

Operations Research

Prof. Jürgen Sauer

Operations Research

Vorlesung im Sommer-Semester 2005

Inhaltsverzeichnis

1. Merkmale des Operations Research.....	6
1.1 Definition	6
1.2 Problemstellungen.....	6
1.3 Problemlösungen.....	8
2. Lineare Planungsrechnung.....	9
2.1 Einführung	9
2.1.1 Bestimmen des Problembereichs	9
2.1.1.1 Sind Entscheidungen Glücksache?	9
2.1.1.2 Vom Zufall zur Methode.....	10
2.1.1.3 Von der Methode zum praktischen Verfahren	11
2.1.2 Beispiele für Optimierungsprobleme	12
2.1.2.1 Produktionsprobleme	12
2.1.2.2 Mischungsprobleme	13
2.1.2.3 Verschnittprobleme.....	15
2.1.2.4 Verteilungsprobleme.....	16
2.1.2.5 Einsatzmöglichkeiten von OR-Modellen im DV-Management	17
2.1.3 Nichtlineare und lineare Optimierungsprobleme	19
2.1.4 Nichtlineare, lineare Optimierungsprobleme und Extremalwert-probleme	21
2.1.4.1 Einführende Beispiele	21
2.1.4.1.1 Optimale Lagerauffüllung vor Preiserhöhungen.....	21
2.1.4.1.2 Die klassische Bestellmengenformel (Losgrößenformel)	22
2.1.4.2 Lagerhaltungsprobleme.....	24
2.1.4.3 Abgrenzung der Extremalwertprobleme zur linearen Optimierung.....	33
2.2 Grafische Lösungsverfahren	34
2.2.1 Normalformen der linearen Planungsrechnung mit 2 Variablen.....	34
2.2.1.1 Die Normalform der Maximumaufgabe.....	34
2.2.1.2 Normalformen der Minimumaufgabe	36
2.2.2 Die Lösungsmenge für Probleme mit 2 Variablen.....	37
2.2.2.1 Probleme mit mehr als einer Lösung.....	37
2.2.2.2 Probleme ohne optimale Lösung.....	38
2.2.2.3 Probleme mit genau einer Lösung.....	38
2.2.2.4 Charakterisierung der Lösungsmenge.....	38
2.2.2.5 Basislösungen und zulässige Basislösungen	41
2.2.2.6 Die optimale Lösung	42
2.2.3 Sonderfälle	44
2.2.3.1 Die ganzzahlige Planungsrechnung	44
2.2.3.2 Die Parametrische Planungsrechnung.....	45
2.3 Das Simplexverfahren.....	48
2.3.1 Einführung: Ein Beispiel aus der Produktion.....	48
2.3.2 Das Simplexverfahren für die Normalform der Maximumaufgabe	53
2.3.2.1 Das mathematische Modell für die Normalform mit n Variablen.....	53
2.3.2.2 Die Basistransformation.....	54
2.3.2.3 Zulässige Basislösungen	56
2.3.2.3.1 Der Nullpunkt	56
2.3.2.3.2 Berechnung zulässiger Basislösungen	57
2.3.2.3.3 Eckpunkte	58
2.3.2.3.4 Das Maximum.....	60
2.3.2.4 Bestimmen der optimalen Lösung.....	61
2.3.3 Das Dualitätsproblem.....	62
2.3.3.1 Ein einführendes Beispiel	62
2.3.3.2 Das Dualitätstheorem.....	64
2.3.3.2.1 Das duale lineare Optimierungsprobleme	64
2.3.3.2.2 Das Optimalitätskriterium	65
2.3.3.2.3 Der Dualitätssatz	66
2.4 Anwendung des Simplexverfahrens.....	68
2.4.1 Das verkürzte Simplex-Tableau.....	68
2.4.1.1 Ein einführendes Beispiel	68

2.4.1.2 Allgemeine Darstellung	69
2.4.1.3 Steepest-Unit Ascent.....	70
2.4.1.4 Die "Greatest Change"-Version	74
2.4.1.5 Sonderfälle	77
2.4.1.5.1 Entartung	77
2.4.1.5.2 Unbegrenzte Lösungen	77
2.4.1.6 Das Simplexverfahren in Java.....	78
2.4.1.7 Simplex-Optimierung mit dem Excel-Solver.....	82
2.4.2.1 ">=" -Restriktionen	84
2.4.2.2 Die duale Simplex-Methode	86
2.4.2.2.1 Das zweidimensionale Verschnittproblem.....	86
2.4.2.2.2 Das eindimensionale Verschnittproblem	89
2.4.2.3 Gleichungen als Restriktionen	93
2.4.2.4 Zusammenfassung: Ablauf des Simplex-Verfahrens	96
2.4.3 Unter- und Obergrenzen von Variablen.....	97
2.4.3.1 Vereinbarungen.....	97
2.4.3.2 Untergrenzen von einzelnen Variablen $x_j \geq l_j$	97
2.4.3.3 Obergrenzen von einzelnen Variablen	99
3. Zweipersonen-Nullsummenspiele.....	104
3.1 Grundlagen.....	104
3.1.1 Matrixspiele	104
3.1.2 Wert eines Spiels und optimale Strategien.....	106
3.1.3 Ermittlung des Sattelpunkts für Matrixspiele.....	107
3.1.4 Gemischte Strategien	110
3.1.5 Bestimmen des Spielwerts bei gemischter Erweiterung	111
3.2 Lösung mit dem Simplex-Verfahren.....	112
3.2.1 Grundlagen.....	112
3.2.2 Beispiele.....	114
4. Graphen und Graphenalgorithmien.....	121
4.1 Theorie und praktische Bedeutung	121
4.1.1 Einführung	121
4.1.2 Anwendungsgebiete aus dem Bereich des OR.....	122
4.1.3 Definition und Darstellung von Graphen.....	123
4.1.4 Darstellung in Rechnerprogrammen	132
4.2 Grundlegende Algorithmen.....	136
4.2.1 Die Ermittlung kürzester bzw. längster Wege nach dem Algorithmus von Ford	136
4.2.2 Netzplantechnik	141
4.2.2.1 Grundlagen.....	141
4.2.2.2 CPM	142
4.2.2.3 MPM	152
4.2.3 Flüsse in Transportnetzen	164
4.2.3.1 Das Problem des maximalen Flusses (Algorithmus von Ford / Fulkerson).....	164
4.2.3.2 Kostenminimaler Gesamtfluß	172
4.2.3.2.1 Einführung	172
4.2.3.2.2 Transportprobleme	173
4.2.3.2.3 Das Umladeproblem.....	182
4.2.3.2.4 Zuordnungsprobleme	190
4.2.3.2.5 Das Travelling-Salesman-Problem	199
4.2.3.2.6 Das Chinese Postman's Problem.....	201
5. Warteschlangenmodelle	205
5.1 Stochastische Prozesse.....	205
5.1.1 Einführende Beispiele	205
5.1.2 Wahrscheinlichkeitstheoretische Grundlagen.....	206
5.1.2.1 Zufallsvariable	206
5.1.2.2 Mehrdimensionale Zufallsvariable, Abhängigkeiten zwischen mehreren Zufallsvariablen	209
5.1.2.3 Stochastische Prozesse.....	210
5.1.2.4 Markov-Prozesse.....	210
5.2 Elementare Wartesysteme.....	228
5.2.1 Beschreibung.....	228
5.2.2 Leistungsgrößen	228
5.2.3 Charakterisierung (Kurzschreibweise) von Wartesystemen	230

5.3 Bedienungssysteme	232
5.3.1 Markov'sche Bedienungssysteme.....	233
5.3.1.1 Stationäre Verteilungen.....	233
5.3.1.1.1 M/M/m-Systeme	233
5.3.1.1.2 Systeme mit beschränktem Zugang.....	238
5.3.1.2 Wartezeitenverteilungen	241
5.3.2 Allgemeine Bedienungssysteme	241
5.4 Warteschlangennetze	242
5.4.1 Grundlagen.....	242
5.4.2 Produktformlösungen.....	245
5.4.2.1 Jackson-Theorem für offene Netze	245
5.4.2.2 Gordon/Newell-Theorem für geschlossene Netze	245
6. Simulation	246
6.1 Definition und Klassifizierung von Simulationsmethoden	246
6.1.1 Definition	246
6.1.2 Klassifizierung	247
6.2 Simulationsstudie	248
6.2.1 Fallbeispiel	248
6.2.2 Manuelle Simulation	249
6.2.3 Allgemeine Vorgehensweise bei einer Simulationsstudie	251
6.2.4 Werkzeuge der Simulation.....	253
6.3 Modelltypen der digitalen Simulation.....	254
6.3.1 Determinierte Modelle	254
6.3.2 Stochastische Modelle.....	255
6.3.2.1 Beschreibung und Vorgehensweise	255
6.3.2.2 Stochastische Methoden in der Simulation diskreter Systeme.....	256
6.3.2.3 Die Erzeugung von Zufallszahlen	258
6.3.2.4 Verteilungen.....	261
6.3.2.4.1 Mathematische Grundlagen	261
6.3.2.4.2 Gleichverteilung.....	262
6.3.2.4.3 Exponentialverteilung und verwandte Verteilungen.....	263
6.3.2.4.4 Normalverteilung und verwandte Verteilungen.....	265
6.3.2.4.5 Allgemeine Methoden zur Erzeugung von Zufallszahlen.....	266
6.4 Diskrete Simulation.....	268
6.4.1 Modellierung und Sichtweise.....	268
6.4.2 Komponenten ereignisorientierter Simulationen	268
6.4.3 Typische Modellklassen.....	271
6.4.3.1 Bedienungs-/ Wartesysteme.....	271
6.4.3.2 Lagerhaltungssysteme	271

Empfohlene Literatur

- Churchman, C.W. / Ackoff, R.L. / Arnoff, E.L.: Operations Research, Oldenbourg, Wien und München, 1964
Desbazeille, G.: Unternehmensforschung, Stuttgart , 1970
Domschke, W. / Drexl, A.: Operations Research, Springer, Berlin, 1990
Müller-Merbach, H.: Operations Research, 3. Auflage, München, 1973
Wille, H. / Gewalt, K. / Weber, H.D.: Netzplantechnik, München Wien 1966
Grams, Timm: Simulation, Mannheim, Leipzig, Wien, Zürich 1992
Pflug, Georg: Stochastische Modelle in der Informatik, Stuttgart 1986
Bolch, Gunter: Leistungsbewertung von Rechensystemen, Stuttgart 1989

1. Merkmale des Operations Research

1.1 Definition

Operations Research ist ein angloamerikanischer Begriff. Synonym ist hierzu der deutsche Begriff: **Optimalplanung**. Eine **Optimalplanung** besitzt folgende Eigenschaften:

- (1) Vorbereitung optimaler Entscheidungen
- (2) Verwendung mathematischer Modelle zur Optimierung

1.2 Problemstellungen

Allgemein sind Gegenstand von OR-Untersuchungen: Probleme der wirtschaftlichen, gesellschaftlichen Praxis, die sich durch mathematische Modelle beschreiben lassen. Solche Probleme können sein:

(1) Zuteilungsprobleme

Eine vorgegebene Leistung ist durch den Einsatz beschränkter Mittel auf wirtschaftliche Weise zu erzielen bzw. mit gegebenen Mitteln ist ein maximaler Ertrag zu erzielen.

Am häufigsten zur Lösung von Zuteilungsproblemen verwendete mathematische Verfahren: Lineare Planungsrechnung (lineare Optimierung, lineare Programmierung)

(2) Konkurrenzprobleme

Entscheidungen des einen Partners werden durch Entscheidungen des anderen Partners beeinflusst.

Konkurrenzprobleme werden mit Hilfe von Spielen (z.B. Zweipersonen-Nullsummenspiele¹) beschrieben, d.h. charakterisiert durch eine bestimmte Anzahl von Spielern, Spielregeln, Gewinn und Verlust.

Das grundlegende mathematische Verfahren ist in der Spieltheorie beschrieben.

(3) Lagerhaltungsprobleme

In diesen Problemen sind die Kosten der Lagerung abzuwägen gegen Auftragskosten bzw. Bestellkosten und Kosten, die durch verzögerte Lieferung wegen Erschöpfung des Lagerbestands entstehen.

Mathematische Hilfsmittel sind hier: Gleichungssysteme und Verfahren der linearen und dynamischen Planungsrechnung.

¹ Vgl. Kap.3

(4) Wartezeitprobleme

Personen und Güter werden durch (eine oder mehrere) Stellen abgefertigt. Von Ausnahmefällen abgesehen, müssen dabei zu bedienende Einheiten oder Bedienungsstellen warten. Es entstehen Kosten. Die Summe dieser Kosten soll einen möglichst geringen Wert annehmen.

Mathematische Verfahren zur Behandlung derartiger Probleme stellen zur Verfügung:

1. Warteschlangentheorie, z.B.

- zur Bestimmung der Anzahl der Bedienungsstellen
- zur Festlegung der Zeitpunkte, zu denen die abzufertigenden Einheiten voraussichtlich eintreffen

2. Optimale Reihenfolgen

Optimale Bestimmung der Reihenfolge in der bereitstehende Einheiten zur Bedienung herangezogen werden. Die Zahl der Bedienungsstellen liegt hier fest, die Reihenfolge ist dagegen beeinflussbar.

Bsp.: Ablaufplanung für Produktserien über eine Reihe von Maschinen mit dem Ziel eine minimale Gesamtlaufzeit zu erreichen.

Dazu verwandte Probleme:

- Das Abstimmen der einzelnen Arbeitsgänge an einem Montageband mit dem Ziel eine minimale Gesamtlaufzeit zu erreichen
- Eine Route für eine Reihe von Orten so festlegen, daß die zurückgelegte Entfernung ein Minimum wird.

(5) Ersatzprobleme

1. Die Leistungsfähigkeit der Einrichtungen nimmt allmählich ab. Es ist der optimale Einsatzpunkt zu bestimmen durch Minimieren der Kosten für eine neue Einrichtung bzw. der Kosten zur Erhaltung der Leistungsfähigkeit der alten Anlage.
2. Die Einrichtungen fallen plötzlich vollständig aus.

Hier ist festzustellen:

Welche Stücke sind zu ersetzen? (z.B. alle außer in der letzten Woche eingesetzten Stücke). Wie oft sollen diese Einrichtungen ersetzt werden?

Die Fragen werden mit Hilfe des folgenden Entscheidungskriteriums beantwortet:

- Kosten, die durch die Einrichtung entstehen
- Kosten, die durch den Ausfall entstehen

1.3 Problemlösungen

Mit Hilfe eines Algorithmus wird das mathematische Modell unter Verwendung der Daten gelöst. OR im weitesten Sinn beschäftigt sich mit Modellbildung und Lösungsfindung (Entwicklung und / oder Anwendung von Algorithmen) sowie Methoden zur Datenermittlung.

Modelle spielen im OR eine zentrale Rolle. Ein Modell ist ein vereinfachtes Abbild eines realen Systems oder Problems. OR benutzt im wesentlichen Entscheidungs- bzw. Optimierungs- – sowie Simulationsmodelle.

Ein Entscheidungs- bzw. Optimierungsmodell ist eine (formale) Darstellung eines Entscheidungs- oder Planungsproblems, das in seiner einfachsten Form mindestens eine Alternativmenge und eine bewertende Zielfunktion enthält. Es wird aufgestellt, um mit geeigneten Verfahren optimale und suboptimale Lösungsvorschläge ermitteln zu können.

Ein **Optimierungsmodell** läßt sich folgendermaßen beschreiben:

Maximiere (oder Minimiere) $z = G(\vec{x})$ unter den Nebenbedingungen²

$$f_i(\vec{x}) \begin{cases} \geq \\ = \\ \leq \end{cases} 0 \quad \text{für } i = 1, 2, \dots, m$$

Dabei sind:

\vec{x} ein Variablenvektor mit n Komponenten³

$G(\vec{x})$ eine Zielfunktion.

Häufig sind aber die unter 1.2 angegebenen Probleme so komplex, daß zu den mathematischen Modellen keine bzw. nur sehr komplizierte, direkte analytische Lösungsmethoden vorliegen. In solchen Fällen setzt man auf Simulationsverfahren. Darunter wird das zielgerichtete Experimentieren an Modellen, die der Wirklichkeit nachgebildet sind verstanden. Durch die Simulation, d.h. die Bearbeitung von Modellen bei zielgerichteter Veränderung der Einflußgrößen, sollen Rückschlüsse auf das reale System möglich werden.

Simulation als Methode des OR wird dann angewandt, wenn sich das Problem nicht durch ein mathematisches Modell beschreiben läßt oder wenn es kein analytisches Lösungsverfahren gibt oder wenn ein solches einen zu hohen Rechenaufwand erfordern würde. Zielsetzung der Simulation ist das Bestimmen sog. Suboptima, z.B. die optimale Bestellmenge, optimale Ersatzzeitpunkte, u. ä. Häufig werden lediglich suboptimale Lösungen erreicht, da die Modelle nur Teilzusammenhänge realer Systeme nachbilden. Simulationen gehören somit zu den heuristischen Verfahren⁴.

² System von m Gleichungen und / oder Ungleichungen

³ Alle im Vektor \vec{x} zusammengefaßten Variablen sind in der Regel nichtnegative reelle bzw. nichtnegative ganzzahlige bzw. binäre Werte.

⁴ Verfahren zur Lösung komplexer Systeme oder Entscheidungsaufgaben. Werden vor allem für solche Optimierungsaufgaben angewandt, bei denen exakte Lösungen mit vertretbarem Rechenaufwand nicht möglich sind.

2. Lineare Planungsrechnung

2.1 Einführung

2.1.1 Bestimmen des Problembereichs

2.1.1.1 Sind Entscheidungen Glücksache?

„**OR** oder **Optimalplanung** bedeuten das Vorbereiten **optimaler Entscheidungen**“. Entscheidungen sind aber häufig von zufälligen Ereignissen abhängig, d.h. "Entscheidungen sind Glücksache".

1) Gegeben ist das folgende Problem:

Aus einem bestimmten Kontigent an Rohmaterial können 2 Erzeugnisse (Produkte) hergestellt werden. Der Stückgewinn des einen Produkts (Produkt 2) ist doppelt so hoch wie der Stückgewinn des anderen Produkts (Produkt 1). Die aus dem gegebenen Rohmaterial mögliche Produktion ist in jedem Fall ganz absetzbar.

Wieviel ist von beiden Produkten zu fertigen, wenn ein optimaler Gewinn erreicht werden soll?

Dispositionsregel (1): Es ist das Produkt zu fertigen, das den höchsten Gewinn abwirft.

Das bedeutet: Nur das Produkt 2 ist zu fertigen. Ist diese Entscheidung richtig? Sie kann u.U. falsch sein.

