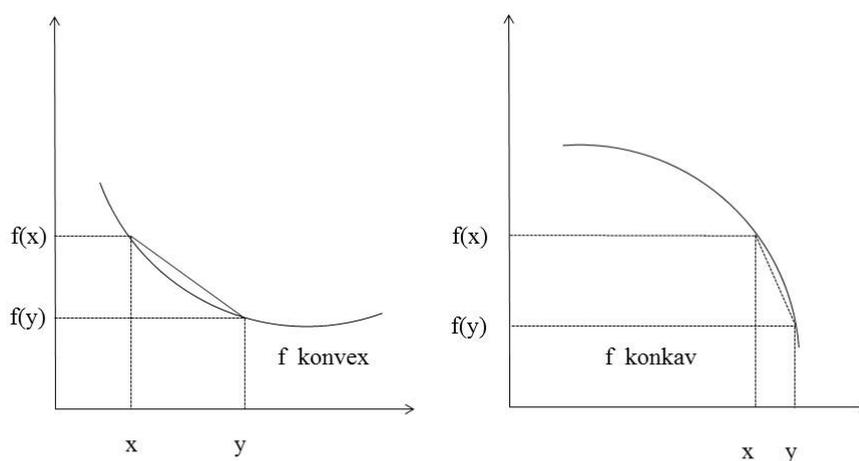
Definition 3.13

Sei $C \subseteq \mathbb{R}^n$ eine konvexe Menge und $f : C \rightarrow \mathbb{R}$ eine reellwertige Funktion. Falls $\forall x, y \in C$ mit $x \neq y$ und für alle $\alpha \in [0, 1]$:

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \quad (3.4)$$

gilt, so heißt f konvex. Gilt in (3.4) sogar $<$ statt \leq für alle $0 < \alpha < 1$, so heißt f strikt konvex. Falls $-f$ eine (strikt) konvexe Funktion ist, so heißt f (strikt) konkav.

Skizze 3.14 ($n = 1$)



Beispiel 3.15

Sei $n = 1$ und

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto f(x) = x^2 \in \mathbb{R}. \end{aligned}$$

Dann gilt $\forall x \neq y \in \mathbb{R}$ und $\forall 0 < \alpha < 1$:

$$\begin{aligned} \alpha f(x) + (1 - \alpha)f(y) &= \alpha x^2 + (1 - \alpha)y^2 \\ &= \alpha(x - y)^2 + 2\alpha y(x - y) + y^2 \\ &> \alpha^2(x - y)^2 + 2\alpha y(x - y) + y^2 \\ &= [\alpha(x - y) + y]^2 \\ &= [\alpha x + (1 - \alpha)y]^2 \\ &= f(\alpha x + (1 - \alpha)y). \end{aligned}$$

Also ist f strikt konvex auf \mathbb{R} .

Satz 3.16

Sei $f : C \rightarrow \mathbb{R}$ mit $C \subseteq \mathbb{R}^n$ konvex eine konvexe Funktion. Ferner besitze f im Punkt $x^* \in C$ ein lokales Minimum. Dann ist $x^* \in C$ auch eine globale Minimalstelle von f (auf C).

Beweis: Da f im Punkt x^* ein lokales Minimum besitzt, existiert ein $\varepsilon > 0$ so, dass für alle $x \in K_\varepsilon(x^*) \cap C$ gilt: $f(x) \geq f(x^*)$. Sei nun $x_0 \in C \setminus K_\varepsilon(x^*)$ beliebig gewählt.

zu zeigen: $f(x_0) \geq f(x^*)$.

Da $x_0 \notin K_\varepsilon(x^*)$ ist, gilt $\|x_0 - x^*\| \geq \varepsilon$. Wähle nun

$$0 < \alpha < \frac{\varepsilon}{\|x_0 - x^*\|} \leq 1 \tag{3.5}$$

und bilde $\tilde{x} := \alpha x_0 + (1 - \alpha)x^*$. Dann liegt \tilde{x} als Konvexkombination zweier Elemente aus C selbst wieder in C . Ferner ist \tilde{x} sogar in $K_\varepsilon(x^*) \cap C$, denn

$$\|\tilde{x} - x^*\| = \|\alpha x_0 + (1 - \alpha)x^* - x^*\| = \alpha\|x_0 - x^*\| < \varepsilon,$$

wegen (3.5). Da nun aber x^* Minimalstelle auf $K_\varepsilon(x^*) \cap C$ ist, gilt

$$\begin{aligned} f(x^*) \leq f(\tilde{x}) &= f(\alpha x_0 + (1 - \alpha)x^*) \\ &\leq \alpha f(x_0) + (1 - \alpha)f(x^*), \end{aligned}$$

wegen der Konvexität von f . Folglich ist

$$0 \leq \alpha f(x_0) + (1 - \alpha)f(x^*) - f(x^*) = \alpha[f(x_0) - f(x^*)].$$

Da aber $\alpha > 0$ gewählt war, ist

$$0 \leq f(x_0) - f(x^*) \Leftrightarrow f(x^*) \leq f(x_0),$$

was zu zeigen war. ■

Korollar 3.17

Ist, unter den Voraussetzungen von Satz 3.16, f eine konkave Funktion und $x^* \in C$ eine lokale Maximalstelle von f , dann ist x^* eine globale Maximalstelle von f (auf C).

Im Rest dieses Abschnittes werden wir Kriterien dafür definieren, dass eine Funktion f konvex ist.

Lemma 3.18

Sei $f : C \rightarrow \mathbb{R}$ mit $C \subseteq \mathbb{R}^n$ konvex eine Funktion. Dann sind die beiden folgenden Aussagen äquivalent:

(i) f ist konvex.

(ii) Für jede Konvexkombination $\sum_{i=1}^p \alpha_i x^{(i)}$ mit $x^{(i)} \in C$ für alle $1 \leq i \leq p$ gilt

$$f\left(\sum_{i=1}^p \alpha_i x^{(i)}\right) \leq \sum_{i=1}^p \alpha_i f(x^{(i)}).$$

Beweis: Übungsaufgabe.

Satz 3.19

Sei

$$\begin{aligned} f : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto f(x) := x^\top A x \end{aligned}$$

eine positiv semidefinite quadratische Form, d. h., es sei $A \in \mathbb{R}^{n \times n}$ eine positiv semidefinite Matrix. Dann ist f konvex.

Beweis: Seien $x, y \in \mathbb{R}^n$ und $0 \leq \alpha \leq 1$. Dann gilt zunächst

$$\begin{aligned} 0 \leq f(x - y) &= (x - y)^\top A (x - y) \\ &= (x^\top - y^\top) A (x - y) \\ &= x^\top A x - x^\top A y - y^\top A x + y^\top A y \end{aligned}$$

und folglich

$$x^\top A y + y^\top A x \leq x^\top A x + y^\top A y. \tag{3.6}$$

Damit ist

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &= [\alpha x + (1 - \alpha)y]^\top A[\alpha x + (1 - \alpha)y] \\ &= \alpha^2 x^\top A x + \alpha(1 - \alpha)[x^\top A y + y^\top A x] + (1 - \alpha)^2 y^\top A y. \end{aligned}$$

Wir wenden (3.6) an und erhalten, dass

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &\leq \alpha^2 x^\top A x + \alpha(1 - \alpha)[x^\top A x + y^\top A y] + (1 - \alpha)^2 y^\top A y \\ &= \alpha^2 x^\top A x + \alpha x^\top A x - \alpha^2 x^\top A x + (1 - \alpha)[\alpha + (1 - \alpha)]y^\top A y \\ &= \alpha x^\top A x + (1 - \alpha)y^\top A y \\ &= \alpha f(x) + (1 - \alpha)f(y). \end{aligned}$$

■

Korollar 3.20

Sei

$$\begin{aligned} f : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto f(x) = x^\top A x \end{aligned}$$

eine negativ semidefinite quadratische Form. Dann ist f konkav.

Lemma 3.21

Sei $f : C \rightarrow \mathbb{R}$ mit $C \subseteq \mathbb{R}^n$ konvex eine konvexe Funktion. Dann ist für alle $\gamma \in \mathbb{R}$ die Menge

$$\{f \leq \gamma\} = \{x \in C : f(x) \leq \gamma\}$$

konvex.

Beweis: Seien $\gamma \in \mathbb{R}$ und $x, y \in \{f \leq \gamma\}$ beliebig gegeben. Dann gilt für alle $0 \leq \alpha \leq 1$ zunächst, dass $\alpha x + (1 - \alpha)y \in C$, da C eine konvexe Menge ist. Darüber hinaus gilt weiter

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &\leq \alpha f(x) + (1 - \alpha)f(y) \\ &\leq \alpha \gamma + (1 - \alpha)\gamma = \gamma. \end{aligned}$$

Also ist $\alpha x + (1 - \alpha)y$ ebenfalls ein Element von $\{f \leq \gamma\}$ und somit $\{f \leq \gamma\}$ konvex. ■

Definition 3.22

Sei $f : \mathcal{M} \rightarrow \mathbb{R}$ mit $\mathcal{M} \subseteq \mathbb{R}^n$ eine Funktion. Dann heißt

$$gr(f, C) = \{(x, f(x)) : x \in \mathcal{M}\}$$

der Graph von f auf \mathcal{M} . Ferner heißt

$$\text{egr}(f, C) = \{(x, y) : x \in \mathcal{M}, f(x) \leq y\}$$

der Epigraph von f auf \mathcal{M} .

Lemma 3.23

Sei $f : C \rightarrow \mathbb{R}$ mit $C \subseteq \mathbb{R}^n$ konvex eine konvexe Funktion. Dann sind die folgenden beiden Aussagen äquivalent

(i) f ist konvex.

(ii) Der Epigraph $\text{egr}(f, C)$ ist eine konvexe Teilmenge des \mathbb{R}^{n+1} .

Beweis: (i) \Rightarrow (ii):

Wir müssen zeigen, dass $\text{egr}(f, C)$ konvex ist, falls f konvex ist. Seien dazu zwei Punkte (x, y) und (u, v) aus $\text{egr}(f, C)$ sowie $0 \leq \alpha \leq 1$ beliebig vorgegeben. Dann gilt

$$\begin{aligned} f(\alpha x + (1 - \alpha)u) &\leq \alpha f(x) + (1 - \alpha)f(u) \\ &\leq \alpha y + (1 - \alpha)v. \end{aligned}$$

Da $\alpha x + (1 - \alpha)u \in C$ ist, folgt aus der Definition von $\text{egr}(f, C)$ daher, dass

$$(\alpha x + (1 - \alpha)u, \alpha y + (1 - \alpha)v) \in \text{egr}(f, C).$$

Nun ist aber

$$(\alpha x + (1 - \alpha)u, \alpha y + (1 - \alpha)v) = \alpha(x, y) + (1 - \alpha)(u, v).$$

Daher ist $\alpha(x, y) + (1 - \alpha)(u, v) \in \text{egr}(f, C)$, was zu zeigen war.

(ii) \Rightarrow (i):

Seien jetzt $x, u \in C$ und $0 \leq \alpha \leq 1$ beliebig vorgegeben. Aus Definition von $\text{egr}(f, C)$ folgt, dass $(x, f(x))$ und $(u, f(u))$ zwei Elemente aus $\text{egr}(f, C)$ sind. Damit ergibt sich aus (ii), dass

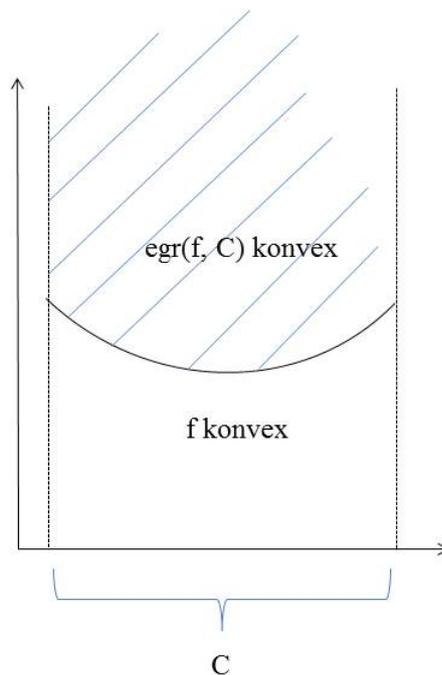
$$(\alpha x + (1 - \alpha)u, \alpha f(x) + (1 - \alpha)f(u)) = \alpha(x, f(x)) + (1 - \alpha)(u, f(u)) \in \text{egr}(f, C).$$

Dies bedeutet nach der Definition von $\text{egr}(f, C)$ aber gerade, dass

$$f(\alpha x + (1 - \alpha)u) \leq \alpha f(x) + (1 - \alpha)f(u)$$

ist. Da x, u und α beliebig gewählt waren, folgt die Konvexität von f . ■

Skizze 3.24 ($n = 1$)



Zum Abschluss dieses Abschnittes geben wir noch analytisch überprüfbare Konvexitätskriterien für stetig differenzierbare Funktionen an.