2) Das vorhandene Rohmaterialkontigent ist:

Rohmaterial 1: 6t
Rohmaterial 2: 10t
Rohmaterial 3: 16t

Davon wird für die Fertigung je eines Stücks benötigt:

Produkt 1: 1t Rohmaterial 1
 2t Rohmaterial 2
 1t Rohmaterial 3
Produkt 2: 1t Rohmaterial 1
 1t Rohmaterial 2
 3t Rohmaterial 3

Fertigt man 5 Stück von Produkt 2, dann bleiben noch übrig:

Rohmaterial 1: 1t
Rohmaterial 2: 5t
Rohmaterial 3: 1t

Diese Rohmaterialien reichen zur Fertigung von Produkt 2 nicht mehr aus, wohl aber zur Herstellung eines Stücks von Produkt 1.

Dispositionsregel (1) ist zu diesem speziellen Fall falsch.

Dispositionsregel (2): Es wird soviel wie möglich von dem Produkt gefertigt, das den höchsten Gewinn abwirft. Vom etwa übrig gebliebenem Rohmaterial wird soviel wie möglich vom anderen Produkt erzeugt.

Auch diese Regel ist nicht allgemeingültig.

3) Zur Fertigung eines Produkts A bzw. B braucht man

Produkt A: 1t Rohmaterial a
 1t Rohmaterial b
 1t Rohmaterial c

Produkt B: 4t Rohmaterial a
 4t Rohmaterial b
 4t Rohmaterial c

Der Gewinn für ein Stück von Produkt B ist wieder doppelt so groß wie für ein Stück von Produkt A (z.B. 200.- DM gegenüber 100.- DM).

Nach Dispositionsregel (2) ist der Gesamtgewinn = 400.- DM. Es könnte ein Höchstgewinn von 600.- DM erreicht werden, wenn 6 Stück von Produkt A hergestellt werden.

Dispositionsregel (3): Fertigung des Produkts, das den niedrigeren Stückgewinn bringt.

Offensichtlich ist es Glücksache, eine Dispositionsregel zu finden, die zum optimalen Ergebnis führt.

2.1.1.2 Vom Zufall zur Methode

Eine Methode in jedem Fall eine optimale Dispositionsregel zu finden, besteht aus folgenden Schritten:

(1) Stelle systematisch sämtliche Möglichkeiten (der Fertigung von Produkten) auf unter Berücksichtigung von

1. Materialanforderungen
2. Beschränkung der Materialkapazität

(2) Rechne für jede Möglichkeit den Gewinn aus

(3) Wähle die, oder, wenn es mehrere gibt, eine optimale (Fertigungs-) Kombination aus.

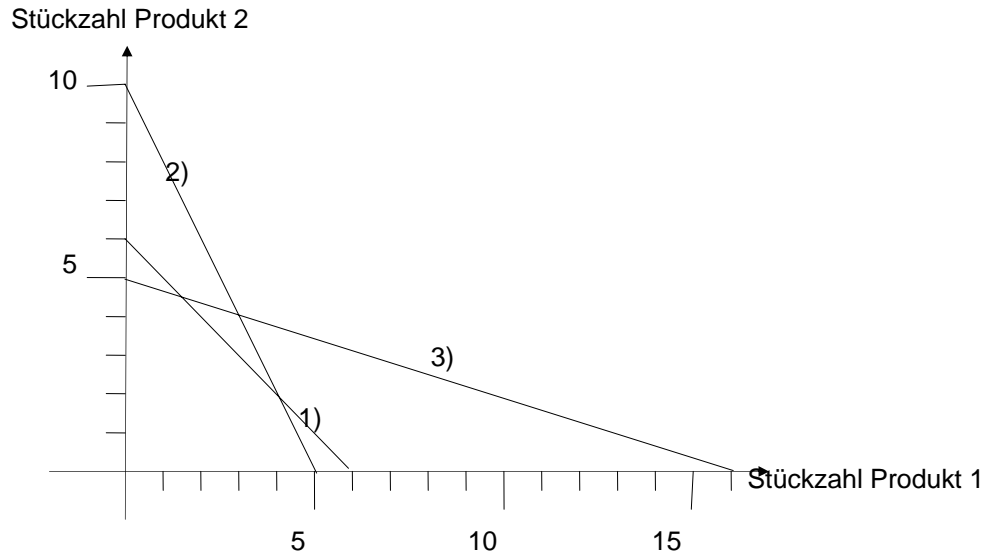
Anwendung auf das gegebene Problem

(1) systematische Zusammenstellung der Möglichkeiten

- 1) Möglichkeiten, die die Rohmaterialkapazität 1 zur Fertigung bietet.
- 2) Möglichkeiten, die die Rohmaterialkapazität 2 zur Fertigung bietet.
- 3) Möglichkeiten, die die Rohmaterialkapazität 3 zur Fertigung bietet.

(2) Bestimmen des Gewinns für die 25 zulässigen (Fertigungs-)Kombinationen.

Annahme: Gewinn für ein Stück des Produkts 1 ... 1 GE
 Gewinn für ein Stück des Produkts 2 ... 2 GE



(3) optimale Kombination

Optimale Lösung: 5 Stück Produkt 2
1 Stück Produkt 1
Gewinn 11 GE

Zusammenfassung: Ein Verfahren wurde gefunden, das jedes beliebige Zahlenbeispiel mit 2 zu fertigenden Produkten und beliebig vielen Rohmaterialien systematisch zu einer gewinnoptimalen Lösung führt, falls es eine solche Lösung gibt.

2.1.1.3 Von der Methode zum praktischen Verfahren

Unbefriedigend an dem bisher entwickelten, methodisch einwandfreien Verfahren ist:

1. Das Verfahren ist nicht mehr in dieser Form durchführbar, wenn mehr als 2 zu fertigende Produkte behandelt werden müssen.
2. Auch bei 2 Produkten kann die Prozedur, die nach den 3 Schritten abläuft, sehr langwierig sein, z.B. dann, wenn die Zahl der grundsätzlich möglichen Fertigungskombinationen sehr groß ist.

Wie muß ein Verfahren beschaffen sein, das diese beiden Schwachpunkte vermeidet, dennoch systematisch in jedem Fall eine optimale Dispositionsregel bestimmt?

Dazu sind 3 Fragen zu klären:

1. Wie stellt man das Problem formelmäßig dar?
(Voraussetzung für die Lösung durch einen formalen Rechenprozeß)
2. Gibt es einen Rechenprozeß, der immer, auch bei beliebig vielen Produkten, systematisch zu einer optimalen Lösung führt?
3. Gibt es einen solchen Rechenprozeß , der mit möglichst wenigen Rechenschritten auskommt?

Die Klärung der unter 2. und 3. ausgeführten Fragen erfordert umfangreiche mathematische Beweisverfahren.

Ergebnis dieser mathematischen Untersuchungen: Es gibt tatsächlich einen Rechenprozeß, der immer zum Ziel führt. Dazu brauchen nur die Punkte eines „Simplex“ für die Berechnung des Optimums herangezogen werden. Eine optimale Lösung liegt in dem Punkt eines „Simplex“. Das auf diese Weise angedeutete Rechenverfahren heißt **Simplexverfahren**.

Die Antwort auf Frage 1 zeigt die folgende mathematische Formulierung des Problems ((1) und (2)):

Mögliche Stückzahl des Produkts 1: x_1

Mögliche Stückzahl des Produkts 2: x_2

Allgemein gilt: $x_1 \geq 0$, $x_2 \geq 0$

Die grafische Darstellung läßt sich dann folgendermaßen beschreiben:

$$\begin{aligned}x_1 + x_2 &\leq 6 \\2x_1 + x_2 &\leq 10 \\x_1 + 3x_2 &\leq 16\end{aligned}$$

Gewinn: $Z = x_1 + 2x_2$

Wähle aus den möglichen Lösungen des Systems von Ungleichungen die heraus, die den Wert der Zielfunktion Z zu einem Maximum macht. Das Auswählen erfolgt nach einem speziellen Rechenformalismus, dem (bereits erwähnten) Simplex-Verfahren.

2.1.2 Beispiele für Optimierungsprobleme

2.1.2.1 Produktionsprobleme

1) vgl. Bsp. zu 2.1.1

2) Ein Unternehmen erstellt 2 Produkte P_1 und P_2 . Für einen Produktionszyklus sind folgende Bedingungen bekannt:

a) In das Produkt P_2 geht je Mengeneinheit (ME) ein Rohstoff R_1 mit 2 ME ein. Dieser Rohstoff wird als Abfallprodukt mit 1 ME je hergestellter ME von P_1 frei. Aus der vergangenen Produktionsperiode stehen noch 4 ME von R_1 zur Verfügung. Die Produktion von P_2 darf die gesamte verfügbare Menge von R_1 nicht überschreiten Falls x_1 die Anzahl der ME vom Produkt P_1 und x_2 die Anzahl der ME von Produkt P_2 ist, gilt:

$$2x_2 \leq 4 + x_1$$

b) Beide Produkte werden zwei verschiedenen Arbeitsgängen auf zwei verschiedenen Maschinen M_1 , M_2 unterworfen. Die zur Verfügung stehende Kapazität in Maschinenstunden (Mh) von M_1 beträgt 8 (Mh) und die von M_2 10 (Mh). Auf M_1 werden für die Erstellung einer ME von P_1 1 (Mh) und P_2 2 (Mh) gebraucht. Bei M_2 sind es 2 Mh für 1 ME von P_1 und 1 Mh für 1 ME von P_2 . Es gilt also:

$$\begin{aligned}x_1 + 2x_2 &\leq 8 \\2x_1 + x_2 &\leq 10\end{aligned}$$

c) Es geht ein weiterer Rohstoff R_2 in die Produktion ein, der im Verhältnis 1 : 3 in je eine ME der Produktion P_1 und P_2 eingeht. Es stehen 12 ME von P_2 zur Verfügung. Es gilt also: $x_1 + 3x_2 \leq 12$

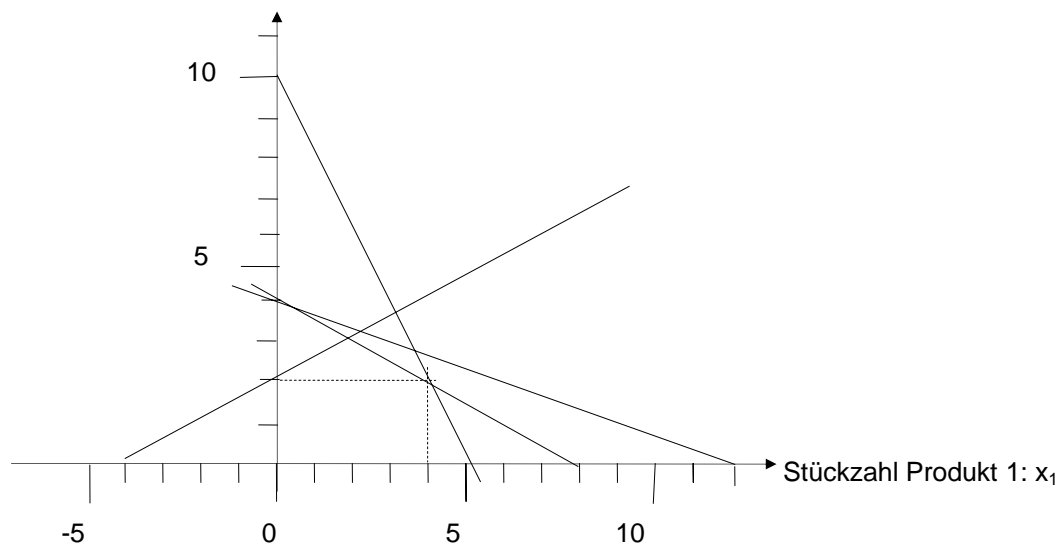
d) Das Ziel dieser Aufgabe ist die Maximierung des Gesamtdeckungsbeitrags. Es wurden die Deckungsbeiträge 3 GE (Geldeinheiten) für je eine ME von P_1 und 4 GE für je eine ME von P_2 ermittelt.

Die Zielfunktion lautet also: $Z_{\max} = 3x_1 + 4x_2$

Zusammengefaßt ergibt sich das folgende mathematische Modell:

- (1) $x_1 \geq 0, x_2 \geq 0$
- (2) $-x_1 + 2x_2 \leq 4$
 $x_1 + 2x_2 \leq 8$
 $2x_1 + x_2 \leq 10$
 $x_1 + 3x_2 \leq 12$
- (3) $Z_{\max} = 3x_1 + 4x_2$

Stückzahl Produkt 2: x_2



Optimale Lösung: $x_1 = 4, x_2 = 2, Z_{\max} = 20$

2.1.2.2 Mischungsprobleme

- Typische Aufgabenstellung

Für eine Schiffsbesatzung soll ein Vitaminpräparat aus den Substanzen S_1 und S_2 hergestellt werden. Der Gehalt an Vitaminen (in ME je g) dieser Substanzen, der Mindestbedarf an Vitaminen in dem herzustellenden Präparat sowie Kosten (in DM je g) sind in der folgenden Tabelle zusammengestellt.

	S_1	S_2	Mindestbedarf
Vitamin A	2	1	16
Vitamin B	1	0	2
Vitamin C	2	3	32
Vitamin D	2	5	40
Kosten	10	8	

Das Präparat soll durch Mischung von S_1 und S_2 so hergestellt werden, daß die vorgegebenen Mindestmengen darin enthalten sind und die Gesamtkosten für die Herstellung minimal werden.

- Mathematische Formulierung des Problems

x_1 ist die Menge in g, die von der Substanz S_1 und x_2 ist die Menge in g, die von der Substanz S_2 benötigt wird.

(1) Nichtnegativitätsbedingungen

$$x_1 \geq 0, \quad x_2 \geq 0$$

(2) Einschränkende Bedingungen

$$2x_1 + x_2 \geq 16$$

$$x_1 \geq 2$$

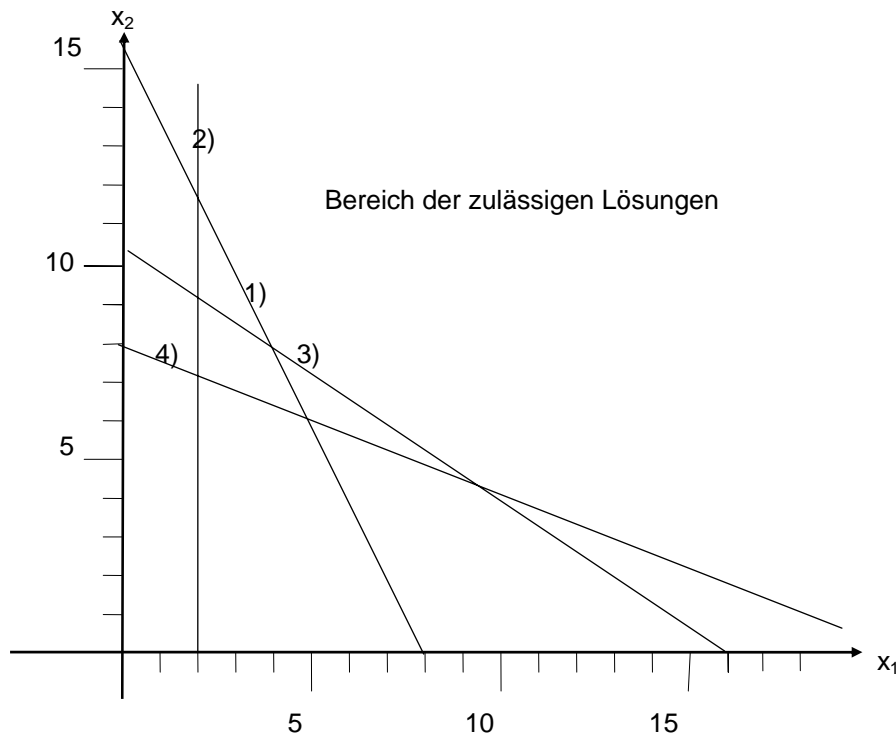
$$2x_1 + 3x_2 \geq 32$$

$$2x_1 + 5x_2 \geq 40$$

(3) Zielfunktion

$$Z_{\min} = 10x_1 + 8x_2$$

- Grafische Darstellung des Problems



1. Der Graf der Zielfunktion stellt hier eine parallele Schar von Geraden (Anstieg $-5/4$) dar.
2. Es wird ein Punkt der Lösungsmenge gesucht, für den der Wert der Zielfunktion am kleinsten ist.

Optimale Lösung

$$\begin{aligned} x_1 &= 4 \\ x_2 &= 8 \\ Z_{\min} &= 104 \end{aligned}$$

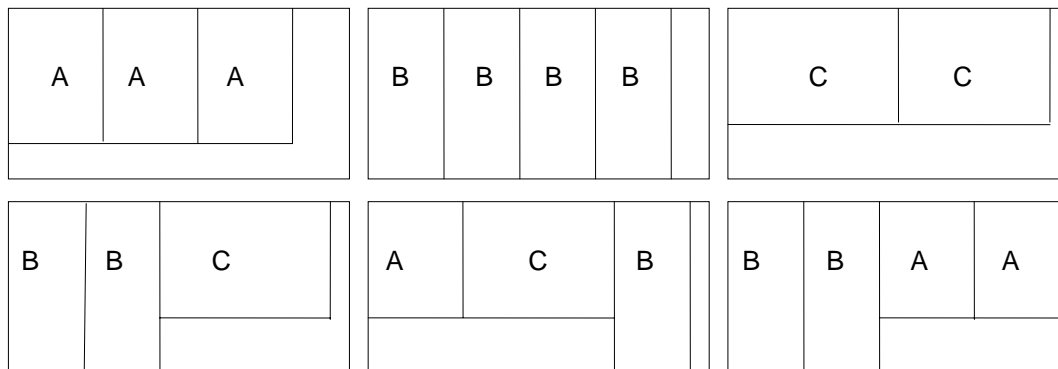
2.1.2.3 Verschnittprobleme

- Typische Aufgabenstellung

Eine Möbelfirma benötigt mindestens:

- 200 Kunststoffplatten der Größe 150 cm * 150 cm (Format A)
- 300 Kunststoffplatten der Größe 200 cm * 110 cm (Format B)
- 100 Kunststoffplatten der Größe 250 cm * 150 cm (Format C)

Der Lieferant stellt nur Platten in der Einheitsgröße 200 cm * 520 cm her. Deshalb muß die Möbelfirma die Platten selbst nachschneiden. Dieses Zuschneiden kann nach folgenden Plänen erfolgen:



Der Zuschnitt nach diesen Schnittplänen soll so erfolgen, daß möglichst wenig Abfall entsteht, d.h. etwas vereinfacht: Die Anzahl der zu bestellenden Platten soll minimal sein.

- Mathematische Formulierung des Problems

$x_1, x_2, x_3, x_4, x_5, x_6$ sind die Anzahl der Platten, die nach den Schnittplänen 1, 2, 3, 4, 5, 6 notwendig sind.

(1) Nichtnegativitätsbedingung

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0$$

(2) Einschränkende Bedingungen

$$\begin{array}{rcl} 3x_1 & & + x_5 + 2x_6 \geq 200 \\ 4x_2 + & 2x_4 + x_5 + 2x_6 & \geq 300 \\ 2x_3 + & x_4 + x_5 & \geq 100 \end{array}$$

(3) Zielfunktion

$$Z_{\min} = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

2.1.2.4 Verteilungsprobleme

- Typische Aufgabenstellung

Für die Realisierung eines Produktionsprogramms stehen mehrere Maschinen zur Verfügung, die gegeneinander austauschbar sind. Die einzelnen Maschinen stellen folgende Produktionskapazitäten bereit:

```

Maschine M1 ..... 180 Minuten (min)
Maschine M2 ..... 100      "   ( " )
Maschine M3 ..... 150      "   ( " )
Maschine M4 ..... 100      "   ( " )

```

Diese Kapazitäten sollen für die Herstellung folgender Erzeugnisse in Anspruch genommen werden:

Erzeugnis E₁ mit 30 Stück

E₁ kann auf M₁ in 7 min, auf M₂ in 5 min, auf M₃ in 4 min oder auf M₄ in 4 min hergestellt werden

Erzeugnis E₂ mit 30 Stück

E₂ kann auf M₁ in 4 min oder auf M₂ in 3 min hergestellt werden

Erzeugnis E₃ mit 50 Stück

E₃ kann auf M₂ in 2 min, auf M₃ in 4 min oder auf M₄ in 5 min hergestellt werden

Erzeugnis E₄ mit 40 Stück

E₄ kann auf M₁ in 6 min, auf M₂ in 5 min oder auf M₄ in 3 min hergestellt werden

E₂ kann also nicht auf M₃ und M₄ hergestellt werden, E₃ nicht auf M₁ und E₄ nicht auf M₃ hergestellt werden.

Mit Hilfe dieser Angaben ist das mathematische Modell für den optimalen Maschinen-Belegungsplan zu ermitteln!

- Mathematische Formulierung des Problems

x_{ij} : Stückzahl des Erzeugnisses E_i, die auf der Maschine M_j hergestellt werden kann

(1) Nichtnegativitätsbedingungen

$x_{ij} \geq 0$ für $1 \leq i \leq 4$ und $1 \leq j \leq 4$

(2) Einschränkende Bedingungen

Erzeugnis	M ₁	M ₂	M ₃	M ₄	Stückzahl
E ₁	7	5	4	4	30
E ₂	4	3	-	-	30
E ₃	-	2	4	5	50
E ₄	6	5	-	3	40
Freie Produktkapazität	180	100	150	100	

$$x_{11} + x_{12} + x_{13} + x_{14} = 30$$

$$x_{21} + x_{22} = 30$$

$$x_{32} + x_{33} + x_{34} = 50$$

$$x_{41} + x_{42} + x_{44} = 40$$

$$7x_{11} + 4x_{21} + 6x_{41} \leq 180$$

$$5x_{12} + 3x_{22} + 2x_{32} + 5x_{42} \leq 100$$

$$\begin{array}{rcl}
4x_{12} + & 4x_{33} & \leq 150 \\
4x_{14} + & 5x_{34} + 3x_{44} & \leq 100
\end{array}$$

(3) Zielfunktion

(Die Gesamtbearbeitungszeit ist zu minimieren)

$$\begin{aligned}
Z_{\min} = & 7x_{11} + 4x_{21} + 6x_{41} + 5x_{12} + 3x_{22} + 2x_{32} + 5x_{42} + 4x_{13} \\
& + 4x_{33} + 4x_{14} + 5x_{34} + 3x_{44}
\end{aligned}$$

2.1.2.5 Einsatzmöglichkeiten von OR-Modellen im DV-Management

1) Optimale Arbeitslastverteilung⁵

In Computer-Verbundsystemen interessiert die optimale Arbeitslastverteilung. Verbundsysteme werden zur gemeinsamen Nutzung von installierter Hardware, bestimmter Anwendungssoftware oder großen Datenbeständen aufgebaut. Zur Formulierung als allgemeines lineares Optimierungsproblem liegen folgende Angaben vor:

In einem Netzwerk mit den Rechnern R_i sollen die Aufträge A_j bearbeitet werden. Es bedeuten

d_{ij} ... Dauer der Bearbeitung der Aufträge A_j auf dem Rechner R_i

r_i ... Kapazität des Rechners

c_i ... Kosten je Einheit für die Bearbeitung auf Rechner R_i

c_{ij} ... Kosten der Übertragung des Auftrags A_j von „seinem Rechner“ R_j zum Rechner R_i (und zurück)

Stelle mit Hilfe dieser Angaben das mathematische Modell für das Optimierungsproblem auf.

1, falls Auftrag A_j auf Rechner R_i bearbeitet wird

$$(1) \ x_{ij} = \{$$

0, andernfalls

$$(2) \ \sum_{j=1}^n d_{ij} \cdot x_{ij} \leq r_i \quad \text{für alle } i$$

$$\sum_{j=1}^n x_{ij} = 1 \quad \text{für alle } j \text{ (Kein Auftrag darf unbearbeitet bleiben)}$$

$$(3) \ Z_{\min} = \sum_{i=1}^m \sum_{j=1}^n (c_i \cdot d_{ij} + c_{ij}) \cdot x_{ij}$$

2) Benchmark-Tests mit möglichst vielen Programmläufen

Benchmark-Tests sind Versuchsläufe mit repräsentativen Programmzusammenstellungen, Datenprofilen und Lastverteilungen zur Leistungsbeurteilung von Hardwarekonfigurationen. Um zu erreichen, daß ein solcher Test aus möglichst vielen Programmläufen besteht, ist ein „lineares Optimierungsmodell“ angebracht. Formuliere dafür das „mathematische Modell“. Berücksichtige dabei:

⁵ vgl.: Stahlknecht, Peter: Einsatzmöglichkeiten von quantitativen Verfahren und OR-Modellen für das DV-Management, OR-Spektrum (1981) 3:65-90

- Für den Test sollen $j = 1, 2, \dots, n$ Programme in Betracht kommen.
- Eingesetzt werden $i = 1, \dots, n$ Systemkomponenten (CPU, Drucker, Bänder, etc.).

Es bedeuten

- a_{ij} : Inanspruchnahme der Komponente i durch Programm j
- b_i : im Test gewünschte höchste Inanspruchnahme der Komponente i

Mathematisches Modell: x_j ist die Anzahl Läufe des Programms j

(1) $x_j \geq 0$ und ganzzahlig

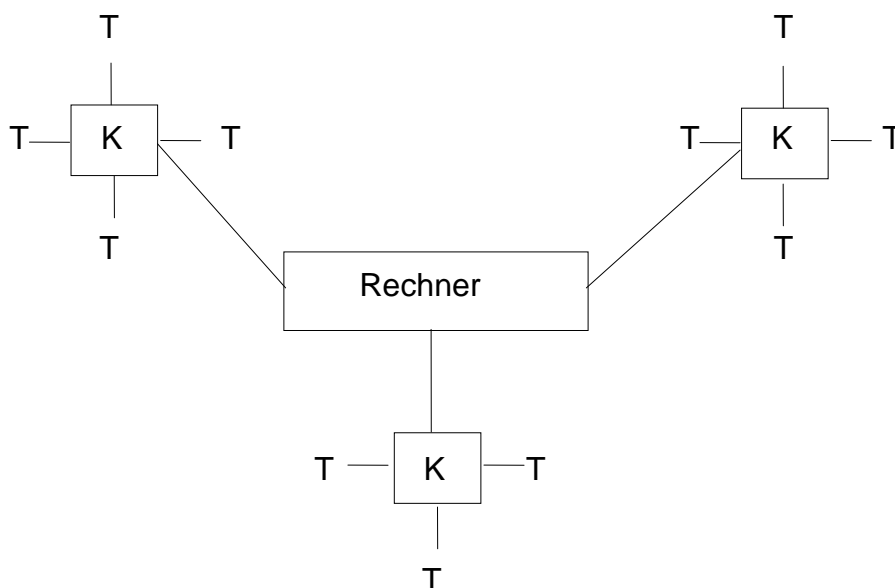
(2) $\sum_{j=1}^n a_{ij} \cdot x_j \leq b_i$ für alle i

(3) $Z_{\max} = \sum_{j=1}^n x_j$

3) Optimale Netzkonfiguration

„m“ Datenstationen (Terminals) T_i sind über „n“ Konzentratoren K_j an einen zentralen Rechner anzuschließen:

- Die Anzahl der Anschlüsse an einen Konzentrador ist beschränkt
- Jede Datenstation ist genau an einen Konzentrador angeschlossen



Jeder Konzentrador hat dabei die Aufgabe, die Nachrichten von mehreren Datenstationen zu sammeln und auf einer Leitung an den zentralen Rechner zu übertragen. Die Weitergabe der Daten auf einer einzigen, dafür schnellen Leitung erspart Übertragungskosten und Anschlüsse in der dem zentralen Rechner vorgeschalteten Steuereinheit für Datenfernverarbeitung. Der Konzentrador muß außerdem Daten zwischenspeichern und Warteschlangen verwalten.

Wie viele Konzentratoren sind aufzustellen, damit die gesamten Übertragungskosten minimiert werden?

Vereinbarungen:

c_{ij} ... Kosten der Verbindung von T_i und K_j
 c_j ... Kosten der Verbindung von K_j nach dem Rechner

Mathematisches Modell:

- (1) $x_{ij} = \begin{cases} 1, & \text{falls } T_i \text{ mit } K_j \text{ verbunden ist} \\ 0, & \text{andernfalls} \end{cases}$
- (2) $\sum_{j=1}^m x_{ij} = 1$ (Jede Datenstation ist an einen Konzentrador angeschlossen)
- $\sum_{i=1}^m x_{ij} = b_j$ (Die Anzahl der Anschlüsse an einen Konzentrador ist beschränkt)
- (3) $Z_{\min} = \sum_{j=1}^n (c_j + \sum_{i=1}^m c_{ij} \cdot x_{ij})$

2.1.3 Nichtlineare und lineare Optimierungsprobleme

- Allgemeine Formulierung des Problems

(1) Nichtnegativitätsbedingung

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, \dots, x_n \geq 0$$

(2) Einschränkende Bedingungen

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &\leq b_1 \\ f(x_1, x_2, \dots, x_n) &\leq b_2 \\ &\dots \\ f(x_1, x_2, \dots, x_n) &\leq b_m \end{aligned}$$

(3) Zielfunktion

Nichtnegativitätsbedingungen und einschränkende Bedingungen bestimmen den Definitionsbereich für die Zielfunktion

$$Z_{\max} = f(x_1, x_2, \dots, x_n)$$

Jedes n -Tupel (x_1, \dots, x_n) , das (1) und (2) erfüllt, heißt zulässige Lösung. Eine zulässige Lösung heißt optimale Lösung, wenn sie das Maximum der Zielfunktion liefert.

- Bemerkungen

1. Soll bei einem Optimierungsproblem nicht das Maximum sondern das Minimum einer Zielfunktion Z bestimmt werden, so kann zu Z durch Multiplikation mit -1 eine neue Zielfunktion angegeben werden, von der dann das Maximum zu berechnen ist.

2. Haben einzelne oder alle Ungleichungen der einschränkenden Bedingungen statt " \leq " " \geq ", so können diese Ungleichungen mit -1 multipliziert werden, damit sie angegebene Richtung erhalten.
3. Wird bei einigen oder allen einschränkenden Bedingungen nur das Gleichheitszeichen zugelassen, so können für eine Gleichung jeweils 2 Ungleichungen angegeben werden:

$f_k(x_1, x_2, \dots, x_n) = b_k$ ist äquivalent zu den beiden Ungleichungen;

$f_k(x_1, x_2, \dots, x_n) \geq b_k$ und $f_k(x_1, x_2, \dots, x_n) \leq b_k$

4. Ist die Nichtnegativitätsbedingung für einzelne oder alle Variable nicht erfüllt, dann kann durch Einführung zweier neuer Variablen diese Bedingung stets erreicht werden.

Gilt z.B.: $x_k > 0$, $x_k = 0$, $x_k < 0$, so erhält man durch Substitution:

$$x_k = x_{k1} - x_{k2} \text{ mit } x_{k1} \geq 0 \text{ und } x_{k2} \geq 0$$

mit der Voraussetzung:

x_{k1} ist eine nichtnegative Variable

x_{k2} " " " "

die gewünschte Form der Nichtnegativitätsbedingungen. Es ergibt sich:

Falls $x_{k1} > x_{k2}$, dann ist $x_k > 0$

Falls $x_{k1} = x_{k2}$, dann ist $x_k = 0$

Falls $x_{k1} < x_{k2}$, dann ist $x_k < 0$

Lineare Optimierung

Alle einschränkenden Bedingungen und die Zielfunktion werden durch lineare Beziehungen beschrieben:

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0, \quad \dots, \quad x_n \geq 0$$

$$(2) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \end{aligned}$$

$$(3) \quad Z_{\max} = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Nichtlineare Optimierungsprobleme

Alle Optimierungsprobleme, die nicht durch das vorstehende mathematische Modell beschrieben werden.

Normalformen der linearen Optimierung

In der Praxis werden zwei Aufgabentypen der linearen Optimierung unterschieden. Sie werden als **Normalformen** bezeichnet.

Normalform der Maximumaufgabe

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0, \quad \dots, \quad x_n \geq 0$$

$$(2) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \end{aligned}$$

mit $b_1 \geq 0, b_2 \geq 0, \dots, b_m \geq 0$

$$(3) \quad Z_{\max} = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Normalform der Minimumaufgabe

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0, \quad \dots, \quad x_n \geq 0$$

$$(2) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\geq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\geq b_2 \\ &\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\geq b_m \end{aligned}$$

mit $b_1 \geq 0, b_2 \geq 0, \dots, b_m \geq 0$

$$(3) \quad Z_{\min} = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

2.1.4 Nichtlineare, lineare Optimierungsprobleme und Extremalwertprobleme

2.1.4.1 Einführende Beispiele

Klassische Extremalwertprobleme werden mit Hilfe der Methoden der Differentialrechnung gelöst.

2.1.4.1.1 Optimale Lagerauffüllung vor Preiserhöhungen

Ein Betrieb verbraucht von einem Rohprodukt je Woche $m = 10.000 \text{ kg}$. Die Menge wird wöchentlich von einem Lieferanten bezogen. Der bisherige Preis betrug $s_0 = 5.- \text{ DM/kg}$ und soll auf $s_1 = 5,50.- \text{ DM/kg}$ erhöht werden. Welche Menge soll der Betrieb noch zum alten Preis einkaufen, um von der Preiserhöhung möglichst wenig betroffen zu sein?

- Wird nur eine kleine Menge zum alten Preis eingekauft, so ist der Vorteil gering
- Wird dagegen eine große Menge gekauft, so geht die durch die Preisdifferenz erzielte Einsparung durch Zinsen für das im Lager gebundene Kapital verloren (Zinssatz $p = 1/4\%$ pro Woche, das entspricht 13% im Jahr).

Mathematischer Ansatz

x ist die zu bestimmende Einkaufsmenge. Dann beträgt die Einsparung: $(s_1 - s_0) \cdot x$.
 x deckt den Bedarf für x/m Wochen. Der durchschnittliche Lagerbestand ist zu

verzinsen: $\frac{x}{m} \cdot \frac{x}{2} \cdot s_0 \cdot \frac{p}{100}$.

Der Gewinn G ist dann: $G = (s_1 - s_0) \cdot x - \frac{s_0 \cdot p}{200 \cdot m} \cdot x^2$

Differenzieren nach x und Nullsetzen der 1. Ableitung führt zur optimalen Menge x_{opt} :

$$\frac{dG}{dx} = (s_1 - s_0) - \frac{s_0 \cdot p}{100 \cdot m} \cdot x$$

$$x_{\text{opt}} = \frac{100 \cdot m}{p} \cdot (s_1 - s_0) = \frac{100 \cdot 10000}{0.25} \cdot (1.1 - 1) = 400000$$

Diese Menge von 400.000 kg deckt einen Bedarf von 40 Wochen. Die Einsparung durch die Preisdifferenz beträgt 200.000.- DM. Davon gehen Zinsen in Höhe von 100.000.- DM ab, so daß ein Nettovorteil von 100.000.- DM bleibt.

Nachweis des Extremums: $\frac{d^2G}{dx^2} = -\frac{s_0 \cdot p}{100 \cdot m}$

Die Lösung ist tatsächlich ein Maximum.

2.1.4.1.2 Die klassische Bestellmengenformel (Losgrößenformel)

Ein Betrieb verbraucht regelmäßig ein bestimmtes Rohprodukt mit der gleichmäßig über das Jahr verteilten Bedarfsmenge $m = 10.000$ Stück. Ein Produkt kostet $s = 10$.- DM/Stück. Für jede Bestellung und die daran anschließende Lieferung und Einlagerung entstehen die (bestellfixen) Kosten $E = 60$.- DM. Wegen dieser Bestell- und Beschaffungskosten werden möglichst selten Bestellungen mit entsprechend großer Menge angestrebt. Jedoch entstehen Lagerkosten und Zinsen für das im Lager gebundene Kapital vom $p = 12$ % p.a., so daß eine möglichst kleine Lagermenge und damit häufig Bestellungen vorteilhaft erscheinen.

Gesucht ist diejenige Bestellung x , für die die Summe beider gegenläufigen Kosten minimal ist.