Lemma 3.25

Sei $C \subseteq \mathbb{R}^n$ eine offene konvexe Menge und sei $f \in C_n^1(C)$ eine stetig differenzierbare Funktion. Dann sind die beiden folgenden Aussagen äquivalent.

- (i) f ist konvex.
- (ii) $\forall x, y \in C$ gilt: $f(x) - f(y) \geq (\nabla f(y))^T(x - y)$.

Beweis: (i) \Rightarrow (ii):

Wegen der Konvexität von f erhalten wir für beliebige Stellen $x, y \in C$ und beliebiges $0 < \alpha < 1$, dass

$$\alpha f(x) + (1 - \alpha)f(y) \geq f(\alpha x + (1 - \alpha)y)$$

ist. Folglich ist

$$\alpha[f(x) - f(y)] \geq f(\alpha x + (1 - \alpha)y) - f(y) \iff f(x) - f(y) \geq \frac{f(\alpha x + (1 - \alpha)y) - f(y)}{\alpha}.$$

Lassen wir nun $\alpha \downarrow 0$ gehen, so ergibt sich

$$f(x) - f(y) \geq \lim_{\alpha \downarrow 0} \frac{f(\alpha x + (1 - \alpha)y) - f(y)}{\alpha} = (\nabla f(y))^\top (x - y)$$

nach Kettenregel.

(ii) \Rightarrow (i):

Seien $x, y \in C$ und $0 < \alpha < 1$ beliebig vorgegeben. Wir setzen $z := \alpha x + (1 - \alpha)y$. Nach Voraussetzung ist $f(x) - f(z) \geq (\nabla f(z))^\top (x - z)$ sowie $f(y) - f(z) \geq (\nabla f(z))^\top (y - z)$. Wir folgern, dass

$$\begin{aligned} \alpha f(x) + (1 - \alpha)f(y) - f(z) &= \alpha[f(x) - f(z)] + (1 - \alpha)[f(y) - f(z)] \\ &\geq \alpha(\nabla f(z))^\top (x - z) + (1 - \alpha)(\nabla f(z))^\top (y - z) \\ &= (\nabla f(z))^\top [\alpha x + (1 - \alpha)y] - (\nabla f(z))^\top [\alpha z + (1 - \alpha)z] \\ &= (\nabla f(z))^\top z - (\nabla f(z))^\top z = 0. \end{aligned}$$

■

Ist f sogar zweimal stetig differenzierbar, so kann ein Konvexitätskriterium an die Hesse-Matrix gestellt werden.

Lemma 3.26

Sei $C \subseteq \mathbb{R}^n$ eine offene konvexe Menge und $f \in \mathcal{C}_n^2(C)$. Dann sind die beiden folgenden Aussagen äquivalent.

(i) f ist konvex.

(ii) Für alle $x \in C$ ist die Hesse-Matrix $Hf(x)$ positiv semidefinit.

Beweis: (i) \Rightarrow (ii):

Seien $x \in C$ und eine Richtung $h \in \mathbb{R}^n$ beliebig gegeben. Da C offen ist, existiert ein $\gamma_0 > 0$ mit $x + \gamma h \in C \forall \gamma \in [0, \gamma_0]$. Wir wenden nun Lemma 3.25 an und erhalten, dass $\forall \gamma \in [0, \gamma_0]$ gilt:

$$f(x + \gamma h) - f(x) - \gamma h^\top \nabla f(x) \geq 0.$$

Aus der mehrdimensionalen Taylor-Formel folgt somit, dass für alle $\gamma \in [0, \gamma_0]$ ein $\lambda \in (0, 1)$ existiert mit

$$h^\top Hf(x + \lambda \gamma h)h \geq 0.$$

Da die Hesse-Matrix nach Voraussetzung stetig ist, folgt für $\gamma \downarrow 0$, dass

$$h^\top Hf(x)h \geq 0.$$

(ii) \Rightarrow (i):

Seien $x, y \in C$ beliebig vorgegeben. Aus der mehrdimensionalen Taylor-Formel folgt, dass ein $\alpha > 0$ existiert mit

$$f(x) - f(y) - (x - y)^\top \nabla f(y) = \frac{1}{2}(x - y)^\top Hf(y + \alpha(x - y))(x - y).$$

Wegen der positiven Semi-Definitheit von Hf auf ganz C folgern wir, dass

$$f(x) - f(y) \geq (x - y)^\top \nabla f(y) = (\nabla f(y))^\top (x - y).$$

Aus Lemma 3.25 folgt somit, dass f auf C konvex ist. ■

Beispiel 3.27

Wir betrachten die Funktion

$$\begin{aligned} f: \mathbb{R}^3 &\rightarrow \mathbb{R} \\ x = (x_1, x_2, x_3)^\top &\mapsto f(x) = x_1^3 + x_2^3 + x_3. \end{aligned}$$

Offenbar ist $f \in \mathcal{C}_3^2(\mathbb{R}^3)$ mit Gradient ∇f und Hesse-Matrix Hf , gegeben durch

$$\begin{aligned} \nabla f(x) &= (3x_1^2, 2x_2, 1)^\top, \\ Hf(x) &= \begin{pmatrix} 6x_1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Damit erhalten wir für $x, z \in \mathbb{R}^3$, dass

$$z^\top Hf(x)z = 6x_1 z_1^2 + 2z_2^2$$

und folglich $z^\top Hf(x)z \geq 0 \forall z \in \mathbb{R}^3 \Leftrightarrow x_1 \geq 0$. Demnach ist f konvex auf der konvexen Teilmenge

$$C = \{x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3 : x_1 \geq 0\}.$$

Dies beschließt den Abschnitt über unrestringierte nicht-lineare Optimierung.

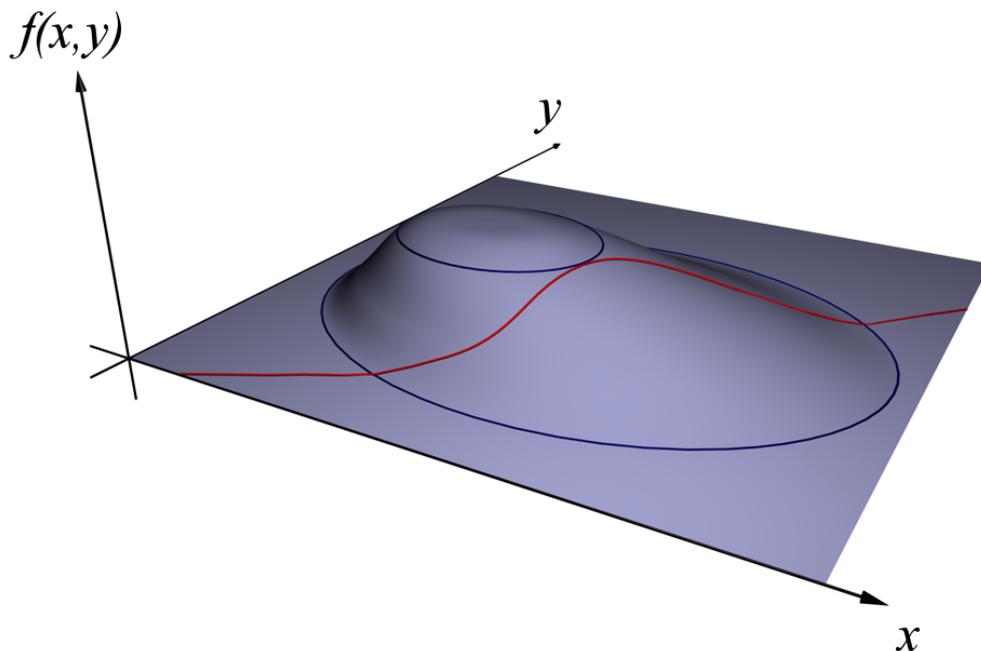
3.3 Nicht-lineare Optimierung unter Nebenbedingungen

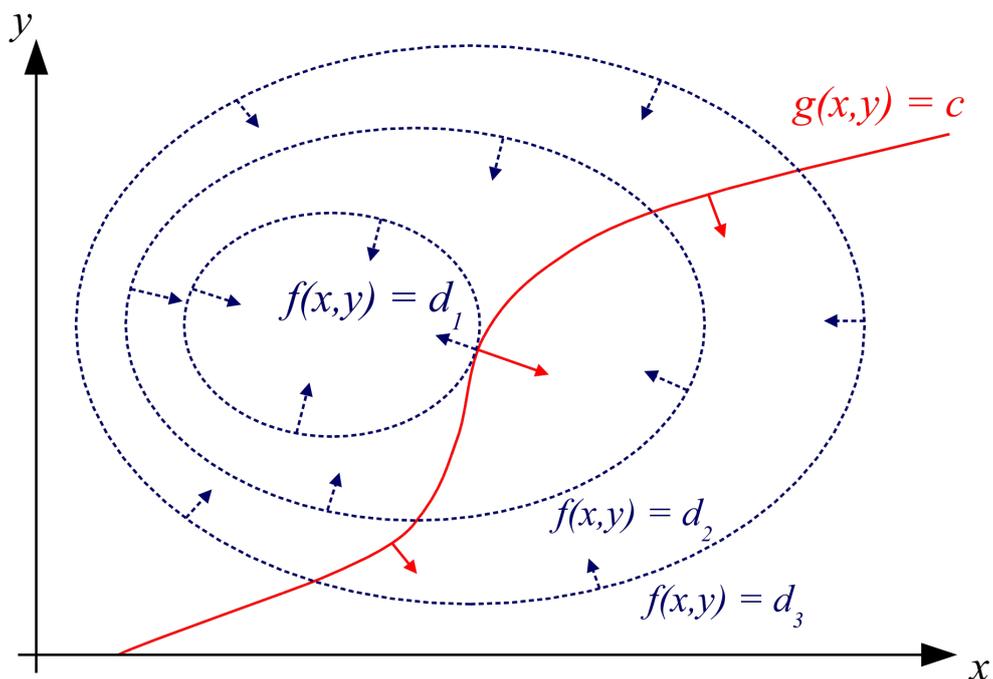
Definition 3.28

Seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ vorgegebene Funktionen und $c \in \mathbb{R}^m$ ein vorgegebener Punkt. Dann besitzt f im Punkt $x^* \in \mathbb{R}^n$ ein lokales Maximum unter der Nebenbedingung $\mathcal{M} = \{x \in \mathbb{R}^n : g(x) = c\}$, falls $x^* \in \mathcal{M}$ ist und es eine ε -Kugel $K_\varepsilon(x^*)$ gibt, so dass für alle $x \in K_\varepsilon(x^*) \cap \mathcal{M}$ gilt: $f(x) \leq f(x^*)$. Gilt statt „ \leq “ sogar „ $<$ “, so heißt x^* strikte lokale Maximalstelle von f auf \mathcal{M} . Analog werden (strikte) lokale Minimalstellen definiert.

Es kann o. B. d. A. $c = 0 \in \mathbb{R}^m$ angenommen werden, denn ansonsten kann $g(\cdot)$ ersetzt werden durch $g(\cdot) - c$.

Heuristik 3.29 ($n = 2, m = 1$)





In Analogie zur graphischen Methode zur Lösung linearer Optimierungsprobleme unter Nebenbedingungen (siehe Kapitel 1) betrachte man für $m = 1$ Höhenlinien von f , d. h., Mengen $\{x \in \mathbb{R}^n : f(x) = d\}$ für verschiedene $d \in \mathbb{R}$. (In der zweiten Graphik sind Höhenlinien von f für $d_1 > d_2 > d_3$ eingezeichnet.) Ziel ist es, d soweit zu vergrößern, bis dass man gerade noch einen zulässigen Punkt $x_0 \in \mathcal{M} \cap \{x \in \mathbb{R}^n : f(x) = d\} = \{x \in \mathbb{R}^n : g(x) = c \text{ und } f(x) = d\}$ findet. Man erkennt grafisch leicht, dass eine Vergrößerung von d so lange möglich ist, bis dass man an einem Punkt x_0 gelangt ist, in dem $\nabla f(x_0)$ parallel zu $\nabla g(x_0)$ ist, d. h., dass ein $\lambda_0 \in \mathbb{R}$ existiert mit

$$\nabla f(x_0) = -\lambda_0 \nabla g(x_0) \iff \nabla f(x_0) + \lambda_0 \nabla g(x_0) = 0 \quad (3.7)$$

(vgl. die Situation bei d_1 in der Grafik). Die Zahl $\lambda_0 \in \mathbb{R}$ heißt Lagrange-Multiplikator. Die Bedingung (3.7), zusammen mit der Bedingung $g(x_0) = c$, lässt sich durch das folgende Gleichungssystem ausdrücken.

$$\nabla L(x, \lambda) = 0 \in \mathbb{R}^{n+1},$$

wobei die Lagrange-Funktion

$$\begin{aligned} L : \mathbb{R}^{n+1} &\rightarrow \mathbb{R} \\ (x, \lambda)^\top &\mapsto L(x, \lambda) \end{aligned}$$

gegeben ist durch $L(x, \lambda) = f(x) + \lambda(g(x) - c)$. Dies erkennt man durch Bildung der partiellen Ableitungen von L , nämlich

$$\frac{\partial L(x, \lambda)}{\partial \lambda} = g(x) - c,$$

$$\frac{\partial L(x, \lambda)}{\partial x_j} = \frac{\partial f(x)}{\partial x_j} + \lambda \frac{\partial g(x)}{\partial x_j}, \quad 1 \leq j \leq n.$$

Definition 3.30 (Jacobi-Matrix)

Sei

$$g : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$x \mapsto g(x) = (g_1(x), \dots, g_m(x))^T \in \mathbb{R}^m$$

eine differenzierbare Funktion. Dann heißt

$$\frac{\partial g}{\partial x}(x) \equiv J_g(x) := \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \cdots & \frac{\partial g_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x) & \cdots & \frac{\partial g_m}{\partial x_n}(x) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Jacobi-Matrix von g an der Stelle x . Falls $m = 1$ ist, so gilt $J_g(x) = (\nabla g(x))^T$.