Mathematischer Ansatz

x ist die jeweilige Bestellmenge, dann sind die jährlichen Bestellkosten: $E \cdot \frac{m}{x}$ und die Zins- und Lagerkosten unter der Annahme eines durchschnittlichen Lagerbestands von $x/2$ (d.h. durchschnittlich ist im Lager ein Kapital von $\frac{x \cdot s}{2}$ gebunden): $\frac{p \cdot s}{200} \cdot x$

Insgesamt betragen die jährlichen Kosten: $K = m \cdot s + E \cdot \frac{m}{x} + \frac{p \cdot s}{200} \cdot x$

Die Kosten sind zu minimieren: $\frac{dK}{dx} = -E \cdot \frac{m}{x^2} + \frac{p \cdot s}{200}$

$$x_{opt} = \sqrt{\frac{200 \cdot E \cdot m}{p \cdot s}} = \sqrt{\frac{200 \cdot 10000 \cdot 60}{12 \cdot 10}} = 1000$$

Optimale Bestellhäufigkeit

$$y = \frac{m}{x}$$

$$K = m \cdot s + E \cdot y + \frac{p \cdot s}{200} \cdot \frac{m}{y}$$

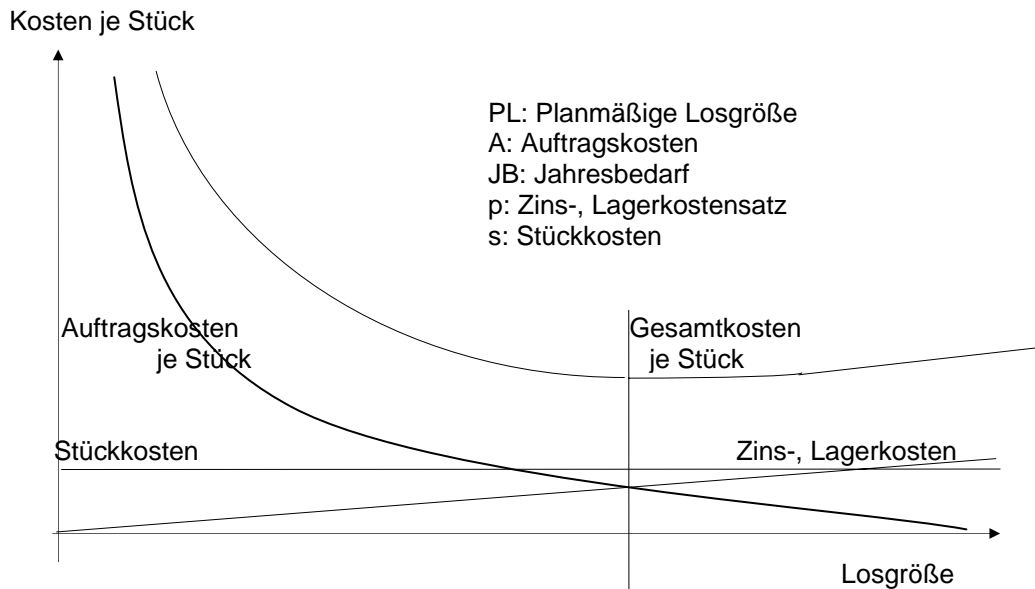
$$\frac{dK}{dy} = E - \frac{p \cdot s \cdot m}{200 \cdot y^2}$$

$$y_{opt} = \sqrt{\frac{m \cdot p \cdot s}{200 \cdot E}} = \sqrt{\frac{10000 \cdot 12 \cdot 10}{200 \cdot 60}} = 10$$

Andler'sche Formel

Die gleichen Formeln, die zur Berechnung optimaler Bestellmengen entwickelt wurden, lassen sich auch für die Bestimmung optimaler Los- oder Seriengrößen in der Serien- und Sortenfertigung anwenden.

Bsp.: Andler'sche Formel $PL = \sqrt{\frac{200 \cdot A \cdot JB}{p \cdot s}}$



2.1.4.2 Lagerhaltungsprobleme

Aufgabe: Vorausbestimmen des Zeitpunkts und des Umfangs von Bestellungen zur optimalen Lagerhaltung

Hierzu sind optimale Entscheidungen so zu bestimmen, daß die mit der Lagerhaltung verbundenen Kosten minimiert werden:

1. Lagerkosten (Unterhaltungskosten des Lagers, Umlaufmittelbindungskosten)
2. Beschaffungskosten (Bestell-) (administrative Kosten der Bestellung)
3. Fehlmengenkosten (Kosten und Verluste, die durch Fehlmengen (unbefriedigeter Bedarf) am Lager verursacht werden)

Optimale Entscheidungen können bzgl. der Zeit oder der Menge getroffen werden (Wann soll wieviel bestellt werden?).

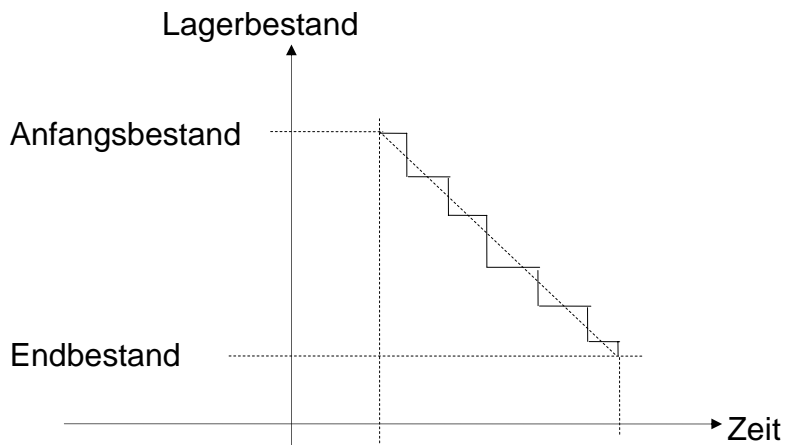
Dabei sind verschiedene Kombinationen der folgenden Bestellregeln denkbar:

1. Das Lager wird nur aufgefüllt, wenn der Lagebestand kleiner als s (ME) ist
2. Das Lager wird alle T Zeiteinheiten aufgefüllt
3. Die Bestellmenge ist q (ME)
4. Die Bestellmenge soll so groß sein, daß der Lagerbestand auf S (ME) gebracht wird.

Das Lagerhaltungsproblem bei dem q und T gesucht werden, heißt Lagerhaltungsproblem mit einer (q,T) -Bestellregel (bzw. (T,q) -Bestellregel).

Grundlegende Lagerhaltungsmodelle

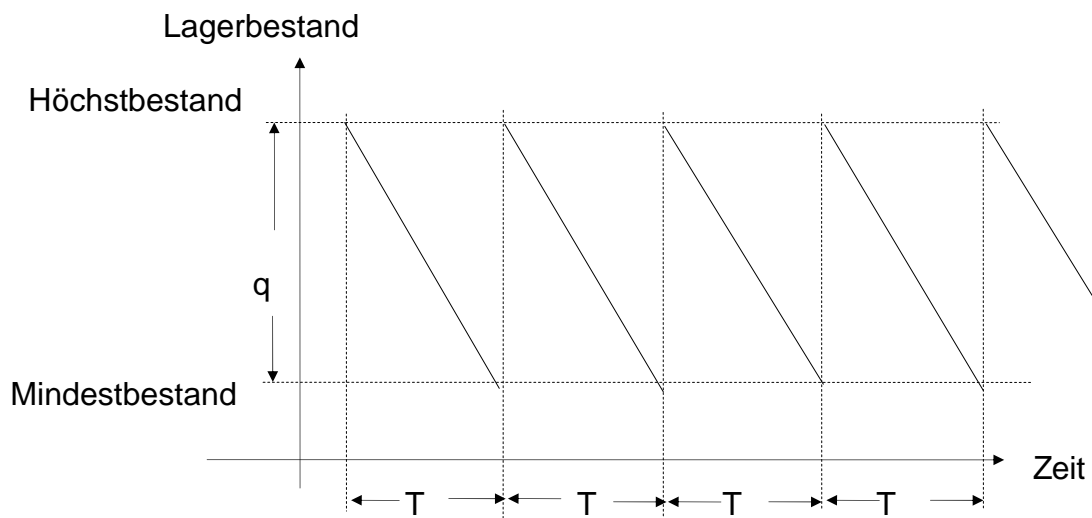
Allen Lagerhaltungsmodellen liegt folgender Verlauf der Bestands während einer Periode T zugrunde:



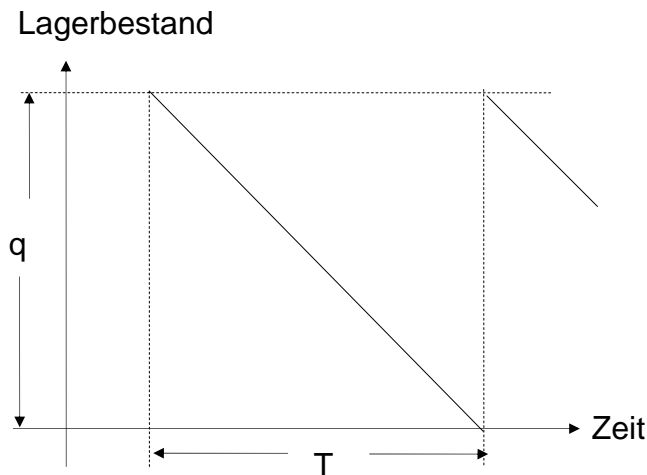
1. Planung mit fester Periode

a) Bedarf und Bestellmenge sind konstant. Die Bestellung läuft ohne Verzögerung.

1) Fehlmengen sind nicht erlaubt.



Problembeschreibung: Betrachtet wird das folgende deterministische Lagerhaltungsmodell



Der Bedarf ist bekannt und zeitlich (z.B. jährlich) konstant. Er soll immer voll befriedigt werden. Bestellungen können jederzeit in beliebiger Höhe getätigt werden. Die Beschaffungszeit ist beliebig konstant und hat keinen Einfluß auf das Modell. Bekannt ist außerdem:

c_B (GE): Bestellkosten je Bestellung

c_L (GE/ME u. ZE): Lagerkosten.

Gesucht ist die optimale (q,T) -Bestellregel, so daß minimale Gesamtkosten entstehen!

Optimale Bestellmenge

$$K = c_B \cdot \frac{b}{q} + \frac{1}{2} \cdot c_L \cdot q$$

$$\frac{dK}{dq} = -c_B \cdot \frac{b}{q^2} + \frac{1}{2} \cdot c_L = 0$$

$$q^2 = \frac{2 \cdot c_B \cdot b}{c_L}$$

$$q = \sqrt{\frac{2 \cdot c_B \cdot b}{c_L}}$$

optimaler Bestellrhythmus

$$q = b \cdot T$$

$$K = \frac{c_B}{T} + \frac{1}{2} \cdot c_L \cdot b \cdot T$$

$$\frac{dK}{dT} = -\frac{c_B}{T^2} + \frac{1}{2} \cdot c_L \cdot b$$

$$T^2 = \frac{2 \cdot c_B}{c_L \cdot b}$$

$$T = \sqrt{\frac{2 \cdot c_B}{c_L \cdot b}}$$

Aufgabe: Ein Erzeuger hat seinen Kunden mit 24.000 (ME) eines Erzeugnisses je Jahr zu beliefern. Der Bedarf ist fest und bekannt. Da dieses Erzeugnis vom Abnehmer in einem Fließbandprozeß verwertet wird, hat er keinen Lagerraum für dieses Erzeugnis vorgesehen. Der Produzent muß bedarfsgerecht liefern. Sollte der Erzeuger die verlangten Mengen nicht liefern, wird er den Auftrag und wahrscheinlich auch sein Geschäft verlieren. Folglich werden keine Fehlmengen zugelassen. Die Lagerkosten betragen 0.10 GE je Stück und Monat, die Vorbereitungskosten je Produktionsserie 350 GE.

Bestimme die optimale Seriengröße, das zugehörige optimale Intervall und das Minimum der Gesamtkosten.

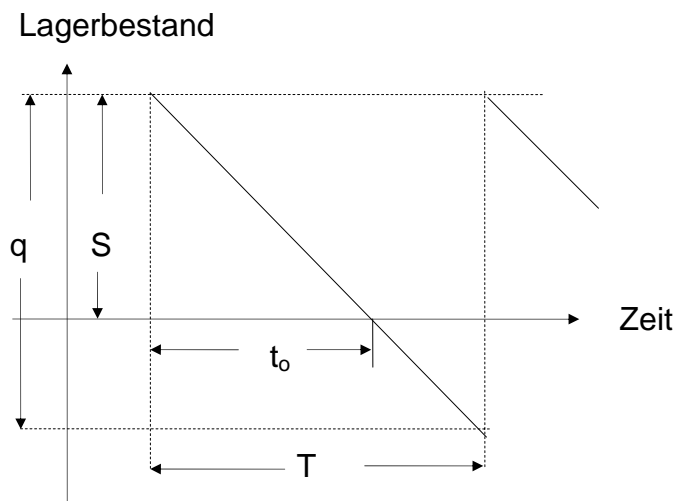
q

T

Minimum der Gesamtkosten: $\sqrt{2 \cdot c_B \cdot c_L \cdot b}$

2) Fehlmengen sind erlaubt

Problembeschreibung: Betrachtet wird das folgende deterministische Lagerhaltungsmodell



Der Fehlmengenkostensatz c_F (GE/ME u. ZE) ist bekannt. Gesucht ist der optimale Bestellrhythmus und der optimale Bestand ((T,S)-Bestellregel), so daß minimale Gesamtkosten entstehen.

Bestellkosten: c_B

Lagerkosten: Da $S = b \cdot t_0$ ist, gilt $\frac{c_L \cdot S \cdot t_0}{2} = \frac{c_L \cdot S^2}{2 \cdot b}$

Fehlmengenkosten: Da $q = b \cdot T$ ist, gilt $c_F \cdot (q - S) \cdot \frac{T - t_0}{2} = (b \cdot T - S) \cdot (T - \frac{S}{b}) \cdot \frac{c_F}{2}$

Bezieht man die Kosten auf eine Zeiteinheit (ZE), so ergeben sich die Lagerkosten je ZE:

$$K = \frac{c_B}{T} + \frac{c_L \cdot S^2}{2 \cdot b \cdot T} + c_F \cdot \frac{b \cdot T - S}{T} \cdot \frac{b \cdot T - S}{2 \cdot b} = \frac{c_B}{T} + \frac{c_L + c_F}{2 \cdot b \cdot T} \cdot S^2 - c_F \cdot S + \frac{c_F \cdot b}{2}$$

$$\frac{\partial K}{\partial T} = -\frac{c_B}{T^2} - \frac{c_L + c_F}{2 \cdot b \cdot T^2} \cdot S^2 + \frac{c_F \cdot b}{2}$$

$$\frac{\partial K}{\partial S} = \frac{c_L + c_F}{b \cdot T} \cdot S - c_F$$

$$S = \frac{c_F \cdot b \cdot T}{c_L + c_F}$$

$$\frac{c_B}{T^2} + \frac{c_L + c_F}{2 \cdot b \cdot T^2} \cdot S^2 = \frac{c_F \cdot b}{2}$$

$$\frac{c_B}{T^2} + \frac{(c_F \cdot b \cdot T)^2}{(c_L + c_F) \cdot 2 \cdot b \cdot T^2} = \frac{c_F \cdot b}{2}$$

$$\frac{c_B}{T^2} = \frac{c_F \cdot b}{2} - \frac{c_F^2 \cdot b}{2 \cdot (c_L + c_F)}$$

$$\frac{c_B}{T^2} = \frac{c_F \cdot b}{2} \cdot \left(1 - \frac{c_F}{c_L + c_F}\right)$$

$$\frac{T^2}{c_B} = \frac{2 \cdot (c_L + c_F)}{c_L \cdot c_F \cdot b}$$

$$T = \sqrt{\frac{2 \cdot c_B \cdot (c_L + c_F)}{c_L \cdot c_F \cdot b}} = \sqrt{\frac{2 \cdot c_B}{b \cdot c_L}} \cdot \sqrt{\frac{c_L + c_F}{c_F}}$$

$$S = \frac{c_F}{c_L + c_F} \cdot \sqrt{\frac{c_L + c_F}{c_F}} \cdot \sqrt{\frac{2 \cdot c_B \cdot b}{c_L}} = \sqrt{\frac{2 \cdot c_B \cdot b}{c_L}} \cdot \sqrt{\frac{c_F}{c_L + c_F}}$$

$$\text{optimale Bestellmenge: } q = b \cdot T = \sqrt{\frac{c_L + c_F}{c_F}} \cdot \sqrt{\frac{2 \cdot c_B \cdot b}{c_L}}$$

Aufgabe: Ein Erzeuger hat seinen Kunden mit 24.000 (ME) eines Erzeugnisses je Jahr zu beliefern. Der Bedarf ist fest und bekannt. Da das Erzeugnis direkt vom Abnehmer verwertet wird, hat der Produzent bedarfsgerecht anzuliefern. Sollte dies ihm nicht gelingen, so wird er mit Fehlmengenkosten in der Höhe von 0.20 (GE je ME und Monat) belastet. Seine Lagerhaltungskosten betragen 0.10 (GE je ME und Monat), die Verarbeitungskosten je Produktionsserie 350 (GE).

a) Bestimme die optimale Seriengröße, das zugehörige optimale Zeitintervall und das Minimum der Gesamtkosten.

q

S

$$\text{Minimum der Gesamtkosten: } \sqrt{\frac{c_F}{c_F + c_L}} \cdot \sqrt{2 \cdot c_B \cdot c_L \cdot b}$$

b) Wie groß werden die voraussichtlichen Fehlmengen am Ende jeder Planungsperiode (optimales Zeitintervall) sein?

Fehlmenge:

3) Der Bedarf ist stochastisch und diskret

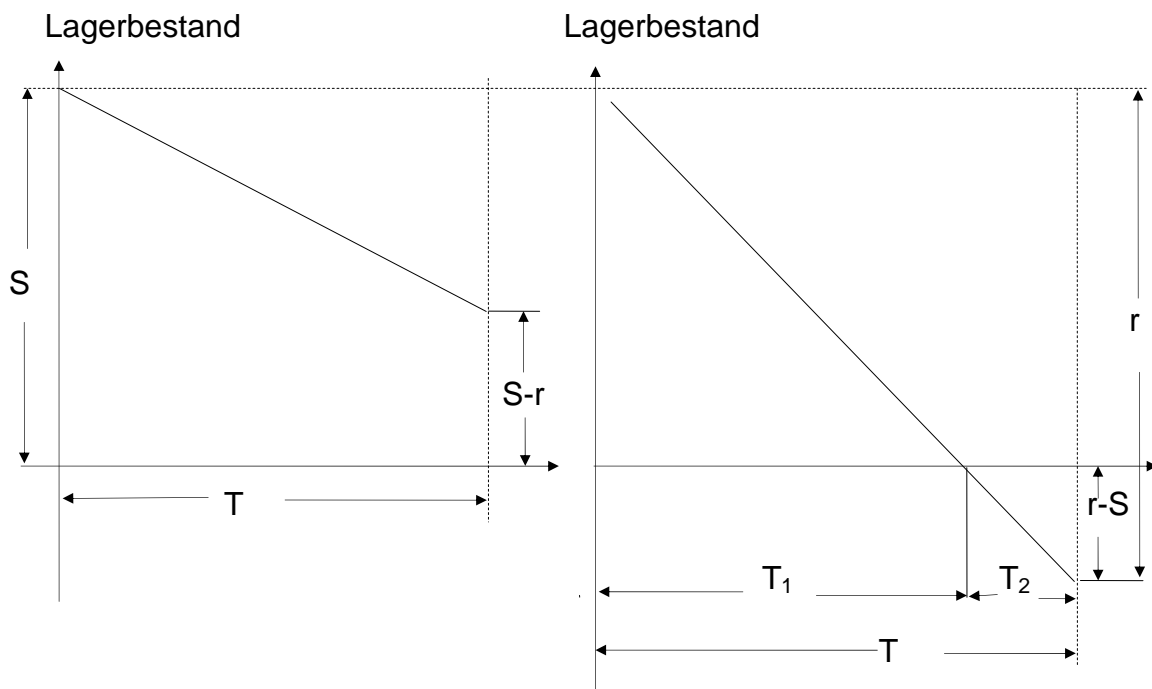
Der Bedarf ist eine Zufallsgröße. Statistische Analysen haben jedoch eine Wahrscheinlichkeitsverteilung des Bedarfs ermittelt.

Bsp.: Der Bedarf je Periode T ist eine Zufallsvariable und $p(r)$ ist die Wahrscheinlichkeit für den Bedarf r . Der Lagerbestand wird am Ende jeder Periode auf ein vorgegebenes Niveau S erneuert.

Gesucht wird jener Wert S_0 von S , der die mittleren Kosten der Planung minimiert. Gegeben ist:

- c_L ...Lagerkostensatz (GE/ME u. ZE)
- c_F ...Fehlmengekostensatz

Kosten der Beschaffung bzw. Bestellung bleiben hier unberücksichtigt.



Lösung

$$\text{Bedarf} < \text{Anfangsbestand: } K_1 = T \cdot c_L \cdot \left(S - \frac{r}{2}\right)$$

$$\text{Bedarf} > \text{Lagervorrat: } K_2 = T_1 \cdot c_L \cdot \frac{S}{2} + T_2 \cdot c_F \cdot \frac{r-S}{2}$$

$$T_1 = \frac{S}{r} \cdot T$$

$$T_2 = \frac{r-S}{r} \cdot T$$

$$K_2 = T \cdot \frac{s^2}{2 \cdot r} \cdot c_L + T \cdot \frac{(r-S)^2}{2 \cdot r} \cdot c_F$$

Erwartungswert der Lagerkosten für eine Periode

$$K(S) = T \cdot c_L \cdot \sum_{r=0}^S (S - \frac{r}{2}) \cdot p(r) + T \cdot c_L \cdot \sum_{r=S+1}^{\infty} \frac{S^2}{2 \cdot r} \cdot p(r) + T \cdot c_F \cdot \sum_{r=S+1}^{\infty} \frac{(r-S)^2}{2 \cdot r} \cdot p(r)$$

Gibt es einen Wert S_0 für den Lagerbestand S , so daß $K(S_0 - 1) > K(S_0)$ und $K(S_0 + 1) > K(S_0)$ ist, dann ist S_0 der optimale Lagerbestand.

Folgende Beziehung ist gegeben:

$$K(S+1) = K(S) + T \cdot (c_F + c_L) \cdot (\sum_{r=0}^S p(r) + (S + \frac{1}{2}) \cdot \sum_{r=S+1}^{\infty} \frac{p(r)}{r}) - T \cdot c_F$$

$$\text{Abkürzung: } L(S) = \sum_{r=0}^S p(r) + (S + \frac{1}{2}) \cdot \sum_{r=S+1}^{\infty} \frac{p(r)}{r}$$

Nachweis der Richtigkeit:

$$T \cdot c_L \cdot \sum_{r=0}^S (S - \frac{r}{2}) \cdot p(r) = T \cdot c_L \sum_{r=0}^S S \cdot p(r) - T \cdot c_L \sum_{r=0}^S \frac{r}{2} \cdot p(r)$$

Die Summe $T \cdot c_L \cdot \sum_{r=0}^S S \cdot p(r)$ ist bis auf $T \cdot c_L \cdot \sum_{r=0}^S p(r)$ in $K(S)$ berücksichtigt.

$$T \cdot c_F \cdot \sum_{r=S+1}^{\infty} \frac{(r-S)^2}{2 \cdot r} \cdot p(r) = T \cdot c_F \cdot \sum_{r=S+1}^{\infty} \frac{r^2 - 2 \cdot r \cdot S + S^2}{2 \cdot r} \cdot p(r)$$

$$-T \cdot c_F \sum_{r=S+1}^{\infty} S \cdot p(r)$$

Die Summe ist bis auf $\sum_{r=S+1}^{\infty} p(r)$ in $K(S)$ berücksichtigt.

$$\sum_{r=0}^S p(r) + \sum_{r=S+1}^{\infty} p(r) = 1$$

$$\sum_{r=S+1}^{\infty} p(r) = 1 - \sum_{r=0}^S p(r)$$

Daraus folgt: $-T \cdot c_F + T \cdot c_F \sum_{r=0}^S p(r)$

$$T \cdot c_F \cdot \sum_{r=S+1}^{\infty} \frac{(S+1)^2}{2 \cdot r} \cdot p(r) = T \cdot c_F \cdot \sum_{r=S+1}^{\infty} \frac{S^2 + 2S + 1}{2 \cdot r} \cdot p(r)$$

Die Summe ist bis auf $T \cdot c_F \cdot \sum_{r=S+1}^{\infty} (S + \frac{1}{2}) \cdot \frac{p(r)}{r}$ in $K(S)$ berücksichtigt.

$$\text{Entsprechendes gilt für: } T \cdot c_L \cdot \sum_{r=S+1}^{\infty} (S + \frac{1}{2}) \cdot \frac{p(r)}{r}$$

Wählt man die Periode zu einer ZE, dann gilt für den ganzzahligen Wert von S_0 , bei dem $K(S)$ das Minimum erreicht, folgende Beziehung:

$$K(S_0) - K(S_0 - 1) = (c_F + c_L) \cdot L(S_0 - 1) - c_F \leq 0$$

$$K(S_0 + 1) - K(S_0) = (c_F + c_L) \cdot L(S_0) - c_F \geq 0$$

$$L(S_0 - 1) \leq \frac{c_F}{c_F + c_L}$$

$$L(S_0) \geq \frac{c_F}{c_F + c_L}$$

S_0 ist also ein ganzzahliger Wert, für den gilt:

$$L(S_0 - 1) \leq \frac{c_F}{c_F + c_L} \leq L(S_0)$$

Aufgabe: Ein Fabrikant, der jedes Jahr eine kleine Serie von Maschinen auf Bestellung produziert, hat festgestellt, daß die Nachfrage seiner Kunden nach einem bestimmten Ersatzteil, welches für den Betrieb seiner Maschinen sehr wichtig ist, dem in der folgenden Tabelle dargestellten Wahrscheinlichkeitsgesetz folgt:

Nachfrage	0	1	2	3	4	5	6	7	8	9	>9
Wahrscheinlichkeit	0.01	0.05	0.08	0.12	0.16	0.16	0.18	0.14	0.08	0.02	0

Die Lagerkosten betragen 100 GE je Stück und die Fehlbestandskosten werden auf 1000 GE je Stück geschätzt. Der Fabrikant möchte den optimalen Lagerbestand je Maschine für dieses Maschinenteil berechnen, damit er dessen Bestellung in die laufende Serienproduktion einbeziehen kann. Zur Berechnung benutzt man zweckmäßigerweise die folgende Tabelle:

S	r	p(r)	p(r)/r	$\sum_{r=S+1}^{\infty} \frac{p(r)}{r}$	$(S + \frac{1}{2}) \cdot \sum_{r=S+1}^{\infty} \frac{p(r)}{r}$	L(S)
0	0	0.01	-	0.2642	0.1321	0.1421
1	1					0.3813
2	2					0.5755
3	3					0.7297
4	4					0.8439
5	5					0.9221
6	6					0.9693
7	7					0.9915
8	8					0.9987
9	9	0.02	0.002	0	0	1

$$\frac{c_F}{c_F + c_L} = \frac{1000}{100 + 1100} = 0.909$$

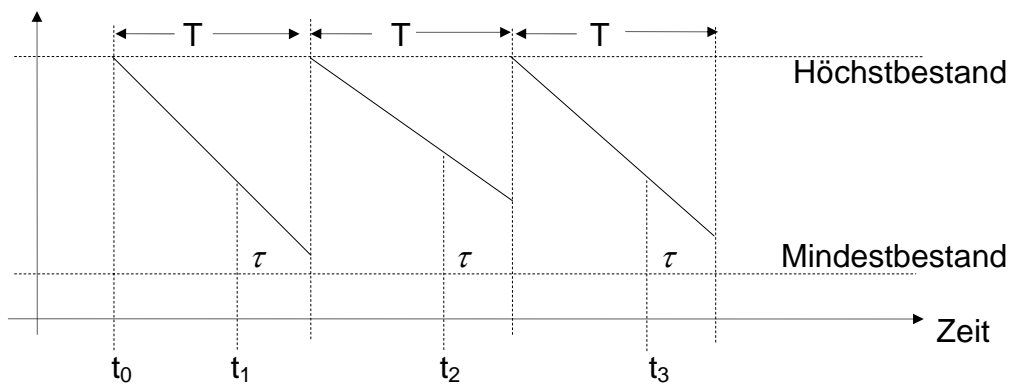
$$L(4) = 0.8348 < 0.909 < 0.921 = L(5)$$

Man sieht, daß $S_0=5$ ist.

b) Der Bedarf ist variabel

Das Lager wird durch variable Bestellmengen q_i bis zu seinem Höchstbestand aufgefüllt. Die q_i müssen zum Zeitpunkt t_i der i -ten Bestellung geschätzt werden (unter Berücksichtigung der Lieferzeit τ)

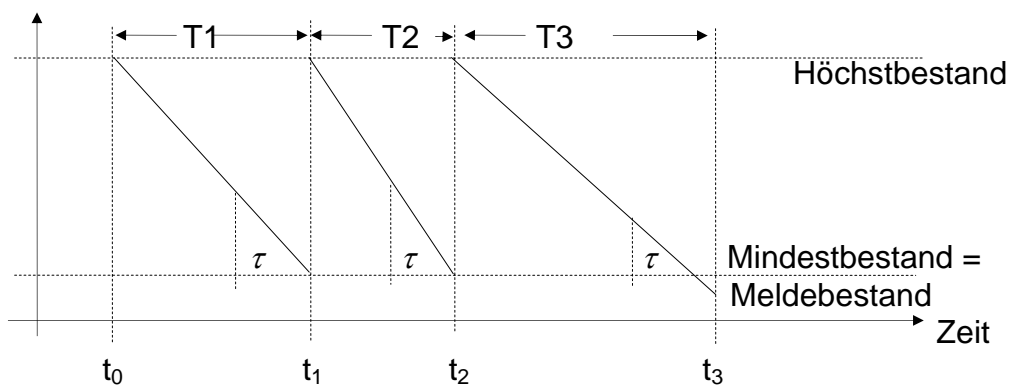
Lagerbestand



2. Planung mit variabler Periode

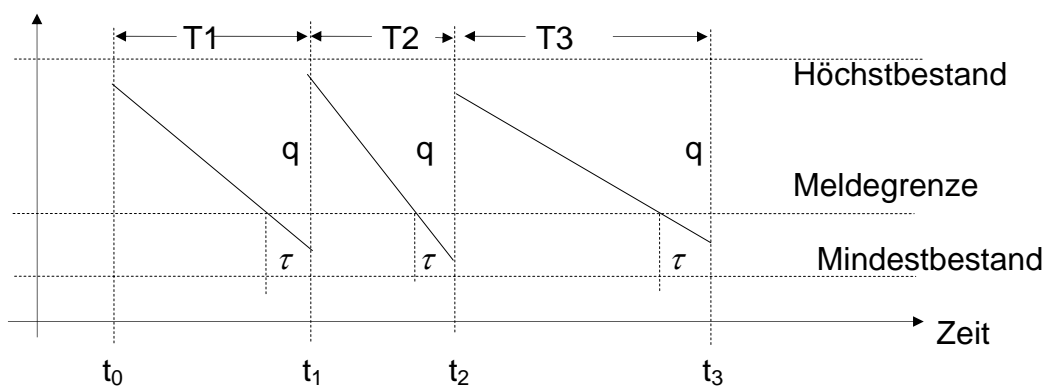
a) variabler Bedarf (determiniert bzw. stochastisch) und feste Bestellzeitpunkte

Lagerbestand



b) variabler Bedarf (determiniert bzw. stochastisch) und konstante Bestellmengen

Lagerbestand



2.1.4.3 Abgrenzung der Extremalwertprobleme zur linearen Optimierung

Methoden der Differentialrechnung zur Extremwertbestimmung sind für lineare Optimierungsprobleme nicht geeignet: Extremwerte treten an den Stellen in der linearen Planungsrechnung auf, für die Funktionen nicht differenzierbar sind. Die lineare Optimierung hat für OR-Probleme eine größere Bedeutung als die klassischen Methoden zum Bestimmen von Extremwerten. Allerdings können auch Lagerhaltungsmodelle zu Optimierungsproblemen führen, wie das folgende gemischte Lagerhaltungs-, Zuteilungsmodell zeigt:

Gelagert werden n Produkte mit dem Bedarf b_i ($1 \leq i \leq n$) je Periode. Fehlmengen sind nicht erlaubt. Weiterhin gelten die Vereinbarungen:

p_i die für eine ME des i -ten Produkts benötigte Lagerfläche in m^2/ME

p Begrenzung der Gesamtlagerfläche in m^2

t_{Si} die Stückzeit des i -ten Produkts in ZE/ME

t_g der zur Verfügung stehende Maschinenzeitfonds

t_{Ai} die Vorbereitungszeit je Los (Bestellmenge) x_i des i -ten Produkts in ZE

c_{Li} Lagerkostensatz (GE je ME und ZE)

c_{Bi} Beschaffungs-, Bestellkostensatz (GE/ME)

Gesucht sind die minimalen Lagerkosten.

Nebenbedingungen

$$(\text{Mittlere Lagerbelegung}) \quad \sum_{i=1}^n \frac{p_i \cdot x_i}{2} \leq p$$

$$(\text{Zeitfondsbegrenzung}) \quad \sum_{i=1}^n (t_{Si} + \frac{t_{Ai}}{x_i}) \cdot b_i \leq t_g$$

t_{Si} , b_i sind bekannt und können in einem sog. Restzeitfond zusammengefaßt werden.

$$t_r = t_g - \sum_{i=1}^n t_{Si} \cdot b_i$$

$$\text{Dann ergibt sich: } \sum_{i=1}^n \frac{t_{Ai} \cdot b_i}{x_i} \leq t_r$$

Zielfunktion

$$Z_{\min} = \sum \frac{c_{Bi} \cdot b_i}{x_i} + \frac{c_{Li} \cdot x_i}{2}$$

Es liegt hier ein nicht lineares Optimierungsproblem vor, da Zielfunktion und eine Nebenbedingung in ihren Variablen nichtlinear sind.

Die lineare Planungsrechnung kann nur dann angewandt werden, wenn Problemstellungen durch lineare Beziehungen hinreichend genau beschrieben werden. Mit Hilfe einer adäquaten mathematischen Theorie (z.B. Simplex-Verfahren) wird das Problem gelöst.

2.2 Grafische Lösungsverfahren

2.2.1 Normalformen der linearen Planungsrechnung mit 2 Variablen

2.2.1.1 Die Normalform der Maximaufgabe

Ein einführendes Beispiel: Produktionsproblem

In einem Schmelzbetrieb werden die Metalle Kupfer und Zink gewonnen. Höchstens kann in den Schmelzöfen 16 t Metall an einem Tag produziert werden, davon höchstens 12 t Zink. Die Kosten für die Produktion einer Tonne Kupfer betragen 3 GE, für eine Tonne Zink 1 GE. Die Kosten sollen 36 GE an einem Tag nicht überschreiten. 1 t Kupfer erzielt 4 GE, 1t Zink erzielt 3 GE Gewinn. Wieviel Tonnen Kupfer und Zink müssen in dem Betrieb an einem Tag produziert werden, damit der Gewinn möglichst groß wird.

Mathematisches Modell

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0$$

$$(2) \quad \begin{array}{rcl} x_1 + x_2 & \leq & 16 \\ & x_2 & \leq 12 \\ 3x_1 + x_2 & \leq & 36 \end{array}$$

$$(3) \quad Z_{\max} = 4x_1 + 3x_2$$

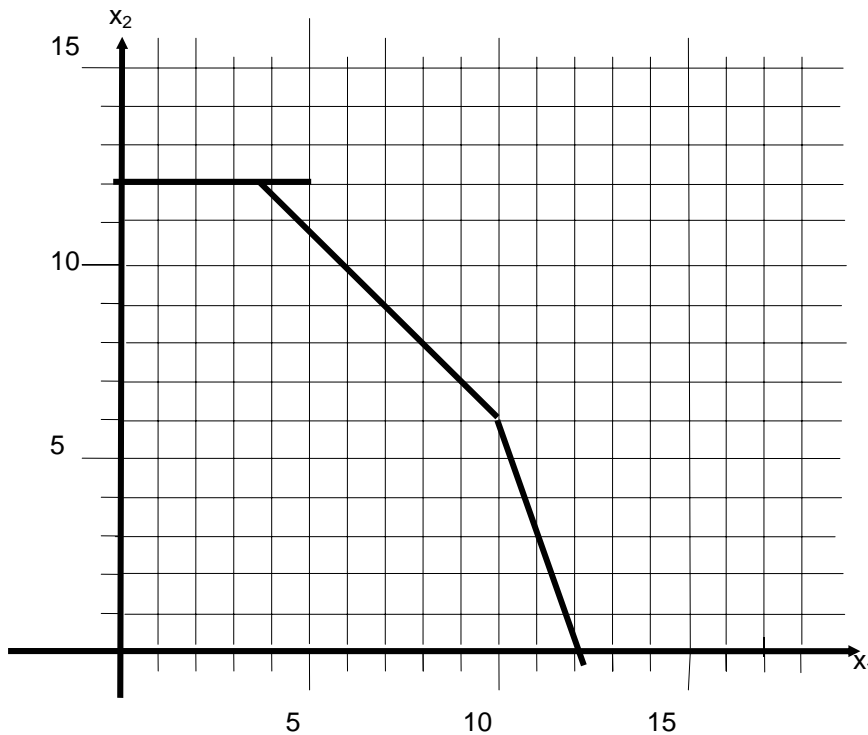
Grafische Lösung

- Menge der Wertepaare (x_1, x_2) , die **(1)** erfüllen
- Menge der Wertepaare, die **(1)** und **(2)** erfüllen. Aus der Menge der zulässigen Lösungen ist das Wertepaar zu bestimmen, für das die Zielfunktion einen maximalen Wert annimmt
- **(3)** kann in folgender Form als Gleichung für eine Schar von Parallelen mit dem Parameter Z aufgefaßt werden

$$x_2 = -4/3x_1 + Z/3$$

Den maximalen Wert der Zielfunktion erhält man durch folgende Wahl aus der Schar der parallelen Geraden:

- Die Gerade muß mindestens einen Punkt mit dem Lösungspolyeder gemeinsam haben
- Der Achsabstand auf der x_2 -Achse muß möglichst groß sein



Allgemeines Lösungsverfahren

(1) $x_1 \geq 0, x_2 \geq 0$

(2) $a_{11}x_1 + a_{12}x_2 \leq b_1$
 $a_{21}x_1 + a_{22}x_2 \leq b_2$

.....
 $a_{m1}x_1 + a_{m2}x_2 \leq b_m$

(3) $Z_{\max} = c_1x_1 + c_2x_2$

Aus der Menge der zulässigen Lösungen ist das Wertepaar (x_1, x_2) zu bestimmen, für das die Zielfunktion einen maximalen Wert annimmt:

$$x_2 = -\frac{c_1}{c_2}x_1 + \frac{Z}{c_2}$$

$\frac{Z}{c_2}$ gibt den Achsabschnitt auf der x_2 -Achse an. Z ist maximal, wenn dieser Achsabschnitt möglichst groß ist.