Satz 3.31 (Satz über implizite Funktionen)

Seien $n > m$ natürliche Zahlen, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Funktion, x_0 ein Punkt im \mathbb{R}^{n-m} , und y_0 ein Punkt im \mathbb{R}^m . Ferner seien $U = U(x_0) \subseteq \mathbb{R}^{n-m}$ und $V = V(y_0) \subseteq \mathbb{R}^m$ offene Umgebungen von x_0 bzw. y_0 derart, dass g in jedem Punkt $(x, y) \in U \times V$ stetig differenzierbar ist. Ferner gelte $\det(J_g(x, y)) \neq 0 \forall (x, y) \in U \times V$ und $g(x_0, y_0) = 0$.

Dann gibt es offene Umgebungen $U_0 \subset U$ mit $x_0 \in U_0$ und $V_0 \subset V$ mit $y_0 \in V_0$ derart, dass durch die Gleichung $g(x, y) = 0$ genau eine Funktion $h : U_0 \rightarrow V_0$ definiert ist mit der Eigenschaft $g(x, h(x)) = 0$ für alle $x \in U_0$. Darüber hinaus ist h stetig differenzierbar, und es gilt:

$$\frac{\partial g}{\partial x}(x, h(x)) + \frac{\partial g}{\partial y}(x, h(x)) \frac{\partial h}{\partial x}(x) = 0. \quad (3.8)$$

Nachweis von (3.8) mit Hilfe der Kettenregel:

Wir betrachten die Funktion

$$\eta : \mathbb{R}^{n-m} \rightarrow \mathbb{R}^n$$

$$x \mapsto (x, h(x)) \in \mathbb{R}^n.$$

Nach Voraussetzung ist $\forall x \in U_0 : g(x, h(x)) = g(\eta(x)) = 0$. Die Kettenregel liefert nun, dass

$$\begin{aligned} 0 &= \frac{\partial g \circ \eta}{\partial x}(x) = J_g(\eta(x))J_\eta(x) \\ &= \left(\frac{\partial g}{\partial x}(x, h(x)), \frac{\partial g}{\partial y}(x, h(x)) \right) \begin{pmatrix} I_{n-m} \\ \frac{\partial h}{\partial x}(x) \end{pmatrix} \\ &= \frac{\partial g}{\partial x}(x, h(x)) + \frac{\partial g}{\partial y}(x, h(x)) \frac{\partial h}{\partial x}(x). \end{aligned}$$

Satz 3.32 (Lagrange-Multiplikatoren)

Seien $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $n > m$ Funktionen, die in einer Umgebung von x_0 stetig differenzierbar seien. Ferner besitze $J_g(x_0)$ Höchstrang m und es gelte $g(x_0) = 0 \in \mathbb{R}^m$. Falls dann f an der Stelle x_0 ein lokales Extremum unter den Nebenbedingungen $g(x) = 0$ besitzt, dann gibt es reelle Zahlen $\lambda_1, \dots, \lambda_m$, die sogenannten Lagrange-Multiplikatoren, so dass gilt:

$$\forall 1 \leq i \leq n : \frac{\partial f}{\partial x_i}(x_0) + \sum_{k=1}^m \lambda_k \frac{\partial g_k}{\partial x_i}(x_0) = 0.$$

Beweis: Wir wenden Satz 3.31 über implizite Funktionen auf die Funktion g an. Da $J_g(x_0)$ den Höchstrang m besitzt, lassen sich m Komponenten von x_0 finden, und zwar o. B. d. A. die ersten m , so dass mit $y_0 := ((x_0)_1, \dots, (x_0)_m)^\top$ gilt:

$$\det \left(\frac{\partial g}{\partial y}(x) \right) \neq 0 \text{ für alle } x \text{ aus einer Umgebung von } x_0. \tag{3.9}$$

Wir bezeichnen im Folgenden

$$\begin{aligned} x_0 &= \begin{pmatrix} y_0 \\ z_0 \end{pmatrix}, \quad y_0 \in \mathbb{R}^m, z_0 \in \mathbb{R}^{n-m}, \\ x &= \begin{pmatrix} y \\ z \end{pmatrix}, \quad y \in \mathbb{R}^m, z \in \mathbb{R}^{n-m}. \end{aligned}$$

Wir beachten, dass $g(y_0, z_0) = 0 \in \mathbb{R}^m$ nach Voraussetzung gilt. Nach Satz 3.31 existiert also eine eindeutig bestimmte Funktion h , die in einer Umgebung $V(z_0) \subseteq \mathbb{R}^{n-m}$ von z_0 definiert ist mit

$$h : V(z_0) \rightarrow \mathbb{R}^m, \quad g(h(z), z) = 0$$

für alle $z \in V(z_0)$ sowie $h(z_0) = y_0$. Darüber hinaus ist h stetig differenzierbar und die Funktion $z \mapsto f(h(z), z)$ besitzt nach Voraussetzung im Punkt z_0 eine lokale Extremstelle, denn

$$f(h(z_0), z_0) = f((y_0, z_0)) = f(x_0).$$

Damit ist aber der Gradient dieser Funktion an der Stelle z_0 gleich Null und es gilt nach der Kettenregel

$$\underbrace{\frac{\partial f}{\partial y}(x_0)}_{\in \mathbb{R}^{1 \times m}} \underbrace{\frac{\partial h}{\partial z}(z_0)}_{\in \mathbb{R}^{m \times n-m}} + \underbrace{\frac{\partial f}{\partial z}(x_0)}_{\in \mathbb{R}^{1 \times n-m}} = 0 \in \mathbb{R}^{1 \times n-m}. \quad (3.10)$$

Aus (3.9) folgt, dass die m Spalten von $\frac{\partial g}{\partial y}(x_0)$ linear unabhängig sind. Deshalb gibt es einen Vektor $\lambda = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$ mit

$$\frac{\partial f}{\partial y}(x_0) + \lambda^\top \frac{\partial g}{\partial y}(x_0) = 0 \in \mathbb{R}^{1 \times m}. \quad (3.11)$$

Es bleibt also nur noch nachzuweisen, dass auch

$$\frac{\partial f}{\partial z}(x_0) + \lambda^\top \frac{\partial g}{\partial z}(x_0) = 0 \in \mathbb{R}^{1 \times n-m}$$

ist. Beachte dazu, dass nach Kettenregel wegen $g(h(z), z) = 0$ gilt

$$\frac{\partial g}{\partial y}(x_0) \frac{\partial h}{\partial z}(z_0) + \frac{\partial g}{\partial z}(x_0) = 0. \quad (3.12)$$

Damit erhalten wir schließlich:

$$\begin{aligned} \frac{\partial f}{\partial z}(x_0) &\stackrel{(3.10)}{=} -\frac{\partial f}{\partial y}(x_0) \frac{\partial h}{\partial z}(z_0) \\ &\stackrel{(3.11)}{=} \lambda^\top \frac{\partial g}{\partial y}(x_0) \frac{\partial h}{\partial z}(z_0) \\ &\stackrel{(3.12)}{=} -\lambda^\top \frac{\partial g}{\partial z}(x_0) \end{aligned}$$

bzw. äquivalenterweise

$$\frac{\partial f}{\partial z}(x_0) + \lambda^\top \frac{\partial g}{\partial z}(x_0) = 0,$$

was zu zeigen war. ■

Korollar 3.33

Notwendige Bedingungen dafür, dass eine reellwertige Funktion f an der Stelle $x_0 \in \mathbb{R}^n$ eine lokale Extremstelle unter den Nebenbedingungen $g(\cdot) = c = (c_1, \dots, c_m)^\top \in \mathbb{R}^m$ mit $m < n$ besitzt, sind gegeben durch

$$\begin{aligned} \frac{\partial L}{\partial x}(x_0, \lambda_0) &= 0 \in \mathbb{R}^{1 \times n} \\ \frac{\partial L}{\partial \lambda}(x_0, \lambda_0) &= 0 \in \mathbb{R}^{1 \times m} \end{aligned}$$

für die Lagrange-Funktion

$$L : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$$

$$(x, \lambda) \mapsto L(x, \lambda) := f(x) + \sum_{k=1}^m \lambda_k (g_k(x) - c_k),$$

wobei $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ und $\lambda = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$.

Beweis: Für alle $1 \leq i \leq n$ ist

$$\frac{\partial L}{\partial x_i}(x, \lambda) = \frac{\partial f}{\partial x_i}(x) + \sum_{k=1}^m \lambda_k \frac{\partial g_k}{\partial x_i}(x),$$

und für alle $1 \leq k \leq m$ ist

$$\frac{\partial L}{\partial \lambda_k}(x, \lambda) = g_k(x) - c_k.$$

Damit folgt die Aussage aus Satz 3.32. ■

Beispiel 3.34

Betrachte die Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x_1, x_2) \mapsto f(x_1, x_2) = 2x_1x_2$$

und die Nebenbedingung $g(x_1, x_2) = 1$ mit

$$g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x_1, x_2) \mapsto g(x_1, x_2) := x_1^2 + x_2^2.$$

Wir benötigen genau einen Lagrange-Multiplikator λ und erhalten die Lagrange-Funktion L , gegeben durch

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda(g(x_1, x_2) - 1)$$

$$= 2x_1x_2 + \lambda(x_1^2 + x_2^2) - \lambda.$$

Damit ergeben sich die notwendigen Bedingungen

- (I) $\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) = 2x_2 + 2x_1\lambda \stackrel{!}{=} 0,$
- (II) $\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = 2x_1 + 2x_2\lambda \stackrel{!}{=} 0,$
- (III) $\frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = x_1^2 + x_2^2 - 1 \stackrel{!}{=} 0.$

Aus (I) und (II) erhalten wir

$$\begin{aligned}2x_2 + 2x_1\lambda &= 2x_1 + 2x_2\lambda \\ \iff (2x_1 - 2x_2)\lambda &= 2x_1 - 2x_2 \\ \iff x_1 = x_2 \text{ oder } \lambda &= 1.\end{aligned}$$

Für $x_1 = x_2$ ergibt (III), dass $x_1 = x_2 = \pm\sqrt{2}/2$. Einsetzen von $x_1 = x_2 = \pm\sqrt{2}/2$ in (I) ergibt $\sqrt{2} + \sqrt{2}\lambda = 0 \iff \lambda = -1$. Einsetzen von $\lambda = 1$ in (I) ergibt andererseits $x_1 = -x_2$.

Es gibt also genau vier Lösungen des Gleichungssystems, nämlich

$$\begin{aligned}P_1 &= (\sqrt{2}/2, \sqrt{2}/2, -1), \\ P_2 &= (-\sqrt{2}/2, -\sqrt{2}/2, -1), \\ P_3 &= (\sqrt{2}/2, -\sqrt{2}/2, 1), \\ P_4 &= (-\sqrt{2}/2, \sqrt{2}/2, 1).\end{aligned}$$

Da die Menge $\mathcal{M} = \{(x_1, x_2)^\top \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ abgeschlossen und beschränkt ist, nimmt die stetige Funktion f nach dem Extremalsatz von Weierstraß (Satz 1.10) auf \mathcal{M} ihr Maximum und ihr Minimum an. Also sind P_1 und P_2 globale Maximalstellen von f auf \mathcal{M} und P_3 und P_4 globale Minimalstellen von f auf \mathcal{M} .

Bemerkung 3.35

- (a) Der Beweis von Satz 3.32 (Lagrange-Multiplikatoren) zeigt, dass die Einführung der Lagrange-Multiplikatoren das restringierte Optimierungsproblem ($\mathcal{M} = \{x \in \mathbb{R}^n : g(x) = c\}$) bezüglich f auf ein Optimierungsproblem ohne Nebenbedingungen bezüglich $z \mapsto f(h(z), z)$ zurückführt. Damit können mit den Techniken aus Abschnitten 3.1 und 3.2 auch hinreichende Bedingungen für lokale Extrema unter Nebenbedingungen hergeleitet werden (siehe Satz 3.38 unten).
- (b) Globalität der ermittelten lokalen Extrema von f auf \mathcal{M} kann analysiert werden, wenn \mathcal{M} eine konvexe Menge ist (siehe Abschnitt 3.2 und Beispiel 3.34).