2.2.1.2 Normalformen der Minimaufgabe

Einführendes Beispiel

Ein Unternehmen benötigt drei Mineralien M_1 , M_2 , M_3 zur Weiterverarbeitung. Diese Mineralien werden in einem Separationsprozeß aus den Rohstoffen R_1 und R_2 gewonnen. Die Ausbeute der Rohstoffe hinsichtlich der drei Mineralien ist verschieden. Die Mengen (in t) der Mineralien, die man jeweils aus einer Tonne Rohstoff gewinnen kann, sind in der folgenden Übersicht angegeben:

Mineralien: Rohstoff:	M_1	M_2	M_3
R_1	0.5	0.3	0.6
R_2	1.5	0.6	0.5

In einem Monat werden vom Mineral M_1 6 t, von Mineral M_2 3 t und vom Mineral M_3 3 t für die weitere Produktion benötigt. Die Gestehungspreise der beiden Rohstoffe R_1 und R_2 sind 20 bzw. 25 DM/t. Welche Rohstoffmengen sollen zur Separation der Mineralien gekauft werden, damit die Rohstoffkosten minimal sind?

Mathematisches Modell

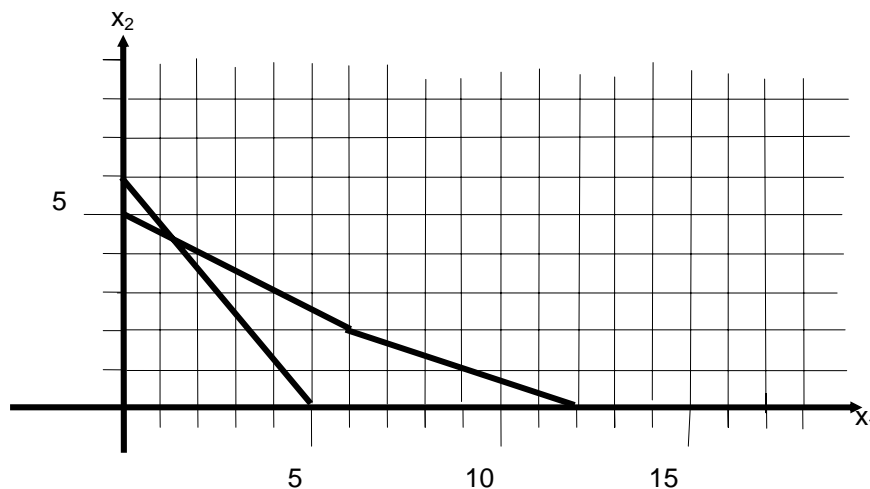
(1) $x_1 \geq 0, x_2 \geq 0$

(2)
$$\begin{aligned} 0.5x_1 + 1.5x_2 &\geq 6 \\ 0.3x_1 + 0.6x_2 &\geq 3 \\ 0.6x_1 + 0.5x_2 &\geq 3 \end{aligned}$$

Umformen:
$$\begin{aligned} x_1 + 3x_2 &\geq 12 \\ x_1 + 2x_2 &\geq 10 \\ 6x_1 + 5x_2 &\geq 30 \end{aligned}$$

(3) $Z_{\min} = 20x_1 + 25x_2$

Grafische Lösung



Optimale Lösung: $x_1 = 10/7, x_2 = 30/7, Z_{\min} = 950/7$

Allgemeines Lösungsverfahren

$$(1) x_1 \geq 0, x_2 \geq 0$$

$$(2) \begin{aligned} a_{11}x_1 + a_{12}x_2 &\geq b_1 \\ a_{21}x_1 + a_{22}x_2 &\geq b_2 \end{aligned}$$

$$\dots \dots \dots a_{m1}x_1 + a_{m2}x_2 \geq b_m$$

$$(3) Z_{\min} = c_1x_1 + c_2x_2$$

Aus der Menge der zulässigen Lösungen ist das Wertepaar (x_1, x_2) zu bestimmen, für das die Zielfunktion einen maximalen Wert annimmt:

$$x_2 = -\frac{c_1}{c_2}x_1 + \frac{Z}{c_2}$$

$\frac{Z}{c_2}$ gibt den Achsabschnitt auf der x_2 -Achse an. Z ist minimal, wenn dieser Achsabschnitt möglichst klein ist.

2.2.2 Die Lösungsmenge für Probleme mit 2 Variablen

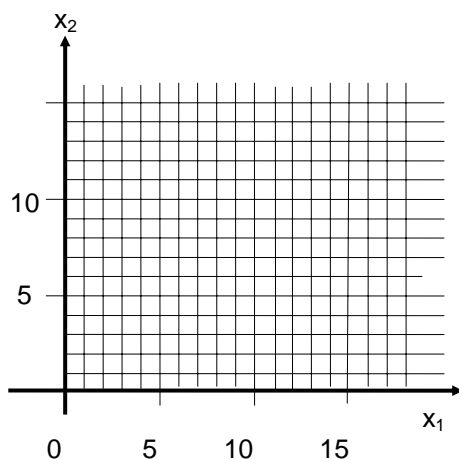
2.2.2.1 Probleme mit mehr als einer Lösung

Maximumaufgabe

$$(1) x_1 \geq 0, x_2 \geq 0$$

$$(2) \begin{aligned} x_1 + x_2 &\leq 16 \\ x_2 &\leq 12 \\ 3x_1 + x_2 &\leq 36 \end{aligned}$$

$$(3) Z_{\max} = 3x_1 + 3x_2$$

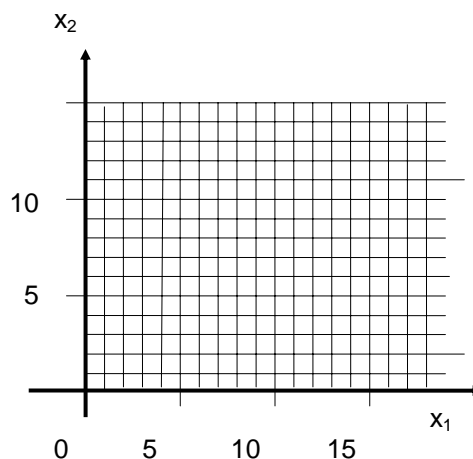


Minimumausgabe

$$x_1 \geq 0, x_2 \geq 0$$

$$\begin{aligned} x_1 + 3x_2 &\geq 12 \\ x_1 + 2x_2 &\geq 10 \\ 6x_1 + 5x_2 &\geq 30 \end{aligned}$$

$$Z_{\min} = 10x_1 + 20x_2$$



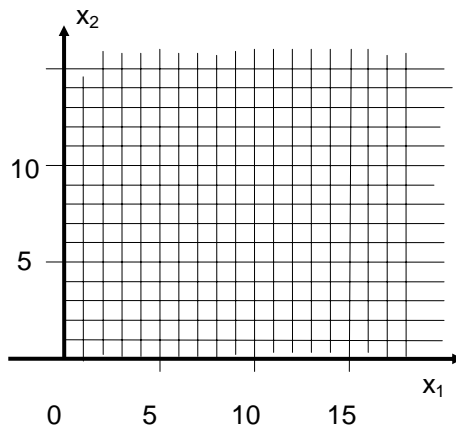
2.2.2.2 Probleme ohne optimale Lösung

Maximumaufgabe

(1) $x_1 \geq 0, x_2 \geq 0$

(2)
$$\begin{aligned} 2x_1 - 5x_2 &\leq 5 \\ 3x_1 - 2x_2 &\leq 24 \\ -2x_1 + x_2 &\leq 4 \end{aligned}$$

(3) $Z_{\max} = 2x_1 + 4x_2$

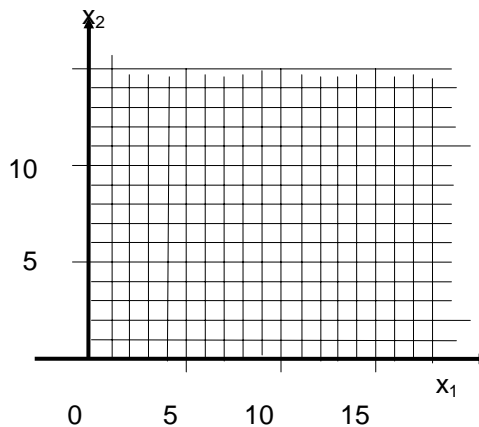


Minimierungsaufgabe

$x_1 \geq 0, x_2 \geq 0$

$$\begin{aligned} x_1 + x_2 &\geq 10 \\ 2x_1 + x_2 &\geq 12 \\ -x_1 &\geq 2 \end{aligned}$$

$Z_{\min} = 4x_1 + 6x_2$



2.2.2.3 Probleme mit genau einer Lösung

Die Menge der optimalen Lösung hat genau ein Element, wenn der Graf der Zielfunktion zur Bestimmung der optimalen Menge mit dem Lösungspolyeder nur einen Punkt gemeinsam hat. Dieser Punkt ist ein Eckpunkt und bestimmt eine zulässige Basislösung.

2.2.2.4 Charakterisierung der Lösungsmenge

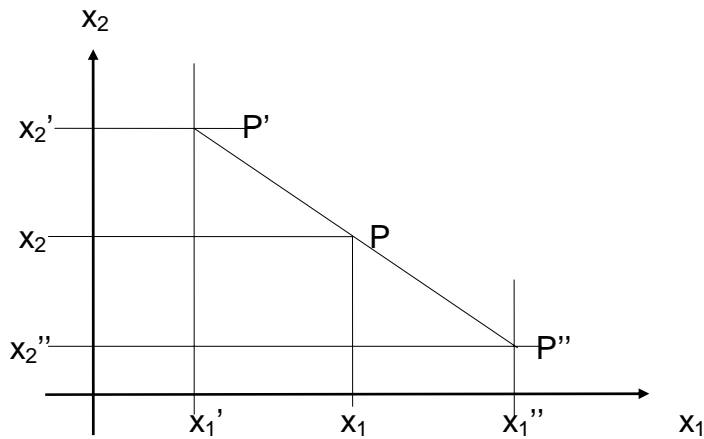
Durch Nichtnegativitätsbedingungen und einschränkende Bedingungen werden jeweils Halbebenen bestimmt, die die Menge der zulässigen Lösungen umfassen. Die so charakterisierten Mengen sind konvexe Punktmenge.

Definition: Eine **Punktmenge** heißt konvex, wenn mit 2 Punkten P' und P'' auch alle Punkte P der Strecke $P'P''$ zu der Menge gehören.

Sind $P' (x_1', x_2')$ und $P'' (x_1'', x_2'')$ Punkte der Menge, dann sind auch alle Punkte $P (x_1, x_2)$ mit den Koordinaten

$$\begin{aligned} x_1 &= k \cdot x_1'' + (1-k)x_1' \\ x_2 &= k \cdot x_2'' + (1-k)x_2' \end{aligned}$$

Punkte der Menge:



$$P'P = k \cdot P'P'' \quad (0 \leq k \leq 1)$$

$$x_1 - x_1' = k(x_1'' - x_1')$$

$$x_2 - x_2' = k(x_2'' - x_2')$$

$$x_1 = k \cdot x_1'' + (1 - k)x_1'$$

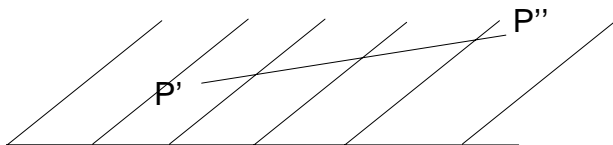
$$x_2 = k \cdot x_2'' + (1 - k)x_2'$$

Beispiele für konvexe Punkte in der Ebene:

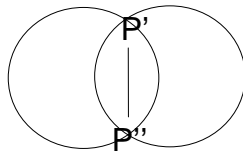
1. Punkte einer Geraden



2. Punkte in einer Halbebene

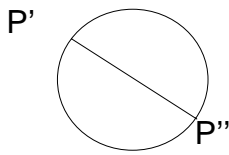


3. Punkte in einer Schnittmenge

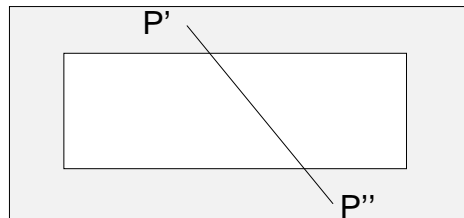


Beispiele für nicht konvexe Punkte in der Ebene

1. Punkte einer Kreislinie

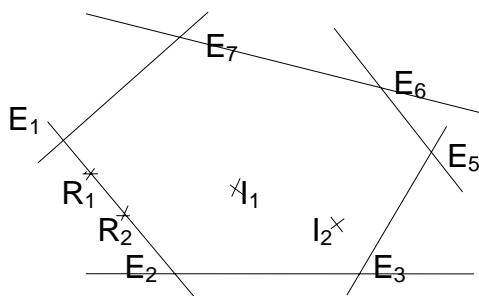


2. Die Punkte einer Fläche zwischen 2 Vierecken



Für konvexe Punktmengen gilt der Satz: Sind M_1, M_2, \dots, M_n konvexe Punktmen

gen, dann ist auch die Schnittmenge $D = M_1 \cap M_2 \cap \dots \cap M_n$ konvex. Da Punkte einer Halbm

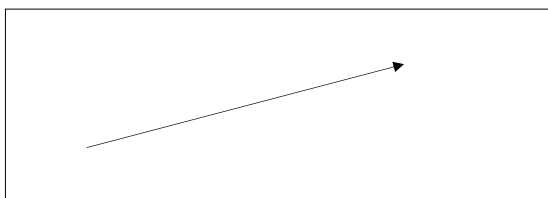


I_1, I_2, \dots : innere Punkte
 R_1, R_2, \dots : Randpunkte, keine Eckpunkte
 E_1, E_2, \dots : Eckpunkte

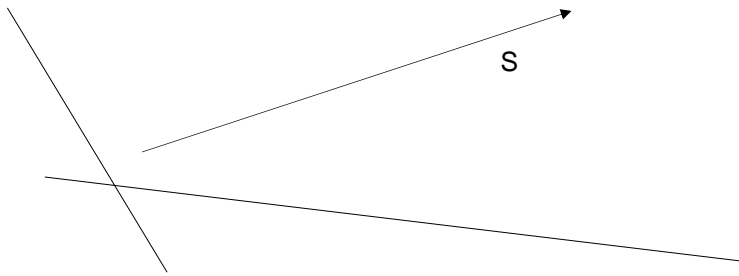
Eine konvexe Punktmenge heißt unbeschränkt, wenn es einen Strahl gibt, so daß alle Punkte des Strahls auch Punkte der konvexen Punktmenge sind. Alle anderen Punktmen

Beispiele:

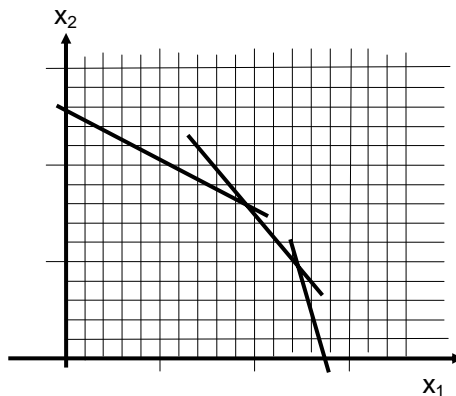
1. Beschränkt konvexe Punktmen



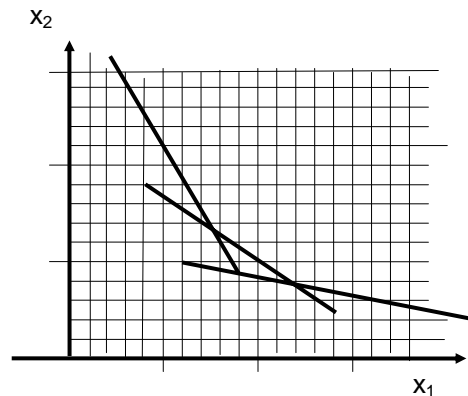
2. Unbeschränkt konvexe Punktmenge



Polyeder sind speziell konvexe Mengen. Eine konvexe Punktmenge ist dann ein Polyeder (oder konvexes Polyeder), wenn die konvexe Punktmenge beschränkt ist, und wenn sie nur endlich viele Eckpunkte hat, z.B.:



Die Punktmenge ist beschränkt und hat endlich viele Eckpunkte. Die Punktmenge ist ein Polyeder.



Die Punktmenge ist unbeschränkt. Sie ist daher kein Polyeder.