Zum Abschluss dieses Kapitels greifen wir die "kanonische Form" des restringierten Optimierungsproblem aus Kapitel 1 noch einmal auf. Das Einführen von Schlupfvariablen führt hier auf die sogenannten "Karush-Kuhn-Tucker" (KKT) - Bedingungen.

Satz 3.36 (Karush-Kuhn-Tucker-Bedingungen)

Seien f und g so wie in Satz 3.32. Gegeben sei das Optimierungsproblem: Maximiere f unter den Nebenbedingungen

$$g(x) \leq c \text{ und } x \geq 0 \in \mathbb{R}^n. \quad (3.13)$$

Sei wieder

$$L : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$$

$$(x, \lambda) \mapsto L(x, \lambda) := f(x) + \sum_{k=1}^m \lambda_k (c_k - g_k(x)).$$

Falls dann x_0 eine lokale Maximalstelle des Optimierungsproblems (3.13) ist, dann existiert ein $\lambda_0 \in \mathbb{R}^m$ so, dass

- (I) $\frac{\partial L}{\partial x}(x_0, \lambda_0) \leq 0 \in \mathbb{R}^{1 \times n}$,
- (II) $\left[\frac{\partial L}{\partial x}(x_0, \lambda_0) \right]^\top x_0 = 0$,
- (III) $\frac{\partial L}{\partial \lambda}(x_0, \lambda_0) = c - g(x_0) \geq 0$,
- (IV) $\left[\frac{\partial L}{\partial \lambda}(x_0, \lambda_0) \right]^\top \lambda_0 = (c - g(x_0))^\top \lambda_0 = 0$,
- (V) $x_0, \lambda_0 \geq 0$.

Die Bedingungen (I) - (V) heißen (notwendige) Karush-Kuhn-Tucker (KKT)-Bedingungen.

Beispiel 3.37

Gesucht werden die Maximalstellen von

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$x = (x_1, x_2) \mapsto f(x) := -x_1^2 - 3x_2^2 + 4x_1 + 6x_2$$

auf $\mathcal{M} = \{x \geq 0 : g(x) \leq 4\}$ mit

$$g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$x \mapsto g(x) := x_1 + 2x_2.$$

Die Lagrange-Funktion ergibt sich zu

$$L(x_1, x_2, \lambda) = -x_1^2 - 3x_2^2 + 4x_1 + 6x_2 + \lambda(4 - x_1 - 2x_2).$$

Daraus ergeben sich die folgenden KKT-Bedingungen.

- (I) $\frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) = -2x_1 + 4 - \lambda \leq 0$,
- (II) $\frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = -6x_2 + 6 - 2\lambda \leq 0$,

$$\text{(III)} \quad x_1 \frac{\partial L}{\partial x_1}(x_1, x_2, \lambda) + x_2 \frac{\partial L}{\partial x_2}(x_1, x_2, \lambda) = x_1(-2x_1 + 4 - \lambda) + x_2(-6x_2 + 6 - 2\lambda) = 0,$$

$$\text{(IV)} \quad \frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = 4 - x_1 - 2x_2 \geq 0,$$

$$\text{(V)} \quad \lambda \frac{\partial L}{\partial \lambda}(x_1, x_2, \lambda) = \lambda(4 - x_1 - 2x_2) = 0,$$

$$\text{(VI)} \quad x_1, x_2, \lambda \geq 0.$$

Aus (I), (II), (VI) und (III) zusammengenommen folgt, dass

$$\text{(VII)} \quad x_1(-2x_1 + 4 - \lambda) = 0 \text{ und}$$

$$\text{(VIII)} \quad x_2(-6x_2 + 6 - 2\lambda) = 0$$

gelten müssen. Eine Lösung wäre $(x_1, x_2, \lambda) = (0, 0, \lambda)$. Aus (V) würde dann aber folgen, dass $\lambda = 0$ ist; der Punkt $(0, 0, 0)$ verletzt aber (I). Also ist $x_1 = 0, x_2 = 0$ ausgeschlossen. Aus (V) erhalten wir $\lambda = 0$ als eine mögliche Lösung. Dann ergeben (VII) und (VIII), dass $x_1 = 2$ und $x_2 = 1$ sein müssen. Es stellt sich heraus, dass $P = (2, 1, 0)$ die einzige Lösung ist. Da f konkav ist, ist P sogar eine globale Maximalstelle von f auf \mathcal{M} .

Satz 3.38

Unter den Voraussetzungen von Satz 3.36 seien f und g zusätzlich zweimal stetig differenzierbar sowie f konkav und g konvex. Falls dann $\mathcal{M} = \{x \geq 0 : g(x) \leq c\}$ nicht leer ist, so sind die KKT-Bedingungen aus Satz 3.36 notwendig und hinreichend dafür, dass x_0 eine globale Maximalstelle von f auf \mathcal{M} ist.

Kapitel 4

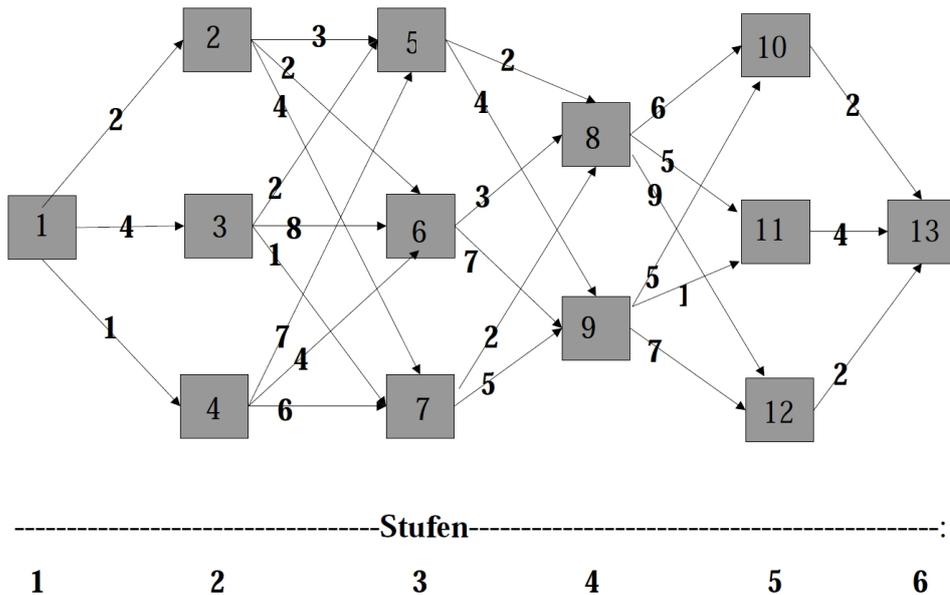
Dynamische und stochastische Optimierung

4.1 Dynamische Programmierung

Die dynamische Programmierung (DP) gehört zu der Klasse der Entscheidungsbaumverfahren. Dabei wird das zu lösende Optimierungsproblem in einzelne Teilprobleme zerlegt, die dann sequentiell gelöst werden. Mathematisch gesehen ist dieses Vorgehen nicht sonderlich aufwändig. Das eigentliche Problem besteht eher darin, zu erkennen, dass das zu lösende Optimierungsproblem zur Kategorie DP gehört.

Beispiel 4.1 (Fahrtstreckenproblem)

Angenommen, eine Reisende möchte von Kalifornien aus in den Bundesstaat New York mit dem Auto fahren. Dabei sind mehrere Reiserouten möglich. Das Ziel der Reisenden sei, die (Gesamt-) Strecke möglichst schnell zurückzulegen. Die relevanten Daten für die Routenplanung sind in der folgenden Grafik zusammengefasst.



Die einzelnen Bundesstaaten, durch die die Reisende fahren kann, sind als Zustände in den quadratischen Kästchen dargestellt und von 1 bis 13 durchnummeriert (Zustand 1 $\hat{=}$ Kalifornien (Start), Zustand 13 $\hat{=}$ New York (Ziel)). In Kalifornien startend muss sich die Reisende überlegen, ob sie zunächst zum Staat 2,3 oder 4 fährt. Die dazu benötigten Zeiteinheiten (ZE) sind an den Pfeilen notiert, die auf die jeweiligen Zustände zeigen. Im zweiten Streckenabschnitt (vgl. die "Stufen" am unteren Rand der Grafik) muss sie sich sodann überlegen, ob sie als nächstes zum Staat 5, 6 oder 7 fährt (Streckenabschnitt (Stufe) 3). Auf diese Weise geht es weiter bis zur letzten Stufe 6 (Ankunft in New York).

Das zu minimierende Zielkriterium ist die Gesamtfahrtzeit, also die Summe der ZE, die in den fünf Streckenabschnitten anfallen. Wählt die Reisende beispielsweise die nördlichste Route (1 \rightarrow 2 \rightarrow 5 \rightarrow 8 \rightarrow 10 \rightarrow 13), so benötigt sie insgesamt $2 + 3 + 2 + 6 + 2 = 15$ ZE.

Frage: Was ist die optimale Route (Minimierung der Gesamtfahrtzeit) ?

Definition 4.2 (Rückwärtsrekursion)

Sei

$f_n(s) :=$ Anzahl der Zeiteinheiten der optimalen Route vom Zustand s in Stufe n bis zum Ziel, $1 \leq s \leq 12, 1 \leq n \leq 5$.

Offenbar ist $f_5(10) = 2, f_5(11) = 4, f_5(12) = 2$.

Ferner sei

$f_n(s, t) :=$ Anzahl der Zeiteinheiten der optimalen Route vom Zustand s in Stufe n bis zum Ziel, wenn man im nächsten Schritt ($n + 1$) den Zustand t wählt.

Bezeichnen wir mit S_n die Menge der möglichen Zustände in der n -ten Stufe, so gilt

$$f_n(s) = \min_{t \in S_{n+1}} f_n(s, t).$$

Schließlich definieren wir noch

$c_n(s, t) :=$ Anzahl ZE vom Zustand s in Stufe n zum Zustand t in Stufe $(n + 1)$.

Wir erhalten den folgenden rekursiven Zusammenhang:

$$f_n(s, t) = c_n(s, t) + f_{n+1}(t),$$

mit dem wir das Problem vermittels einer Rückwärtsrekursion lösen können. Es bietet sich dabei an, das Verfahren in Tabellenform durchzuführen.

Beispiel 4.3 (Beispiel 4.1 fortgesetzt)

(Die jeweils optimale Entscheidung wird mit t^* gekennzeichnet.)

Stufe 5:

$s \in S_5$	$f_5(s)$	t^*
10	2	13
11	4	13
12	2	13

Stufe 4:

Wir beachten, dass $f_4(s, t) = c_4(s, t) + f_5(t)$ ist.

$s \in S_4$	$t = 10$	$t = 11$	$t = 12$	$f_4(s)$	t^*
8	$6 + f_5(10) = 8$	$5 + f_5(11) = 9$	$9 + f_5(12) = 11$	8	10
9	$5 + f_5(10) = 7$	$1 + f_5(11) = 5$	$7 + f_5(12) = 9$	5	11

Stufe 3:

Wir beachten, dass $f_3(s, t) = c_3(s, t) + f_4(t)$ ist.

$s \in S_3$	$t = 8$	$t = 9$	$f_3(s)$	t^*
5	$2 + f_4(8) = 10$	$4 + f_4(9) = 9$	9	9
6	$3 + f_4(8) = 11$	$7 + f_4(9) = 12$	11	8
7	$2 + f_4(8) = 10$	$5 + f_4(9) = 10$	10	8 oder 9

Stufe 2:

Wir beachten, dass $f_2(s, t) = c_2(s, t) + f_3(t)$ ist.

$s \in S_2$	$t = 5$	$t = 6$	$t = 7$	$f_2(s)$	t^*
2	$3 + f_3(5) = 12$	$2 + f_3(6) = 13$	$4 + f_3(7) = 14$	12	5
3	$2 + f_3(5) = 11$	$8 + f_3(6) = 19$	$1 + f_3(7) = 11$	11	5 oder 7
4	$7 + f_3(5) = 16$	$4 + f_3(6) = 15$	$6 + f_3(7) = 16$	15	6

Stufe 1:

Wir beachten, dass $f_1(s, t) = c_1(s, t) + f_2(t)$ ist.

$s \in \mathcal{S}_1$	$t = 2$	$t = 3$	$t = 4$	$f_1(s)$	t^*
1	$2 + f_2(2) = 14$	$4 + f_2(3) = 15$	$1 + f_2(4) = 16$	14	2

Liest man die Tabellen nun von unten nach oben, so erhält man die Zeit-optimale Route

$$1 \rightarrow 2 \rightarrow 5 \rightarrow 9 \rightarrow 11 \rightarrow 13$$

mit einer Gesamtfahrtzeit von 14 ZE.