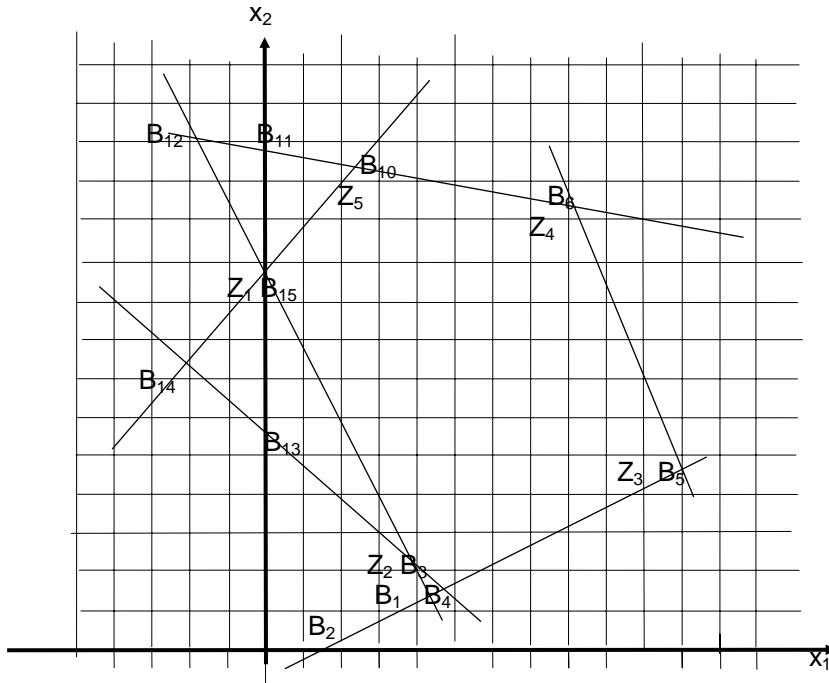
Die Lösungsmenge eines Problems der linearen Planungsrechnung ist ein Polyeder.

2.2.2.5 Basislösungen und zulässige Basislösungen

Zulässige Lösungen: Sie werden bestimmt durch (1) und (2) der allgemeinen Form eines Problems der linearen Planungsrechnung.

Basislösungen: Jeweils 2 nicht parallele Grenzgeraden bestimmen einen Schnittpunkt. Alle Wertepaare heißen Basislösungen.

Zulässige Basislösungen: Basislösungen, die auch zulässige Basislösungen sind. Die Menge der zulässigen Basislösungen wird durch die Eckpunkte des Lösungspolyeders bestimmt.



2.2.2.6 Die optimale Lösung

Bisher wurde erkannt: Die Zielfunktion eines Problems der linearen Planungsrechnung nimmt ihr Optimum für mindestens einen Eckpunkt des Lösungspolyeders, d.h. für mindestens eine zulässige Basislösung an.

Beweis:

$P'(x_1', x_2')$ und $P''(x_1'', x_2'')$ sind Randpunkte des Polyeders. Für diese Punkte hat die Zielfunktion ($Z = c_1x_1 + c_2x_2$) die Werte:

$$Z' = c_1x_1' + c_2x_2'$$

$$Z'' = c_1x_1'' + c_2x_2''$$

Ein beliebiger Punkt P' der Strecke $P'P''$ hat die Koordinaten:

$$x_1 = k \cdot x_1'' + (1 - k)x_1'$$

$$x_2 = k \cdot x_2'' + (1 - k)x_2'$$

Alle Punkte $P(x_1, x_2)$ der Strecke $P'P''$ sind zulässige Lösungen eines linearen Optimierungssystems, da die Menge der zulässigen Lösungen eines linearen Optimierungsproblems konvex ist. Damit ergibt sich für Z

$$\begin{aligned} Z &= c_1x_1 + c_2x_2 = c_1(kx_1'' + (1 - k)x_1') + c_2(kx_2'' + (1 - k)x_2') \\ &= k(c_1x_1'' + c_2x_2'') + (1 - k)(c_1x_1' + c_2x_2') \\ &= kZ'' + (1 - k)Z' \end{aligned}$$

Es gibt 3 Möglichkeiten:

1. $Z' = Z''$

Aus $Z = kZ'' + (1 - k)Z'$ folgt: $Z - Z' = k(Z'' - Z')$. Für $Z' = Z''$ folgt weiter $Z = Z'$. Das bedeutet: Z hat für die Wertepaare aller Punkte der Strecke $P'P''$ den gleichen Wert.

2. $Z'' < Z'$

Aus $Z - Z' = k(Z'' - Z')$ folgt für $Z'' < Z'$, daß $k(Z'' - Z')$ negativ ist. Das bedeutet: $Z - Z' < 0$ oder $Z < Z'$. Die Zielfunktion erreicht für das Wertepaar des Randpunkts P' ihr Maximum.

3. $Z'' > Z'$

Aus $Z = kZ'' + (1 - k)Z'$ folgt $Z - Z'' = (1 - k)(Z' - Z'')$. Das bedeutet: Für $Z'' > Z'$ ist $Z - Z'' < 0$ oder $Z < Z''$. Die Zielfunktion erreicht in diesem Fall für das Wertepaar des Randpunkts P'' ihr Maximum.

Die Zielfunktion erreicht in allen Fällen ihr Maximum in einem Eckpunkt, d.h. mindestens in einem Eckpunkt E . Wäre das nicht so, so könnte man durch den Punkt E eine Randstrecke legen und die gleiche Überlegung für diese Strecke ausführen. Ein Problem der linearen Planungsrechnung kann also das Maximum (oder Minimum) nur in einem Eckpunkt des Polyeders erreichen.

2.2.3 Sonderfälle

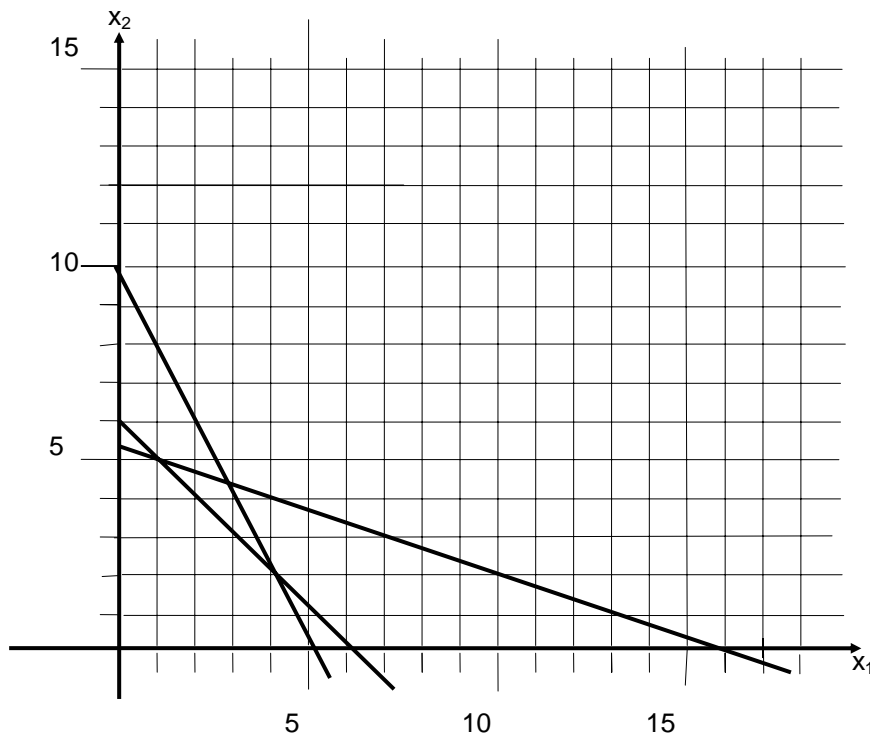
2.2.3.1 Die ganzzahlige Planungsrechnung

1. Beispiel

(1) $x_1 \geq 0, x_2 \geq 0 \quad (x_1, x_2) \in \mathbb{Z} \times \mathbb{Z}$

(2)
$$\begin{aligned} x_1 + x_2 &\leq 6 \\ 2x_1 + x_2 &\leq 10 \\ x_1 + 3x_2 &\leq 16 \end{aligned}$$

(3) $Z_{\max} = x_1 + 2x_2$



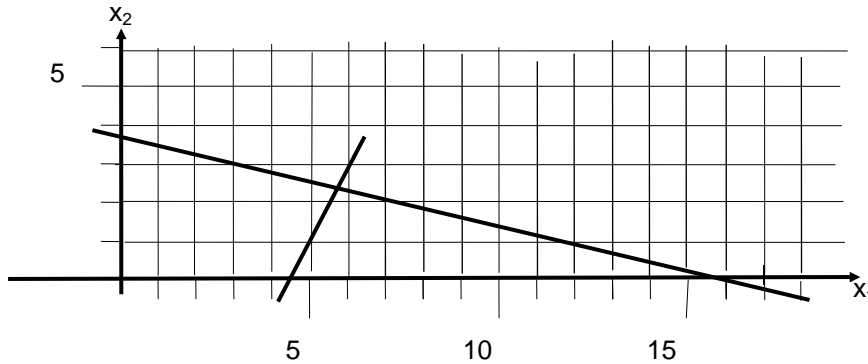
Optimale Lösung: $x_1 = 1, x_2 = 5, Z_{\max} = 11$

2. Beispiel

(1) $x_1 \geq 0, x_2 \geq 0 \quad (x_1, x_2) \in \mathbb{Z} \times \mathbb{Z}$

(2)
$$\begin{aligned} 2x_1 - x_2 &\leq 9 \\ 2x_1 + 8x_2 &\leq 31 \end{aligned}$$

(3) $Z_{\max} = 2x_1 + 5x_2$



Optimale Lösung ohne Berücksichtigung der Ganzzahligkeit: $x_1 = 5\frac{13}{18}$, $x_2 = 2\frac{4}{9}$,

$Z_{opt} = 23\frac{2}{3}$. Runden führt zu $x_1 = 5$, $x_2 = 2$, $Z_{opt} = 20$.

Optimale Lösung mit Berücksichtigung der Ganzzahligkeitsbedingung: $x_1 = 3$, $x_2 = 3$, $Z_{max} = 21$

2.2.3.2 Die Parametrische Planungsrechnung

a) Ein einführendes Beispiel

Ein Unternehmen produziert ein Gut T in 2 verschiedenen Qualitäten T_1 und T_2 . Die Herstellung erfolgt so, daß jedes Stück nacheinander auf den Maschinen M_1 und M_2 bearbeitet wird. Die Bearbeitungszeiten für ein Stück jeder Qualität sind für die beiden Maschinen verschieden. Der Stückgewinn (hier: die Differenz zwischen Verkaufspreis und Materialkosten) ist bekannt. Die Arbeitszeit wird in Minuten gemessen. Das Unternehmen legt diesen Planungen eine wöchentliche Arbeitszeit von s Minuten zugrunde:

	T_1	T_2
Bearbeitungszeit M_1	a_{11}	a_{12}
M_2	a_{21}	a_{22}
Stückgewinn (GE)	c_1	c_2
Hergestellte Menge	x_1	x_2

Wieviel von jeder Qualität muß erzeugt werden, um den Gewinn zu maximieren?

Mathematisches Modell

(1) $x_1 \geq 0$, $x_2 \geq 0$

(2) $a_{11}x_1 + a_{12}x_2 \leq s$
 $a_{21}x_1 + a_{22}x_2 \leq s$

(3) $Z_{max} = c_1x_1 + c_2x_2$

Grafische Lösung:

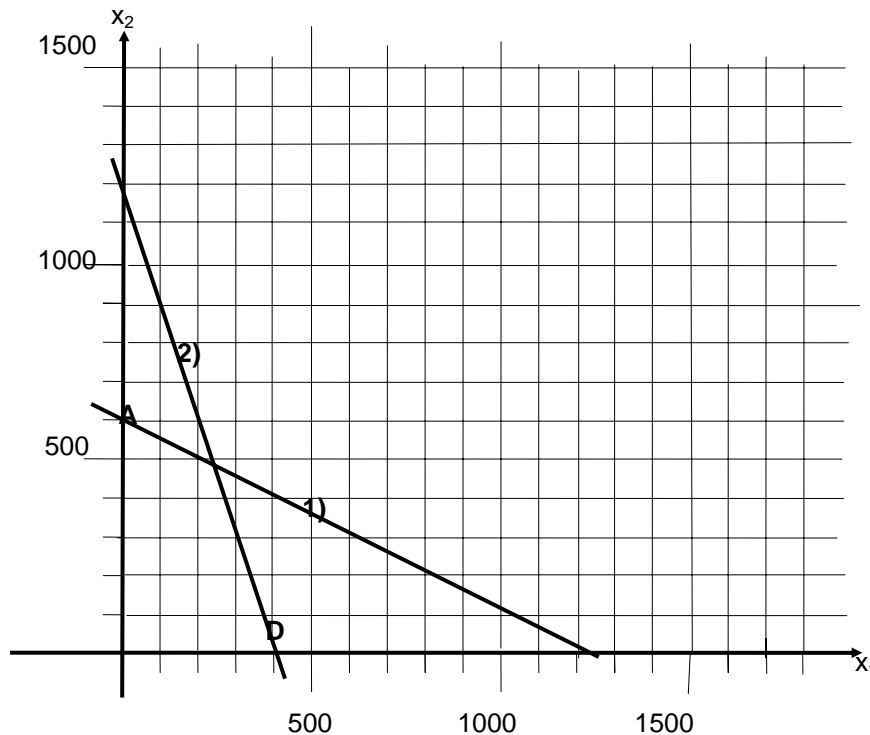
Zur Lösung werden folgende Zahlenwerte herangezogen:

$$a_{11} = 2, a_{12} = 4, c_1 = 4, a_{21} = 6, a_{22} = 2, c_2 = 3$$

$$(2) \quad 2x_1 + 4x_2 \leq 2400$$

$$6x_1 + 2x_2 \leq 2400$$

$$(3) \quad Z_{\max} = 4x_1 + 3x_2$$



Folgende Fälle sind zu unterscheiden:

1. $\frac{c_1}{c_2} \leq \frac{1}{2}$ (Der Graf der Zielfunktion verläuft sehr flach. Das Gewinnmaximum liegt im Punkt A)
2. $\frac{c_1}{c_2} = \frac{1}{2}$ (Der Graf der Zielfunktion besitzt die gleiche Steigung wie **1**)
3. $\frac{1}{2} \leq \frac{c_1}{c_2} \leq 3$ (Das Gewinnmaximum liegt im Schnittpunkt von **1** und **2**)
4. $\frac{c_1}{c_2} = 3$ (Der Graf der Zielfunktion hat die gleiche Steigung wie **2**)
5. $\frac{c_1}{c_2} \geq 3$ (Der Graf der Zielfunktion verläuft sehr steil. Das Gewinnmaximum liegt im Punkt D.)

b) Verallgemeinerung

In welchen Koeffizienten können sich die Koeffizienten c_1 und c_2 in der Zielfunktion ändern, ohne daß sich das entsprechende optimale Wertepaar (x_1, x_2) ändert?

Erweiterung der Zielfunktion mit den Parametern p_1 und p_2 :

$$Z = (c_1 + p_1)x_1 + (c_2 + p_2)x_2$$

(Ebenfalls denkbar: $Z = c_1 p_1 x_1 + c_2 p_2 x_2$)

Anstieg des Zielfunktionsgraphen: $m = -\frac{c_1 + p_1}{c_2 + p_2} \wedge (c_2 + p_2) \neq 0$

Bedingung für den Erhalt des optimalen Wertepaares: $m_2 \leq -\frac{c_1 + p_1}{c_2 + p_2} \leq m_1$

m_1, m_2 : Anstiege der Begrenzungsgraden, die sich in P schneiden.

Beispiele:

Maximumaufgabe

(1) $x_1 \geq 0, x_2 \geq 0$

(2) $x_1 + x_2 \leq 16$

$x_2 \leq 12$

$3x_1 + x_2 \leq 36$

Minimumaufgabe

$x_1 \geq 0, x_2 \geq 0$

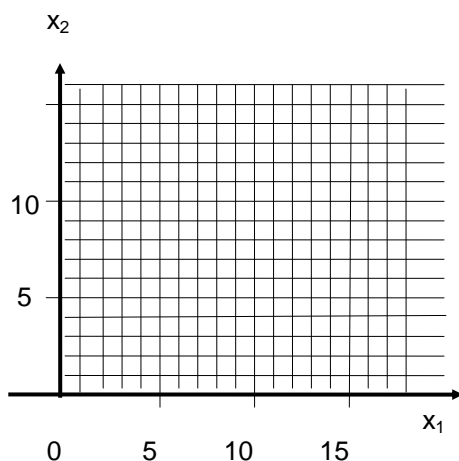
$2x_1 + x_2 \geq 16$

$x_2 \geq 2$

$2x_1 + 3x_2 \geq 32$

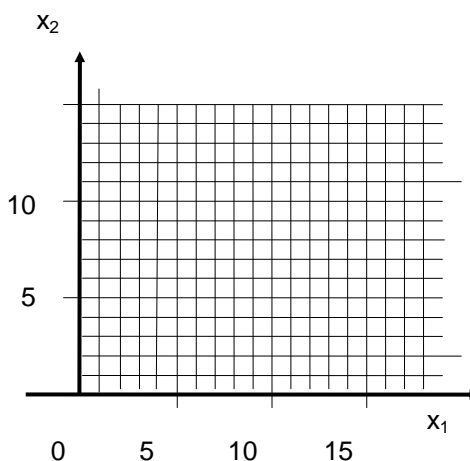
$2x_1 + 5x_2 \geq 40$

(3) $Z_{\max} = (4 + p)x_1 + (3 + p)x_2 \quad Z_{\min} = (10 + p)x_1 + (8 + p)x_2$



Optimale Lösung für $p = 0$:

$x_1 = 10, x_2 = 6, Z_{\max} = 58$



Optimale Lösung für $p = 0$:

$x_1 = 4, x_2 = 8, Z_{\min} = 104$

Bedingung für den Erhalt des Optimums:

$-3 \leq \frac{4+p}{3+p} \leq -1$

$1 \leq \frac{4+p}{3+p} \leq 3$

$-2 \leq \frac{10+p}{8+p} \leq -\frac{2}{3}$

$\frac{2}{3} \leq \frac{10+p}{8+p} \leq 2$

Daraus folgt: $p \geq -5/2$

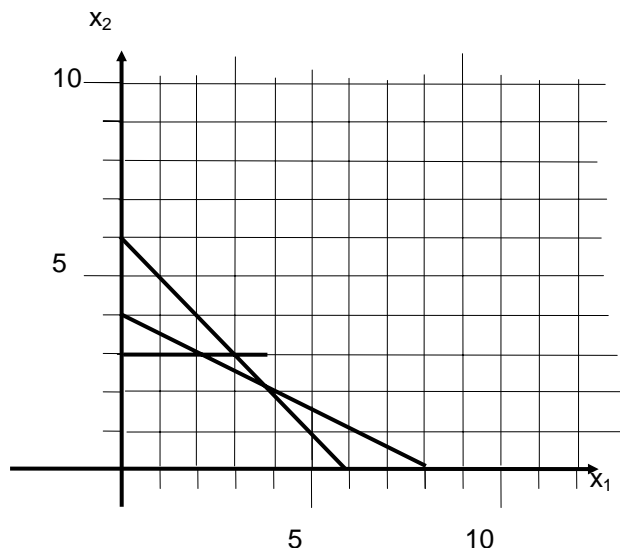
$p \geq -6$

2.3 Das Simplexverfahren

2.3.1 Einführung: Ein Beispiel aus der Produktion

Auf einer Maschine können 2 Stoffe verschiedener Webart hergestellt werden. In einer Zeiteinheit können maximal 8 m der Stoffart S_1 oder 4 m der Stoffart S_2 oder eine entsprechende Kombination beider Stoffarten produziert werden. Auf einer zweiten Maschine werden höchstens 6 m Stoff in der gleichen Zeiteinheit gedruckt. Für die Verpackung beider Produkte sind 2 Arbeiter zuständig. Der Arbeiter, der nur S_1 verpackt, schafft in der Zeiteinheit maximal 5 m. Der Arbeiter, der nur die zweite Stoffart verpackt, schafft in der Zeiteinheit höchstens 3 m. Die Verkaufspreise sind zu 3.- DM für den Meter S_1 und zu 8.- DM für den Meter S_2 festgelegt. Gesucht ist ein Produktionsplan, der einen maximalen Umsatz gestattet?