Definition 4.4 (Vorwärtsrekursion)

Obschon die Rückwärtsrekursion das gängige Verfahren zur Lösung von DP-Problemen ist, existiert auch ein weiteres rekursives Verfahren, nämlich die Vorwärtsrekursion. Betrachte dazu $g_n(s) :=$ Anzahl ZE der optimalen Route vom Start bis in den Zustand s in Stufe $(n + 1)$.

In Beispiel 4.1 ist zum Beispiel $g_1(2) = 2$, $g_1(3) = 4$ und $g_1(4) = 1$. Wir suchen nun also $g_5(13)$ mit der zugehörigen Route. Bezeichne dazu ferner

$g_n(t, s) :=$ Anzahl ZE der optimalen Route vom Start bis in den Zustand s in Stufe $(n + 1)$, wenn man in der n -ten Stufe im Zustand t war. Nun erhalten wir

$$g_n(s) = \min_{t \in \mathcal{S}_n} g_n(t, s).$$

Sei schließlich wiederum

$c_n(t, s) :=$ Anzahl ZE vom Zustand t in Stufe n zum Zustand s in Stufe $(n + 1)$.

Wir erhalten hier den rekursiven Zusammenhang

$$g_n(t, s) = c_n(t, s) + g_{n-1}(t).$$

Damit kann z. B. das Fahrstreckenproblem aus Beispiel 4.1 ebenfalls gelöst werden (siehe Übungsaufgabe).

Definition 4.5 (Charakteristische Merkmale der dynamischen Programmierung)

- (1) Das Problem kann in Stufen zerlegt werden.
- (2) Pro Stufe gibt es eine Anzahl von Zuständen, die die jeweiligen Situationen darstellen, in denen sich das betrachtete System befinden kann, wenn es diese Stufe erreicht.
- (3) Pro Stufe muss eine Entscheidung getroffen werden, die vom gerade aktuellen Zustand des Systems in einen Zustand der nächsten Stufe führt.
- (4) Jede Entscheidung ist mit einem Preis (Kostenwert) verbunden.

- (5) Geht man vom momentanen Zustand aus, so sind die Entscheidungen in den verbleibenden Stufen unabhängig von den getroffenen Entscheidungen, die zum momentanen Zustand geführt haben (Markov-Eigenschaft).

Beispiel 4.6 (Beladungsproblem)

Ein Schiff soll mit Waren beladen werden. Es stehen dabei N unterschiedliche Warentypen mit jeweiligem Einheitsgewicht w_i (in Tonnen (t)) und Einheitswert v_i (in GE), $1 \leq i \leq N$, zur Verfügung. Die Gesamtkapazität des Schiffs betrage W Tonnen.

Ziel: Belade das Schiff so, dass der Gesamtwert der Ladung maximal wird.

Sei dazu x_i die Anzahl der auf das Schiff zu ladenden Einheiten vom Typ $1 \leq i \leq N$. Dann gilt es, die Funktion

$$f(x_1, \dots, x_N) = \sum_{i=1}^N x_i v_i \quad (4.1)$$

unter den Nebenbedingungen

$$\sum_{i=1}^N x_i w_i \leq W, \quad x_i \in \mathbb{N}_0 \forall 1 \leq i \leq N \quad (4.2)$$

zu maximieren.

Erste Lösung: Fasse das durch (4.1) und (4.2) definierte Maximierungsproblem als ein ILP auf und verwende die Techniken aus Kapitel 2 (Branch & Bound Algorithmus).

Hier: Lösung durch DP.

Dazu ist es hilfreich, sich vorzustellen, wie der eigentliche Beladungsprozess abläuft.

Schritt 1: Wähle x_1 und lade x_1 Einheiten des ersten Warentyps auf das Schiff. Die verbleibende Ladekapazität beträgt dann $W - x_1 w_1$ Tonnen.

Schritt 2: Falls $W - x_1 w_1 \geq w_2$, wähle x_2 und lade x_2 Einheiten des zweiten Warentyps auf das Schiff. Die verbleibende Ladekapazität beträgt dann $W - x_1 w_1 - x_2 w_2$ Tonnen, usw.

Also wird auf Stufe n entschieden, wie viele Einheiten x_n vom Warentyp n verladen werden sollen ($1 \leq n \leq N$). Die Zustände in der n -ten Stufe sind die jeweils sich ergebenden Restladekapazitäten.

Konkretes Beispiel mit $N = 4$ und $W = 10$ Tonnen:

n	$w_n[t]$	$v_n[GE]$
1	2	4
2	8	10
3	3	6
4	4	8

Relevante Größen für die Rückwärtsrekursion:

$f_n(s) \hat{=}$ Maximaler Wert der Restladung bestehend aus den Warentypen n bis N , wenn in der Stufe n noch s Tonnen als Restladekapazität vorhanden sind.

$c_n(s, t) \hat{=}$ Wert der Zuladung beim Übergang von Zustand s in Stufe n zu Zustand t in Stufe $(n+1)$, d. h., $c_n(s, t) = (s - t)v_n/w_n$.

$f_n(s, t) \hat{=}$ Maximaler Wert der Restladung bestehend aus den Warentypen n bis N , wenn in der Stufe n noch s Tonnen als Restladekapazität vorhanden sind und man sich entscheidet, in Stufe $(n+1)$ in den Zustand t zu wechseln.

Gesucht: $f_1(10)$ mit den zugehörigen Entscheidungen.

Es gilt: $f_n(s) = \max_{t \in \mathcal{S}_{n+1}} f_n(s, t)$, wobei $f_n(s, t) = c_n(s, t) + f_{n+1}(t)$ ist.

Für die Praxis sind die folgenden Umbezeichnungen hilfreich.

$c_n(s, x_n) \hat{=}$ Wert der Zuladung von x_n Einheiten des Warentyps n , also $c_n(s, x_n) = v_n x_n$.

$f_n(s, x_n) \hat{=}$ $f_n(s, t)$, aber es wird statt des zu erreichenden Zustands t die Anzahl x_n von Einheiten des Warentyps n angegeben, die verladen werden sollen.

Dann lautet die Rückwärtsrekursionsformel:

$$f_n(s, x_n) = c_n(s, x_n) + f_{n+1}(s - x_n w_n)$$

Starttableau der Rückwärtsrekursion (Stufe $N = 4$):

$s \in \mathcal{S}_4$	$f_4(s)$	x_4^*
0	0	0
1	0	0
2	0	0
3	0	0
4	8	1
5	8	1
6	8	1
7	8	1
8	16	2
9	16	2
10	16	2

Stufe 3: $w_3 = 3, v_3 = 6, f_3(s, x_3) = c_3(s, x_3) + f_4(s - x_3 w_3)$

$s \in \mathcal{S}_3$	$x_3 = 0$	$x_3 = 1$	$x_3 = 2$	$x_3 = 3$	$f_3(s)$	x_3^*
10	$0 + f_4(10) = 16$	$6 + f_4(7) = 14$	$12 + f_4(4) = 20$	$18 + f_4(1) = 18$	20	2
8	$0 + f_4(8) = 16$	$6 + f_4(5) = 14$	$12 + f_4(2) = 12$	%	16	0
6	$0 + f_4(6) = 8$	$6 + f_4(3) = 6$	$12 + f_4(0) = 12$	%	12	2
4	$0 + f_4(4) = 8$	$6 + f_4(1) = 6$	%	%	8	0
2	$0 + f_4(2) = 0$	%	%	%	0	0
0	$0 + f_4(0) = 0$	%	%	%	0	0

Stufe 2: $w_2 = 8, v_2 = 10, f_2(s, x_2) = c_2(s, x_2) + f_3(s - x_2 w_2)$

$s \in \mathcal{S}_2$	$x_2 = 0$	$x_2 = 1$	$f_2(s)$	x_2^*
10	$0 + f_3(10) = 20$	$10 + f_3(2) = 10$	20	0
8	$0 + f_3(8) = 16$	$10 + f_3(0) = 10$	16	0
6	$0 + f_3(6) = 12$	%	12	0
4	$0 + f_3(4) = 8$	%	8	0
2	$0 + f_3(2) = 0$	%	0	0
0	$0 + f_3(0) = 0$	%	0	0

Stufe 1: $w_1 = 2, v_1 = 4, f_1(s, x_1) = c_1(s, x_1) + f_2(s - x_1 w_1)$

$s \in \mathcal{S}_1$	$x_1 = 0$	$x_1 = 1$	$x_1 = 2$	$x_1 = 3$	$x_1 = 4$	$x_1 = 5$
10	$0 + f_2(10) = 20$	$4 + f_2(8) = 20$	$8 + f_2(6) = 20$	$12 + f_2(4) = 20$	$16 + f_2(2) = 16$	$20 + f_2(0) = 20$

Also ist $f_1(10) = 20$.

Mögliche optimale Beladungen $(x_1^*, x_2^*, x_3^*, x_4^*)$:

$$\begin{aligned}(0, 0, 2, 1) & [2 \times 6 + 1 \times 8 = 12 + 8 = 20] \\(1, 0, 0, 2) & [1 \times 4 + 2 \times 8 = 4 + 16 = 20] \\(2, 0, 2, 0) & [2 \times 4 + 2 \times 6 = 8 + 12 = 20] \\(3, 0, 0, 1) & [3 \times 4 + 1 \times 8 = 12 + 8 = 20] \\(5, 0, 0, 0) & [5 \times 4 = 20]\end{aligned}$$

MATLAB liefert für das ILP mit Zielfunktion $f(x_1, x_2, x_3, x_4) = -4x_1 - 10x_2 - 6x_3 - 8x_4$ unter den Nebenbedingungen

$$x_i \geq 0 \quad \forall 1 \leq i \leq 4,$$

$$x_i \in \mathbb{Z} \quad \forall 1 \leq i \leq 4, \quad \text{sowie}$$

$$2x_1 + 8x_2 + 3x_3 + 4x_4 \leq 10$$

die Minimalstelle $x_1^* = 1, x_2^* = x_3^* = 0, x_4^* = 2$ mit $f(x_1^*, x_2^*, x_3^*, x_4^*) = -20$ (GE).

Dies ist (nur) eine der durch DP ermittelten optimalen Beladungen.

Beispiel 4.7 (Verpackungsproblem, Beispiel 0.3 fortgesetzt)

Wir greifen das Beispiel 0.3 (Verpackungsproblem, $N = 10$ Artikel unterschiedlicher Größe, fünf zu verwendende Verpackungsgrößen, alle N Artikel sollen verpackt werden können) noch einmal auf. Das Zielkriterium ist hier die Minimierung der Gesamt-Verpackungskosten (unter diesen Nebenbedingungen). Die Artikelgrößen (und damit die Schachtelpreise $c_1 > c_2 > \dots > c_N = c_{10}$) sind absteigend geordnet.

Ziel: Lösung dieses restringierten Optimierungsproblems mit Hilfe von DP.

Zur Bestimmung der Stufen und der Zustände betrachten wir wie in Beispiel 4.6 wieder den eigentlichen Ablauf des Verpackungsprozesses. Auf jeden Fall wird mindestens eine Schachtel der Größe 1 benötigt, denn ansonsten kann der erste Artikel nicht verpackt werden.

Frage: Soll man mehrere Schachteln dieser Größe verwenden?

Die gleiche Frage stellt sich in analoger Weise, wenn die ersten $(n - 1)$ Artikel verpackt sind, aber noch nicht alle fünf Verpackungsgrößen vergeben worden sind. Die Stufen $1 \leq n \leq N = 10$

entsprechen hier also den einzelnen Artikeln und die Zustände den noch verbleibenden Schachtelgrößen. Wir definieren daher

$f_n(s) \hat{=}$ Gesamtkosten für die beste Packstrategie für die Artikel n bis 10, wenn noch s Schachtelgrößen zur Verfügung stehen.

$c_n(s, x_n) \hat{=}$ Kosten für x_n Schachteln der Größe n , d.h., $c_n(s, x_n) = c_n x_n$.

$f_n(s, x_n) \hat{=}$ Gesamtkosten für die beste Packstrategie für die Artikel n bis 10, wenn in Stufe n noch s Schachtelgrößen zur Verfügung stehen und man sich auf dieser Stufe entscheidet, x_n Schachteln der Größe n zu verwenden.

Gesucht: $f_1(5)$

Rekursionsformel:

$$f_n(s, x_n) = c_n x_n + f_{n+x_n}(s - 1)$$

Hier werden also alle Werte $f_k(s - 1)$ für $k \geq n$ in der Rekursion benötigt.

Wählen wir als Zahlenbeispiel

$$(c_1, c_2, \dots, c_{10}) = (32, 30, 20, 17, 15, 10, 9, 5, 3, 2).$$

Stufe 10:

s	$f_{10}(s)$	x_{10}^*
1	2	1

In Stufe 10 gibt es nur einen Zustand, der betrachtet werden muss, denn für $s \geq 2$ können die zuvor getroffenen Entscheidungen nicht optimal gewesen sein.

Stufe 9:

s	$f_9(s)$	x_9^*
1	6	2
2	5	1

Stufe 8:

Wir beachten, dass $f_8(s) = 5x_8 + f_{8+x_8}(s - 1)$ ist.

s	$x_8 = 1$	$x_8 = 2$	$x_8 = 3$	$f_8(s)$	x_8^*
1	%	%	15	15	3
2	$5 + f_9(1) = 11$	$10 + f_{10}(1) = 12$	%	11	1
3	$5 + f_9(2) = 10$	%	%	10	1

Übungsaufgabe:

Vervollständigung der Rückwärtsrekursion. Als Ergebnis erhalten wir $f_1(5) = 153$ mit optimaler Packstrategie $x_1^* = 2, x_3^* = 1, x_4^* = 2, x_6^* = 2, x_8^* = 3$.

4.2 Stochastische dynamische Optimierung

In Abschnitt 4.1 wurde angenommen, dass auf jeder Stufe des dynamisch zu lösenden Optimierungsproblems alle Rahmenbedingungen (d. h., alle Parameter des jeweiligen Zustands des Systems und seiner Übergänge) a priori festgelegt und genau bekannt sind. Dies muss indes in der Praxis nicht stets so sein. Zum Beispiel könnte in dem Fahrstreckenproblem aus Beispiel 4.1 die Reisezeit von unvorhersehbaren bzw. zufälligen Faktoren wie Wetter, Bildung von Staus etc. abhängen. Um solche Unsicherheiten berücksichtigen zu können, muss das Modell aus Definition 4.5 um zufällige Komponenten erweitert werden.

Definition 4.8 (Stochastisches dynamisches Optimierungsproblem)

Im Gegensatz zu der Situation in Definition 4.5 modellieren wir den Zustand s_{n+1} in Stufe $n+1$ bei gegebenem Zustand s_n in Stufe n und getroffener Entscheidung (Aktion) a_n in Stufe n nicht mehr als deterministische Funktion $s_{n+1} = z_n(s_n, a_n)$, sondern als Realisierung $s_{n+1} = I_{n+1}(\omega)$ einer Zufallsvariable I_{n+1} , so dass (I_1, \dots, I_N) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert sind und die Markoff-Eigenschaft besitzen. D. h., die Verteilung von I_{n+1} hängt von $s_n = I_n(\omega)$ und von der Aktion a_n ab.

Falls die Zustandsmenge \mathcal{S}_{n+1} in Stufe $n+1$ höchstens abzählbar ist, so ist das Wahrscheinlichkeitsgesetz, dem das betrachtete System unterliegt, durch Übergangswahrscheinlichkeiten

$$\{p_n(s_n, a_n, s') : s' \in \mathcal{S}_{n+1}\}$$

gegeben, $0 \leq n \leq N$. Die Funktion

$$\begin{aligned} p_n(s_n, a_n, \cdot) : \mathcal{S}_{n+1} &\rightarrow [0, 1] \\ s' &\mapsto p_n(s_n, a_n, s') \end{aligned}$$

ist eine Zähldichte auf $(\mathcal{S}_{n+1}, 2^{\mathcal{S}_{n+1}})$.

Die Komponenten eines stochastischen dynamischen Programmierungsproblems sind:

- (i) der Planungshorizont $N \in \mathbb{N}$, der die Anzahl der Stufen $0 \leq n \leq N$ festlegt.
- (ii) der Zustandsraum \mathcal{S} (eine höchstens abzählbare, nicht-leere Menge). Die nicht-leeren Teilmengen $\mathcal{S}_0 = \{s_0\}, \mathcal{S}_1, \dots, \mathcal{S}_N$ sind die Zustandsmengen der Stufen $0, 1, \dots, N$.
- (iii) der Aktionenraum \mathcal{A} (ebenfalls eine höchstens abzählbare, nicht-leere Menge). Die nicht-leeren endlichen Teilmengen $\mathcal{A}_n(s) : 0 \leq n < N, s \in \mathcal{S}_n$, sind die vom Zustand und der

Stufe abhängigen zulässigen Aktionen. Insbesondere ist $a_n \in \mathcal{A}_n(s), 0 \leq n < N$. Ferner definieren wir $\mathcal{D}_n := \{(s, a) \in \mathcal{S}_n \times \mathcal{A} : a \in \mathcal{A}_n(s)\}, 0 \leq n < N$.

- (iv) das Übergangswahrscheinlichkeitsgesetz p . Für alle $0 \leq n < N$ und alle $(s, a) \in \mathcal{D}_n$ ist $p_n(s, a, \cdot)$ eine Zähldichte (Wahrscheinlichkeitsfunktion) auf \mathcal{S}_{n+1} . Der Wert $p_n(s, a, s')$ gibt die (bedingte) Wahrscheinlichkeit an, mit der $s' \in \mathcal{S}_{n+1}$ angenommen wird, gegeben $(s, a) \in \mathcal{D}_n$.
- (v) die (einstufige) Gewinnfunktion $r_n : \mathcal{D}_n \rightarrow \mathbb{R}$. Der Wert $r_n(s, a)$ gibt den einstufigen Gewinn an, der bei Wahl der Aktion $a \in \mathcal{A}_n(s)$ entsteht, $0 \leq n < N$.
- (vi) die terminale Gewinnfunktion $V_N : \mathcal{S}_N \rightarrow \mathbb{R}$. Der Wert $V_N(s), s \in \mathcal{S}_N$, gibt den terminalen Gewinn (Gewinn im Endzustand) an.

Das Tupel $(N, \mathcal{S}, \mathcal{A}, p, r, V_N)$ heißt (endlich-stufiges) stochastisches dynamisches Optimierungsproblem.

Definition 4.9 (Entscheidungsfunktion, -strategie)

Gegeben sei ein stochastisches dynamisches Optimierungsproblem im Sinne von Definition 4.8. Eine Funktion $f_n : \mathcal{S}_n \rightarrow \mathcal{A}$ mit $f_n(s) \in \mathcal{A}_n(s), 0 \leq n < N, s \in \mathcal{S}_n$, die auf der Stufe n jedem Zustand $s \in \mathcal{S}_n$ eine zulässige Aktion $a = f_n(s)$ zuordnet, bezeichnet man als Entscheidungsfunktion. Ferner heißt ein Tupel $\delta = (f_0, \dots, f_{N-1})$ von Entscheidungsfunktionen eine (Entscheidungs-) Strategie. Die Menge aller (zur Konkurrenz zugelassenen) Strategien werden in der Folge mit Δ bezeichnet.

Definition 4.10 ((erwarteter) Gesamtgewinn)

Unter den Maßgaben von Definition 4.8 seien eine Strategie $\delta = (f_0, \dots, f_{N-1})$ sowie ein Anfangszustand s_0 vorgegeben.

- (a) Für gegebene Realisierungen $I_1(\omega) = s_1, \dots, I_N(\omega) = s_N$ ist der (realisierte) Gesamtgewinn gegeben als

$$R_{s_0, \delta}(s_1, \dots, s_N) := \sum_{n=0}^{N-1} r_n(s_n, f_n(s_n)) + V_N(s_N).$$

- (b) Der erwartete Gesamtgewinn $\bar{R}_\delta(s_0)$ ist gegeben durch

$$\begin{aligned} \bar{R}_\delta(s_0) &= \mathbb{E}_{s_0, \delta}[R_{s_0, \delta}(I_1, \dots, I_N)] \\ &= \sum_{s_1 \in \mathcal{S}_1} \dots \sum_{s_N \in \mathcal{S}_N} R_{s_0, \delta}(s_1, \dots, s_N) \cdot \mathbb{P}_{s_0, \delta}(I_1 = s_1, \dots, I_N = s_N). \end{aligned}$$

Dabei gilt wegen der Markoff-Eigenschaft des Systems, dass

$$\mathbb{P}_{s_0, \delta}(I_1 = s_1, \dots, I_N = s_N) = \prod_{n=0}^{N-1} p_n(s_n, f_n(s_n), s_{n+1}).$$

(c) Zur Herleitung einer Rekursionsformel definieren wir zudem die folgenden Größen. Für $0 \leq n \leq N - 1$ sei

$$\bar{R}_{n, \delta}(s_n) := \sum_{s_{n+1} \in \mathcal{S}_{n+1}} \dots \sum_{s_N \in \mathcal{S}_N} R_{s_n, \delta}(s_{n+1}, \dots, s_N) \cdot \mathbb{P}_{s_n, \delta}(I_{n+1} = s_{n+1}, \dots, I_N = s_N),$$

wobei

$$R_{s_n, \delta}(s_{n+1}, \dots, s_N) = \sum_{t=n}^{N-1} r_t(s_t, f_t(s_t)) + V_N(s_N),$$

$$\mathbb{P}_{s_n, \delta}(I_{n+1} = s_{n+1}, \dots, I_N = s_N) = \prod_{t=n}^{N-1} p_t(s_t, f_t(s_t), s_{t+1}).$$

Offenbar bezeichnet damit $\bar{R}_{n, \delta}(s_n)$ den erwarteten Gesamtgewinn auf den Stufen $n, n + 1, \dots, N$ in Abhängigkeit vom Zustand s_n auf Stufe n und bei Anwendung der Strategie δ . Ferner gilt

$$\bar{R}_\delta(s_0) = \bar{R}_{0, \delta}(s_0).$$

In der Folge sei stets angenommen, dass

$$\forall 0 \leq n \leq N - 1 : \forall s \in \mathcal{S}_n : \forall \delta \in \Delta : \bar{R}_{n, \delta}(s) \in \mathbb{R}.$$

Lemma 4.11 (Rekursionsformel)

Unter den Voraussetzungen von Definition 4.10 gilt $\forall 0 \leq n \leq N - 1 : \forall s \in \mathcal{S}_n : \forall \delta = (f_0, \dots, f_{N-1}) \in \Delta$:

$$\bar{R}_{n, \delta}(s) = r_n(s, f_n(s)) + \sum_{s' \in \mathcal{S}_{n+1}} p_n(s, f_n(s), s') \cdot \bar{R}_{n+1, \delta}(s'), \quad (4.3)$$

wobei $\bar{R}_{N, \delta}(s) := V_N(s)$ für $s \in \mathcal{S}_N$ gesetzt wird.

Beweis: Die Aussage folgt aus den elementaren Rechenregeln für (bedingte) Erwartungswerte. ■

Definition 4.12 (optimale Strategien)

Unter den Voraussetzungen von Definition 4.10 heißt eine Strategie $\delta^* \in \Delta$ optimal (bzgl. des Anfangszustands s_0), falls

$$\forall \delta \in \Delta : \bar{R}_\delta(s_0) \leq \bar{R}_{\delta^*}(s_0).$$

Der maximale erwartete Gesamtgewinn ist dann gegeben durch

$$V_0(s_0) := \sup_{\delta \in \Delta} \bar{R}_\delta(s_0).$$

Wir definieren ferner

$$\forall 0 \leq n \leq N - 1 : V_n(s) := \sup_{\delta \in \Delta} \bar{R}_{n,\delta}(s), s \in \mathcal{S}_n. \quad (4.4)$$

Die Funktionen V_0, \dots, V_N heißen Wertfunktionen.

Der folgende Satz ist das Hauptergebnis dieses Abschnitts.

Satz 4.13 (Optimalitätsgleichung)

Unter den Voraussetzungen von Definition 4.10 gelten die folgenden Aussagen.

(i) Für alle $0 \leq n \leq N - 1$ und alle $s \in \mathcal{S}_n$ ist

$$V_n(s) = \max_{a \in \mathcal{A}_n(s)} \left\{ r_n(s, a) + \sum_{s' \in \mathcal{S}_{n+1}} p_n(s, a, s') \cdot V_{n+1}(s') \right\}. \quad (4.5)$$

(ii) Jede aus den die rechte Seite von (4.5) maximierenden Aktionen $f_n^*(s)$ gebildete Strategie

$\delta^* = (f_0^*, \dots, f_{N-1}^*)$ ist optimal.

Beweis: Wir führen eine Rückwärtsinduktion durch.

Induktionsanfang ($n = N - 1$):

Nach (4.3) und (4.4) gilt für alle $s \in \mathcal{S}_{N-1}$:

$$\begin{aligned} V_{N-1} &= \sup_{\delta = (f_0, \dots, f_{N-1}) \in \Delta} \bar{R}_{N-1,\delta}(s) \\ &= \max_{a \in \mathcal{A}_{N-1}(s)} \left\{ r_{N-1}(s, a) + \sum_{s' \in \mathcal{S}_N} p_{N-1}(s, a, s') V_N(s') \right\}. \end{aligned} \quad (4.6)$$

Offenbar ist (4.6) gerade (4.5) für $n = N - 1$. Da die Mengen $\mathcal{A}_{N-1}(s)$, $s \in \mathcal{S}_{N-1}$, nach Voraussetzung (siehe Definition 4.8.(iii)) endlich sind, existiert eine Entscheidungsfunktion f_{N-1}^* , für die die rechten Seiten in (4.6) ihr Maximum annehmen. Somit gilt

$$\forall s \in \mathcal{S}_{N-1} : V_{N-1}(s) = \bar{R}_{N-1,\delta^*}(s) \quad (4.7)$$

für alle Strategien der Form $\delta^* = (f_0, \dots, f_{N-2}, f_{N-1}^*)$.

Induktionsschritt (hier o. B. d. A. $N - 1 \rightarrow N - 2$, ansonsten analog):

Sei nun $n = N - 2$. Mit (4.3) und (4.4) folgt zunächst für alle $s \in \mathcal{S}_{N-2}$:

$$\begin{aligned}
V_{N-2}(s) &= \sup_{\delta=(f_0, \dots, f_{N-1}) \in \Delta} \bar{R}_{N-2, \delta}(s) \\
&= \sup_{\delta \in \Delta} \left\{ r_{N-2}(s, f_{N-2}(s)) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, f_{N-2}(s), s') \cdot \bar{R}_{N-1, \delta}(s') \right\}.
\end{aligned}$$

Wir nutzen nun die Induktionsvoraussetzung aus und erhalten

$$\begin{aligned}
V_{N-2}(s) &\leq \sup_{\delta \in \Delta} \left\{ r_{N-2}(s, f_{N-2}(s)) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, f_{N-2}(s), s') V_{N-1}(s') \right\} \\
&= \max_{a \in \mathcal{A}_{N-2}(s)} \left\{ r_{N-2}(s, a) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, a, s') V_{N-1}(s') \right\}. \quad (4.8)
\end{aligned}$$

Sei nun f_{N-2}^* eine aus den Maximalpunkten von (4.8) gebildete Entscheidungsfunktion. Dann gilt

$$V_{N-2}(s) \leq r_{N-2}(s, f_{N-2}^*(s)) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, f_{N-2}^*(s), s') \cdot V_{N-1}(s') = \bar{R}_{N-2, \delta^*}(s)$$

für alle Strategien der Form $\delta^* = (f_0, \dots, f_{N-3}, f_{N-2}^*, f_{N-1}^*)$.

Andererseits gilt aber:

$$\begin{aligned}
V_{N-2}(s) &= \sup_{\delta \in \Delta} \bar{R}_{N-2, \delta}(s) \\
&\geq \bar{R}_{N-2, \delta^*}(s) \\
&= r_{N-2}(s, f_{N-2}^*(s)) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, f_{N-2}^*(s), s') \cdot \bar{R}_{N-1, \delta^*}(s') \\
&= r_{N-2}(s, f_{N-2}^*(s)) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, f_{N-2}^*(s), s') \cdot V_{N-1}(s') \\
&= \max_{a \in \mathcal{A}_{N-2}(s)} \left\{ r_{N-2}(s, a) + \sum_{s' \in \mathcal{S}_{N-1}} p_{N-2}(s, a, s') \cdot V_{N-1}(s') \right\}.
\end{aligned}$$

Also ist für alle $s \in \mathcal{S}_{N-2}$ sowohl $V_{N-2}(s) \leq \bar{R}_{N-2, \delta^*}(s)$ als auch $V_{N-2}(s) \geq \bar{R}_{N-2, \delta^*}(s)$. Demnach ist $V_{N-2}(s) = \bar{R}_{N-2, \delta^*}(s)$ für alle Strategien der Form $\delta^* = (f_0, \dots, f_{N-3}, f_{N-2}^*, f_{N-1}^*)$, die gemäß (4.5) gebildet werden. ■

Der folgende Algorithmus implementiert das Ergebnis von Staz 4.13 für die praktische Anwendung.

Algorithmus 4.14 (Wertiteration)

- (1) Setze $v'(s) := V_N(s)$ für alle $s \in \mathcal{S}_N$.
- (2) For $n = N - 1$ downto 0 do
- (a) Setze $v(s) = v'(s)$ für alle $s \in \mathcal{S}_{n+1}$.
- (b) Für alle $s \in \mathcal{S}_n$ berechne

$$v'(s) = \max_{a \in \mathcal{A}_n(s)} \left\{ r_n(s, a) + \sum_{s' \in \mathcal{S}_{n+1}} p_n(s, a, s') v(s') \right\}$$

und bestimme eine Entscheidungsfunktion f_n^* aus Maximalpunkten $f_n^*(s) \in \mathcal{A}_n(s)$.

end for

- (3) Output: $V_0(s_0) = v'(s_0)$ sowie $\delta^* = (f_0^*, \dots, f_{N-1}^*)$.

Definition 4.15 (Kontrollproblem)

Vielen dynamischen Optimierungsproblemen liegt eine exogene Einflussgröße zugrunde, die im Wesentlichen das Übergangswahrscheinlichkeitsgesetz bestimmt. Zum Beispiel spielt in den Wirtschaftswissenschaften die Nachfrage aus der Sicht eines Betriebes die Rolle einer solchen exogenen Einflussgröße, die von Lagerbestand und Bestellmenge unabhängig ist.

Modellierung der exogenen Einflussgröße:

Es seien Y_0, Y_1, \dots, Y_{N-1} stochastisch unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$, wobei Y_n Werte in \mathcal{Y}_n annimmt, und definiere

$$q_n(y) := \mathbb{P}(Y_n = y), \quad y \in \mathcal{Y}_n, \quad 0 \leq n \leq N - 1.$$

Annahme: Die Stochastik des Gesamtsystems wird einzig und allein durch Y_0, \dots, Y_{N-1} induziert (keine weiteren stochastischen Einflüsse). Damit hängt

$$s_{n+1} =: g_n(s_n, a_n, y_n) \in \mathcal{S}_{n+1}$$

von $s_n \in \mathcal{S}_n$, $a_n \in \mathcal{A}_n(s_n)$ und der Realisierung $y_n = Y_n(\omega) \in \mathcal{Y}_n$ ab. Also ist $g_n : \mathcal{D}_n \times \mathcal{Y}_n \rightarrow \mathcal{S}_{n+1}$.

Ein Kontrollmodell ist ein Tupel $(N, \mathcal{S}, \mathcal{A}, \mathcal{Y}, q, g, r, V_N)$, wobei

- (i) $N, \mathcal{S}, \mathcal{A}, r$ und V_N wie in Definition 4.8 sind.
- (ii) \mathcal{Y} (Wertebereich der exogenen Einflussgrößen) eine höchstens abzählbare Menge ist und $\mathcal{Y}_0, \mathcal{Y}_1, \dots, \mathcal{Y}_{N-1}$ (die Stufen-spezifischen Wertebereiche) nicht-leere Teilmengen von \mathcal{Y} sind.
- (iii) $q_n : \mathcal{Y}_n \rightarrow [0, 1]$ eine Zähl-dichte ist (Wahrscheinlichkeitsfunktion von Y_n), für alle $0 \leq n \leq N - 1$.

(iv) $g_n : \mathcal{D}_n \times \mathcal{Y}_n \rightarrow \mathcal{S}_{n+1}$, $0 \leq n \leq N - 1$ eine Übergangsfunktion ist, so dass $s_{n+1} = g_n(s_n, a_n, y_n)$ für $Y_n(\omega) = y_n$ (realisierter Wert von Y_n).

Das Kontrollmodell lässt sich als stochastisches dynamisches Optimierungsmodell schreiben. Definiere dazu für alle $0 \leq n \leq N - 1$:

$$p_n(s, a, s') = \sum_{y \in \mathcal{Y}_n : g_n(s, a, y) = s'} q_n(y),$$

für $(s, a) \in \mathcal{D}_n$ und $s' \in \mathcal{S}_{n+1}$. Die entsprechende Optimalitätsgleichung lautet damit:

$$V_n(s) = \max_{a \in \mathcal{A}_n(s)} \left\{ r_n(s, a) + \sum_{y \in \mathcal{Y}_n} q_n(y) V_{n+1}(g_n(s, a, y)) \right\},$$

und Satz 4.13.(ii) gilt analog.

Beispiel 4.16 (Stopp-Problem)

Angenommen, eine Hausbesitzerin möchte ihr Haus verkaufen. Sie erhält zu den (diskreten) Zeitpunkten $0, 1, \dots, N - 1$ jeweils ein Angebot der Höhe y_n GE, das als Realisierung einer Zufallsvariable Y_n mit Werten in $\{0, \dots, M\}$ aufgefasst werden kann, $n \in \{0, 1, \dots, N - 1\}$.

Annahme: $(Y_n)_{0 \leq n \leq N-1}$ sind stochastisch unabhängig und identisch verteilt mit

$$\forall y \in \{0, \dots, M\} : P(Y_0 = y) =: q(y).$$

Nimmt die Verkäuferin zum Zeitpunkt n das vorliegende Angebot y_n an, so erhält sie den Verkaufserlös y_n GE, und das Entscheidungsproblem ist abgeschlossen. Lehnt sie das Angebot ab, so entstehen ihr Kosten in Höhe von c GE (z. B., erneutes Inserat, laufende Betriebskosten, etc.). Außerdem kann sie auf ein Angebot zu keinem späteren Zeitpunkt mehr zurückgreifen. Dies ist ein Kontrollproblem mit den folgenden Parametern.

- (i) $N \in \mathbb{N}$ (Stufen $n = 0, 1, \dots, N$).
- (ii) $\mathcal{S} \equiv \mathcal{S}_n = \{0, \dots, M, +\infty\}$, $0 \leq n \leq N$. Die Zahl $s_n \in \{0, \dots, M\}$ bezeichnet die Höhe des zum Zeitpunkt n vorliegenden Angebots, und $s_n = +\infty$ ist ein zusätzlicher Zustand, der für den bereits vollzogenen Verkauf des Hauses steht.
- (iii) $\mathcal{A} \equiv \mathcal{A}_n(s) = \{0, 1\}$, $0 \leq n \leq N - 1$, $s \in \mathcal{S}$. Die Aktion $a = 0$ steht für Ablehnung des vorliegenden Angebots, die Aktion $a = 1$ für dessen Annahme. Für $s = +\infty$ sind Aktionen ohne inhaltliche Bedeutung.
- (iv) $\mathcal{Y} \equiv \mathcal{Y}_n = \{0, \dots, M\}$, $0 \leq n \leq N - 1$.
- (v) $q_n(y) \equiv q(y)$, $0 \leq n \leq N - 1$, $y \in \mathcal{Y}$.

(vi)

$$g_n(s, a, y) = \begin{cases} y, & s \leq M \text{ und } a = 0, \\ +\infty, & \text{sonst.} \end{cases}$$

(vii)

$$r_n(s, a) = \begin{cases} -c, & s \leq M \text{ und } a = 0, \\ s, & s \leq M \text{ und } a = 1, \\ 0, & \text{sonst.} \end{cases}$$

(viii) $V_n(s) \equiv 0, \forall s \in \mathcal{S}$.

Wir setzen $V_n(+\infty) = 0$ für alle $0 \leq n \leq N$ und erhalten die Optimalitätsgleichung

$$V_n(s) = \max \left\{ s, -c + \sum_{y=0}^M q_n(y) V_{n+1}(y) \right\}, \quad s \leq M \text{ und } 0 \leq n \leq N-1.$$

Dabei steht der erste Ausdruck in der geschweiften Klammer für Annehmen ($a_n = 1$) und der zweite für Ablehnen ($a_n = 0$). Sei nun

$$s_n^* = -c + \sum_{y=0}^M q(y) V_{n+1}(y), \quad 0 \leq n \leq N-1.$$

Dann ist die aus den Entscheidungsfunktionen

$$f_n^*(s) = \begin{cases} 1, & s \geq s_n^* \\ 0, & s < s_n^*, \end{cases}$$

mit $0 \leq n \leq N-1$ gebildete Strategie $\delta^* = (f_0^*, \dots, f_{N-1}^*)$ nach Satz 4.13.(ii) optimal. Ein "gutes" Angebot ($s \geq s_n^*$) wird angenommen, während bei einem "schlechten" Angebot ($s < s_n^*$) weiter gewartet wird.

Numerisches Beispiel:

Seien $\mathcal{Y} = \{300, 350, 400, 450\}$ [1000 EUR], sowie

$$q(300) = 0.2, \quad q(350) = 0.3, \quad q(400) = 0.4, \quad q(450) = 0.1.$$

Ferner betrachten wir einen Planungshorizont von $N = 10$ (Wochen), und setzen $c = 0$ GE (Kosten bei Ablehnung).

Optimalitätsgleichung:

$$V_n(s) = \max \left\{ s, \sum_{s' \in \{300, 350, 400, 450\}} q(s') \cdot V_{n+1}(s') \right\}$$

	$s = 300$	$s = 350$	$s = 400$	$s = 450$
$V_9(s)$	300*	350	400	450
$V_8(s)$	370	370	400*	450
$V_7(s)$	390	390	400*	450
$V_6(s)$	400	400	400*	450
$V_5(s)$	405	405	405	450*
$V_4(s)$	409,5	409,5	409,5	450*
$V_3(s)$	413,55	413,55	413,55	450*
$V_2(s)$	417,20	417,20	417,20	450*
$V_1(s)$	420,48	420,48	420,48	450*
$V_0(s)$	423,43	423,43	423,43	450*

Markiert ist in jeder Zeile das niedrigste mögliche Angebot, das mit der optimalen Strategie δ^* akzeptiert wird.

Beispiel 4.17 (Aktienoptionen)

Eine Investorin habe die Option für den Kauf einer Aktie in Höhe von c Geldeinheiten. Das Optionsrecht kann sie zu den (diskreten) Zeitpunkten $n = 0, 1, \dots, N - 1$ ausüben, danach verfällt die Option. Der (zufällige) Börsenkurs I_n der Aktie zum Zeitpunkt n sei modelliert durch $I_n = s_0 + \sum_{t=0}^{n-1} Y_t$. Dabei seien $(Y_t)_{t \geq 0}$ stochastisch unabhängige, identisch verteilte Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit

$$\mathbb{P}(Y_0 = y) =: q(y), \quad y \in \mathbb{Z}.$$

Gesucht ist eine optimale Kaufstrategie für die Investorin. Dies ist ein Kontrollmodell mit den folgenden Komponenten:

- (i) Planungshorizont $N \in \mathbb{N}$.
- (ii) $\mathcal{S} \equiv \mathcal{S}_n = \mathbb{Z} \cup \{\infty\}$, $0 \leq n \leq N$. Die Zahl $s_n < \infty$ bezeichnet den Kaufpreis (Börsenkurs) zum Zeitpunkt n , und $s_n = \infty$ ist ein zusätzlicher Zustand, der für den bereits vollzogenen Kauf steht.
- (iii) $\mathcal{A} \equiv \mathcal{A}_n(s) = \{0, 1\}$ für $0 \leq n \leq N$ und $s \in \mathcal{S}$ ist der binäre Aktionsraum. Dabei bedeutet $a_n = 1$ "Kaufrecht zum Zeitpunkt n ausüben".
- (iv) $\mathcal{Y} \equiv \mathcal{Y}_n = \mathbb{Z}$ (Wertebereich von Y_n), $0 \leq n \leq N - 1$.
- (v) $q_n(y) \equiv q(y)$, $0 \leq n \leq N - 1$, $y \in \mathbb{Z}$.
- (vi)

$$g_n(s, a, y) = \begin{cases} s + y, & s < \infty \text{ und } a = 0, \\ \infty, & \text{sonst.} \end{cases}$$

(vii)

$$r_n(s, a) = \begin{cases} s - c, & s < \infty \text{ und } a = 1, \\ 0, & \text{sonst.} \end{cases}$$

(viii) $V_N(s) \equiv 0 \forall s \in \mathcal{S}_N$.

Zur Ermittlung der zugehörigen Optimalitätsgleichung beachten wir, dass $V_n(\infty) = 0$ für alle $0 \leq n \leq N$ gilt. Wir erhalten damit:

$$\begin{aligned} \forall s \in \mathbb{Z}: \quad V_n(s) &= \max\{s - c, J_n(s)\}, \quad \text{wobei} \\ J_n(s) &= \sum_{y \in \mathbb{Z}} q(y) V_{n+1}(s + y). \end{aligned}$$

Der erste Ausdruck lässt sich interpretieren als der momentan erzielbarer Gewinn bei Ausübung des Optionsrechts ($a_n = 1$), und $J_n(s)$ als der erwartete Gewinn bei einer späteren Ausübung des Optionsrechts ($a_n = 0$). Im Falle eines "hohen" Aktienkurses wird das Optionsrecht ausgeübt, anderenfalls nicht.

Lemma 4.18

Für alle $0 \leq n \leq N - 1$ und alle $s \in \mathbb{Z}$ gilt

$$(i) \quad 0 \leq J_n(s + 1) - J_n(s) \leq 1,$$

$$(ii) \quad 0 \leq V_n(s + 1) - V_n(s) \leq 1.$$

Beweis: Der Beweis wird per Rückwärtsinduktion geführt.

Induktionsanfang ($n = N - 1$):

Wegen $V_N(s) = 0 \forall s \in \mathbb{Z}$ gilt $J_{N-1}(s) = 0$ für alle $s \in \mathbb{Z}$ und daher $V_{N-1}(s) = \max\{s - c, 0\}$, $s \in \mathbb{Z}$, was (i) und (ii) für $n = N - 1$ impliziert.

Induktionsschritt ($n + 1 \rightarrow n$):

Angenommen, (i) und (ii) gelten für ein $(n + 1) < N$. Da q eine Zähldichte ist, erhalten wir

$$0 \leq J_n(s + 1) - J_n(s) = \sum_{y \in \mathbb{Z}} q(y) \{V_{n+1}(s + 1 + y) - V_{n+1}(s + y)\} \leq 1,$$

und somit (i) für den Index n . Ferner ist

$$\begin{aligned} V_n(s + 1) - V_n(s) &= \max\{s + 1 - c, J_n(s + 1)\} - \max\{s - c, J_n(s)\} \\ &\leq \max\{1, J_n(s + 1) - J_n(s)\} \leq 1 \end{aligned}$$

und analog

$$V_n(s+1) - V_n(s) \geq \min\{1, J_n(s+1) - J_n(s)\} \geq 0.$$

Damit folgt die Aussage für alle $0 \leq n \leq N-1$ und alle $s \in \mathbb{Z}$. ■

Korollar 4.19

Lemma 4.18 impliziert, dass J_n langsamer (bzw. nicht schneller) in s wächst als $(s-c)$. Seien nun

$$\forall 0 \leq n \leq N-1: \quad \mathcal{M}_n := \{s \mid s-c \geq J_n(s)\} \text{ und}$$

$$s_n^* := \begin{cases} \min \mathcal{M}_n, & \mathcal{M}_n \neq \emptyset, \\ \infty, & \mathcal{M}_n = \emptyset. \end{cases}$$

Dann sind optimale Entscheidungsfunktionen gegeben durch

$$\forall 0 \leq n \leq N-1: f_n^*(s) := \begin{cases} 1, & s \geq s_n^*, \\ 0, & s < s_n^*. \end{cases}$$

und eine optimale Strategie ist gegeben durch $\delta^* = (f_0^*, f_1^*, \dots, f_{N-1}^*)$. Der Wert s_n^* ist also ein Schwellenwert für den Börsenkurs der Aktie zum Zeitpunkt n , ab dem das Kaufrecht ausgeübt werden sollte ($a_n = 1$).

Beispiel 4.20 (Beispiel 0.4 fortgesetzt)

Wir betrachten noch einmal das Ersetzungsproblem für eine Maschine aus Beispiel 0.4. Dies lässt sich als stochastisches dynamisches Optimierungsproblem formalisieren, mit den folgenden Komponenten.

- (i) Planungshorizont $N \in \mathbb{N}$ (Stufen: $0 \leq n \leq N$).
- (ii) $\mathcal{S} \equiv \mathcal{S}_n = \{0, \dots, M\}$, $0 \leq n \leq N$. Der Wert s_n bezeichnet den Zustand (Verschleiß) der Maschine zum Zeitpunkt n . Dabei bedeutet $s_n = 0$, dass die Maschine neu ist.
- (iii) $\mathcal{A} \equiv \mathcal{A}_n(s) = \{0, 1\}$, $0 \leq n \leq N-1$, $s \in \mathcal{S}$. Dabei bedeutet $a = 1$ „Maschine ersetzen“ und $a = 0$ „Maschine weiter betreiben“.
- (iv)

$$p_n(s, a, s') = \begin{cases} q(s, s') & \text{für } a = 0, \\ q(0, s') & \text{für } a = 1. \end{cases}$$

Dabei ist $q(\cdot, \cdot)$ eine Übergangswahrscheinlichkeitsfunktion, so dass

$$\mathbb{P}(I_{n+1} = s' \mid I_n = s) = q(s, s')$$

ist.

(v)

$$r_n(s, a) = \begin{cases} -c(s) & \text{für } a = 0, \\ -k - c(0) & \text{für } a = 1. \end{cases}$$

Dabei gibt $c(s)$ die Betriebskosten (in GE) für eine Maschine in Zustand s an, und k (in GE) die Ersetzungskosten. Die negativen Vorzeichen bedeuten, dass Kosten als negative Gewinne interpretiert werden.

(vi) $V_N(s) \in \mathbb{R}$, $s \in \mathcal{S}$, ist der Restwert der Maschine am Ende des Planungshorizontes.

Die zugehörige Optimalitätsgleichung lautet gemäß (4.5) aus Satz 4.13 für alle $0 \leq n \leq N - 1$:

$$V_n(s) = \max\left\{-c(s) + \sum_{s'=0}^M q(s, s')V_{n+1}(s'), -k - c(0) + \sum_{s'=0}^M q(0, s')V_{n+1}(s')\right\},$$

wobei der erste Term zur Aktion $a_n = 0$ und der zweite Term zur Aktion $a_n = 1$ gehört.

Es ist in diesem Kontext natürlicher, die erwarteten Gesamtkosten $G_n(s) := -V_n(s)$ zu betrachten. Dann gilt

$$\begin{aligned} G_n(s) &= \min\{J_n(s), k + J_n(0)\}, \quad \text{wobei} & (4.9) \\ J_n(s) &:= c(s) + \sum_{s'=0}^M q(s, s')G_{n+1}(s'). \end{aligned}$$

Der erste Term in (4.9) gibt die erwarteten Folgekosten bei Nicht-Ersetzung an, der zweiten Term die erwarteten Folgekosten bei Ersetzung.

Seien nun

$$\begin{aligned} \forall 0 \leq n \leq N - 1: \quad \mathcal{M}_n &= \{s \mid J_n(s) \geq k + J_n(0)\} \quad \text{und} \\ s_n^* &= \begin{cases} \min \mathcal{M}_n, & \mathcal{M}_n \neq \emptyset, \\ \infty, & \mathcal{M}_n = \emptyset. \end{cases} \end{aligned}$$

Dann sind optimale Entscheidungsfunktionen $f_0^*, f_n^*, \dots, f_{N-1}^*$ gegeben durch

$$f_n^*(s) = \begin{cases} 1, & s \geq s_n^*, \\ 0, & s < s_n^*. \end{cases}$$

Man ersetzt die Maschine zu dem Zeitpunkt n , an dem der Verschleiß s_n erstmals die Eingriffsgrenze s_n^* erreicht oder überschritten hat.

Dabei wurden die folgenden Annahmen implizit vorausgesetzt.

(A1) $c(s) \leq c(s + 1)$

(verschlissene Maschinen erzeugen höhere Betriebskosten),

(A2) $V_N(s) \geq V_N(s + 1)$

(verschlissene Maschinen haben niedrigeren Restwert),

(A3) $\forall 0 \leq i \leq M: \sum_{s'=i}^M q(s, s') \leq \sum_{s'=i}^M q(s + 1, s')$

(der Zustand einer Maschine verschlechtert sich systematisch).

Literaturverzeichnis

- Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press.
- Eiselt, H. A. and C.-L. Sandblom (2012). *Operations research. A model-based approach. 2nd ed.* Berlin: Springer.
- Ellinger, T., G. Beuermann, and R. Leisten (2003). *Operations research. Eine Einführung. 6., durchgesehene Aufl.* Berlin: Springer.
- Heuser, H. (2009). *Lehrbuch der Analysis. Teil 1. 17. Auflage.* Wiesbaden: Vieweg+Teubner.
- Heyman, D. P. and M. J. Sobel (1982). Stochastic models in operations research. Vol. I: Stochastic processes and operating characteristics. McGraw-Hill Series in Quantitative Methods for Management. New York etc.: McGraw-Hill Book Company.
- Heyman, D. P. and M. J. Sobel (1984). Stochastic models in operations research. Vol. II: Stochastic optimization. McGraw-Hill series in quantitative methods for management. New York etc.: McGraw-Hill Book Company.
- Jarre, F. and J. Stoer (2004). *Optimierung*. Berlin: Springer.
- Jungnickel, D. (2015). *Optimierungsmethoden. Eine Einführung. 3.Auflage.* Heidelberg: Springer Spektrum.
- Kathöfer, U. and U. Müller-Funk (2008). *Operations Research. 2. Auflage.* UTB GmbH.
- Kohlas, J. (1977). Stochastische Methoden des Operations Research. Leitfäden der angewandten Mathematik und Mechanik. Bd. 40. Teubner Studienbücher: Mathematik. Stuttgart: B.G. Teubner.
- Nickel, S., O. Stein, and K.-H. Waldmann (2014). *Operations Research. 2. Auflage.* Berlin: Springer.