Grafische Lösung



Mathematisches Modell

(1) $x_1 \geq 0, x_2 \geq 0$

(2)
$$\begin{aligned} x_1 + 2x_2 &\leq 8 \\ x_2 &\leq 3 \\ x_1 + x_2 &\leq 6 \\ x_1 &\leq 5 \end{aligned}$$

(3) $Z_{\max} = 3x_1 + 8x_2$

Algebraisches Lösungsverfahren

1. Schritt: Einführung von Schlupfvariablen

Zweck: Überführung der Ungleichungen der Restriktionen in Gleichungen. Die Schlupfvariablen erfüllen wegen „ \leq “ die Nichtnegativitätsbedingungen

Zum mathematischen Modell ist dann die folgende Problembeschreibung mit den Schlupfvariablen (Gleichungssystem) äquivalent:

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0$$

$$(2) \quad x_1 + 2x_2 + y_1 = 8 \quad (2a)$$

$$+ x_2 + y_2 = 3 \quad (2b)$$

$$x_1 + x_2 + y_3 = 6 \quad (2c)$$

$$x_1 + y_4 = 5 \quad (2d)$$

$$(3) \quad Z_{\max} = 3x_1 + 8x_2 + 0y_1 + 0y_2 + 0y_3 + 0y_4$$

2. Schritt: Festlegung einer einfach zu bestimmenden Lösung

Die einschränkenden Bedingungen bestehen aus 4 Gleichungen mit 6 Variablen, davon können 4 Variable in Abhängigkeit von 2 Variablen beschrieben werden.

Anordnung: $x_1 = 0, \quad x_2 = 0$

Daraus folgt: $y_1 = 8 \quad y_2 = 3 \quad y_3 = 6 \quad y_4 = 5 \quad Z = 0$

Die Variablen, die den Wert 0 haben, heißen „**Nichtbasisvariable** (NBV)“. Die Variablen, die nicht den Wert 0 haben, heißen „**Basisvariable**“.

3. Schritt: Verbesserung der Lösung

Ist die gefundene Lösung eine optimale Lösung?

Aus $Z_{\max} = 3x_1 + 8x_2 + 0y_1 + 0y_2 + 0y_3 + 0y_4$ folgt: Z wird größer, wenn x_1 und x_2 positive Werte annehmen.

Die Verbesserung der Lösung kann erreicht werden durch Vergrößerung von x_2 . Grenzen sind gesetzt durch die einschränkenden Bedingungen:

Aus (2a) folgt: x_2 kann nicht größer als 4 sein

Aus (2b) folgt: x_2 kann nicht größer als 3 sein

Aus (2c) folgt: x_2 kann nicht größer als 6 sein

Da alle diese Bedingungen erfüllt sein müssen, ist $x_2 = 3$ der größtmögliche Wert. Daher benutzt man (2b) zum Eliminieren von x_2 aus dem übrigen Gleichungen. Das Gleichungssystem nimmt folgende Gestalt an:

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad y_3 \geq 0, \quad y_4 \geq 0$$

$$(2) \quad x_1 + y_1 - 2y_2 = 2 \quad (2a')$$

$$x_2 + y_2 = 3 \quad (2b')$$

$$x_1 - y_2 + y_3 = 3 \quad (2c')$$

$$x_1 + y_4 = 5 \quad (2d')$$

$$(3) \quad Z_{\max} - 24 = 3x_1 - 8y_2$$

Für den größtmöglichen Wert von x_2 ergibt sich:

$$x_1 = 0 \quad x_2 = 3 \quad y_3 = 6 \quad y_4 = 5 \quad Z = 24$$

Hinweis: $x_1 = 0 \quad x_2 = 3$ zeigt auf einen Eckpunkt.

Fortsetzung der Verfahrensweise mit dem 3. Schritt: Der Wert von Z kann vergrößert werden, indem x_1 einen möglichst großen positiven Wert annimmt.

Für wird aus (2a'), (2c'), (2d') der größtmögliche Wert bestimmt:

Aus (2a') folgt: $x_1 \leq 2$

Aus (2c') folgt: $x_1 \leq 3$

Aus (2d') folgt: $x_1 \leq 5$

Größtmöglicher Wert ist $x_1 = 2$

(2a') wird daher benutzt, um die Variable aus den übrigen Gleichungen zu eliminieren. Dann nimmt das Gleichungssystem folgende Gestalt an:

$$(1) \quad x_1 \geq 0, \quad x_2 \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad y_3 \geq 0, \quad y_4 \geq 0$$

$$(2) \quad \begin{aligned} x_1 + y_1 - 2y_2 &= 2 \\ x_2 + y_2 &= 3 \\ -y_1 + y_2 + y_3 &= 1 \\ -y_1 + 2y_2 + y_4 &= 3 \end{aligned}$$

$$(3) \quad Z_{\max} - 30 = -3y_1 - 2y_2$$

Lösung: Für den größtmöglichen Wert von x_1 ergibt sich $x_1 = 2, \quad x_2 = 3$

Hinweis: $x_1 = 2, \quad x_2 = 3$ zeigt auf einen Eckpunkt.

Die Fortsetzung des Verfahrens mit dem 3. Schritt ergibt: Aus $Z_{\max} - 30 = -3y_1 - 2y_2$ folgt: Eine weitere Verbesserung des Zielfunktionswertes ist nicht mehr möglich.

Optimale Lösung: $x_1 = 2, \quad x_2 = 3, \quad Z_{\max} = 30$

Je Zeiteinheit sollen von S_1 „2m“, von S_2 „3m“ Stoff hergestellt werden. Damit würde ein maximaler Umsatz von 30.- DM erzielt werden.

Zusammenfassende Darstellung

BV	x_1	x_2	y_1	y_2	y_3	y_4	b_i
y_1	1	2	1	0	0	0	8
y_2	0	1	0	1	0	0	3
y_3	1	1	0	0	1	0	6
y_4	1	0	0	0	0	1	5
Z	3	8	0	0	0	0	0
y_1	1	0	1	-2	0	0	2
x_2	0	1	0	1	0	0	3
y_3	1	0	0	-1	1	0	3
y_4	1	0	0	0	0	1	5
Z	3	0	0	-8	0	0	-24
x_1	1	0	1	-2	0	0	2
x_2	0	1	0	1	0	0	3
y_3	0	0	-1	1	1	0	4
y_4	0	0	-1	-2	0	1	3
Z_{opt}	0	0	-3	-2	0	0	-30

Auch das Hinzufügen der einer weiteren, dritten Variable ändert nichts an dem allgemeinen Lösungsschema:

(1) $x_1 \geq 0, x_2 \geq 0$

(2)
$$\begin{aligned} x_1 + 2x_2 + x_3 &\leq 8 \\ &+ x_2 &\leq 3 \\ x_1 + x_2 + x_3 &\leq 6 \\ x_1 &\leq 5 \end{aligned}$$

(3) $Z_{\max} = 3x_1 + 8x_2 + x_3$

Äquivalentes Gleichungssystem

(1) $x_1 \geq 0, x_2 \geq 0, y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_4 \geq 0$

(2)
$$\begin{aligned} x_1 + 2x_2 + x_3 + y_1 &= 8 \\ &+ x_2 + y_2 &= 3 \\ x_1 + x_2 + x_3 + y_3 &= 6 \\ x_1 + y_4 &= 5 \end{aligned}$$

(3) $Z_{\max} = 3x_1 + 8x_2 + x_3 + 0y_1 + 0y_2 + 0y_3 + 0y_4$

Lösungsverfahren

Mit der Ausgangslösung $x_1 = 0, x_2 = 0$ und der 1. Verbesserung $x_2 = 3$ ergibt sich folgendes äquivalentes Gleichungssystem:

(2)
$$\begin{aligned} x_1 + x_3 + y_1 - 2y_2 &= 2 \\ &+ x_2 + y_2 &= 3 \\ x_1 + x_3 - y_2 + y_3 &= 3 \end{aligned}$$

$$\begin{aligned} x_1 + y_4 &= 5 \\ (3) \quad Z_{\max} - 24 &= 3x_1 + x_3 + 0y_1 - 8y_2 + 0y_3 + 0y_4 \end{aligned}$$

Die 2. Verbesserung ergibt: $x_1 = 2$

$$\begin{aligned} (2) \quad x_1 + x_3 + y_1 - 2y_2 &= 2 \\ x_2 + y_2 &= 3 \\ -y_1 + y_2 + y_3 &= 1 \\ -x_3 - y_1 + 2y_2 + y_4 &= 3 \\ (3) \quad Z_{\max} - 30 &= -3y_1 - 2y_2 - 2x_3 + 0y_3 + 0y_4 \end{aligned}$$

Greift man aus der Tabellendarstellung den ersten und 3. Abschnitt heraus, dann lässt sich folgendes Schema ableiten:

BV	x_1^T $x_1 \quad x_2$		x_2^T x_3	y_1^T $y_1 \quad y_2$ y_4		y_2^T y_3	b_i
y_1 y_1 y_2	1 A_{11} 0	2 1	1 A_{12} 0	1 E_1 0	0 1	0 N_1 0	8 b_1 3
y_3 y_2 y_4	1 A_{21} 1	1 0	1 A_{22} 1	0 N_2 0	0 0	1 E_2 1	6 b_2 5
Z	c_1^T 3 8		c_2^T 1	n^T 0 0		n^T 0 0	0
x_1 x_1 x_2	1 E_1 1	0 0	1 $A_{11}^{-1} \cdot A_{12}$ 0	1 A_{11}^{-1} 0	-2 1	0 N_1 0	2 $A_{11}^{-1} \cdot b_1$ 3
y_3 y_2 y_4	0 N_2 0	1 $A_{22} - A_{21} \cdot A_{11}^{-1} \cdot A_{12}$ 0	0 -1	-1 $-A_{22} \cdot A_{11}^{-1}$ -1	1 2	1 E_2 1	⁶ $b_2 - A_{21} A_{11}^{-1} b_1$ 3
Z_{opt}	n^T $c_2^T - c_1^T \cdot A_{11}^{-1} \cdot A_{12}$ 0 0		-2	$-c_1^T \cdot A_{11}^{-1}$ -3 -2		n^T 0 0	$c_1^T A_{11}^{-1} b_1$ -30

Die Darstellung verweist auf eine allgemeine Basistransformation (Übergang von einer Basis zu einer anderen Basis des Vektorraums). Zusätzlich in das Schema, in dem die allgemeine Basistransformation abläuft, wurde hier noch die Zielfunktion aufgenommen.

⁶ $b_2 - A_{21} A_{11}^{-1} b_1$

2.3.2 Das Simplexverfahren für die Normalform der Maximumaufgabe

2.3.2.1 Das mathematische Modell für die Normalform mit n Variablen

1. Format (ausführliche Schreibweise)

$$(1) \quad x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$$

[illegible]

$$(3) \quad Z_{\max} = C_1X_1 + C_2X_2 + \dots + C_nX_n$$

2. Format (Darstellung in Matrizenform)

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & \dots & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & \dots & \dots & a_{mn} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ \dots \\ \dots \\ x_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ \dots \\ \dots \\ b_m \end{pmatrix} \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ \dots \\ \dots \\ c_n \end{pmatrix}$$

Mit diesen Bezeichnungen lautet das mathematische Modell:

(1) $\mathbf{x} \geq \mathbf{n}$

(2) $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b} \wedge \mathbf{b} \geq \mathbf{n}$

$$(3) \quad Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}$$

Mit der Einführung von Schlupfvariablen ergibt sich dann die folgende äquivalente Darstellung (Überführung des Ungleichungssystems in ein Gleichungssystem)

(1) $\mathbf{x} \geq \mathbf{n}$, $\mathbf{y} \geq \mathbf{n}$

(2) $\mathbf{A} \cdot \mathbf{x} + \mathbf{E} \cdot \mathbf{y} = \mathbf{b}$

$$(3) \quad Z_{\max} = \mathbf{c}^T \cdot \mathbf{x} + \mathbf{n}^T \cdot \mathbf{y}$$

3. Definitionen

Vektorraum

Die Menge der $(m \times n)$ -Matrizen - und damit auch die Menge der Vektoren mit m Elementen - bilden einen Vektorraum.

Folgende Eigenschaften besitzt ein Vektorraum über dem Körper der reellen Zahlen:

- (1) Für je zwei Elemente, z.B. \mathbf{A} und \mathbf{B} aus dem Vektorraum, ist eindeutig ein Element $\mathbf{A} + \mathbf{B} = a_{ij} + b_{ij} = c_{ij} = \mathbf{C}$ zugeordnet.

Für die Addition gilt

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$$

$$\mathbf{A} + \mathbf{0} = \mathbf{A}$$

$$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$$

- (2) Jedem Element, z.B. \mathbf{A} aus dem Vektorraum, wird durch die Multiplikation mit k (reelle Zahl) eindeutig ein Element $k \cdot \mathbf{A} = k \cdot a_{ij} = \mathbf{B}$ zugeordnet. Für die

Multiplikation gelten folgende Gesetze:

$$k \cdot \mathbf{A} = \mathbf{A} \cdot k$$

$$(k_1 \cdot k_2) \cdot \mathbf{A} = k_1 \cdot (k_2 \cdot \mathbf{A})$$

$$k \cdot (\mathbf{A} + \mathbf{B}) = k \cdot \mathbf{A} + k \cdot \mathbf{B}$$

$$(k_1 + k_2) \cdot \mathbf{A} = k_1 \cdot \mathbf{A} + k_2 \cdot \mathbf{A}$$

Dimension eines Vektorraums

Ein Vektorraum hat die Dimension n , wenn es n linear unabhängige Vektoren gibt.

Die Vektoren heißen linear unabhängig, wenn $\sum_{j=1}^n k_j \cdot \mathbf{x}_j = \mathbf{0}$ nur für $k_j = 0$ ($j = 1, \dots, n$) erfüllt ist.

Basis eines Vektorraums

Hat ein Vektorraum die Dimension n , so heißt jede Menge B von linear unabhängigen Vektoren Basis des Vektorraums. Die Vektoren einer Basis heißen Basisvektoren.

2.3.2.2 Die Basistransformation

Der Übergang von einer Basis B_1 zu einer anderen Basis B_2 heißt **Basistransformation**. Werden mehrere Vektoren ausgetauscht, so heißt der Übergang allgemeine Basistransformation.

Bsp.: Gegeben ist eine Basis des Vektorraums $B_1 = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{i-1}, \mathbf{b}_i, \dots, \mathbf{b}_n\}$. \mathbf{a} ist ein Vektor des Vektorraums ($\mathbf{a} \neq \mathbf{0}$) und kann folgendermaßen dargestellt werden:

$$\mathbf{a} = a_1 \cdot \mathbf{b}_1 + a_2 \cdot \mathbf{b}_2 + \dots + a_i \cdot \mathbf{b}_i + \dots + a_n \cdot \mathbf{b}_n \quad (1)$$

Wegen $\mathbf{a} \neq \mathbf{0}$ ist mindestens einer der Koeffizienten a_1, a_2, \dots, a_n ungleich 0. Es wird angenommen: $a_i \neq 0$. Dann ist auch B_2 eine Basis des Vektorraums:

$B_2 = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{i-1}, \mathbf{a}, \dots, \mathbf{b}_n\}$. \mathbf{a} kann nicht als Linearkombination folgender Vektoren geschrieben werden: $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \dots$

Wegen $\mathbf{a} = a_1 \cdot \mathbf{b}_1 + a_2 \cdot \mathbf{b}_2 + \dots + a_i \cdot \mathbf{b}_i + \dots + a_n \cdot \mathbf{b}_n$ mit a_i ungleich Null sind die Vektoren $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{i-1}, \mathbf{a}, \dots, \mathbf{b}_n$ linear unabhängig. (Die Bedingung, daß \mathbf{b}_i mit \mathbf{a} getauscht werden kann, ist a_i ungleich Null).

Wie kann ein beliebiger Vektor \mathbf{x} in der neuen Basis dargestellt werden?

1. Basisdarstellung: $\mathbf{x} = x_1 \cdot \mathbf{b}_1 + x_2 \cdot \mathbf{b}_2 + \dots + x_i \cdot \mathbf{b}_i + \dots + x_n \cdot \mathbf{b}_n$ (2)

2. Basisdarstellung: $\mathbf{x} = \bar{x}_1 \cdot \mathbf{b}_1 + \bar{x}_2 \cdot \mathbf{b}_2 + \dots + \bar{x}_i \cdot \mathbf{a} + \dots + \bar{x}_n \cdot \mathbf{b}_n$ (3)

Aus (1) kann \mathbf{b}_i berechnet werden: $\mathbf{b}_i = -\frac{1}{a_i} \cdot (a_1 \cdot \mathbf{b}_1 + a_2 \cdot \mathbf{b}_2 + \dots - \mathbf{a} + \dots + a_n \cdot \mathbf{b}_n)$

Durch Einsetzen von (3) in (2) ergibt sich:

$$\mathbf{x} = x_1 \cdot \mathbf{b}_1 + x_2 \cdot \mathbf{b}_2 + \dots + x_i \cdot \left(-\frac{a_1}{a_i} \cdot \mathbf{b}_1 - \frac{a_2}{a_i} \cdot \mathbf{b}_2 - \dots + \frac{1}{a_i} \cdot \mathbf{a} - \dots - \frac{a_n}{a_i} \cdot \mathbf{b}_n \right) + \dots + x_n \cdot \mathbf{b}_n$$

Durch Koeffizientenvergleich folgt:

$$\bar{x}_1 = x_1 - \frac{x_i}{a_i} \cdot a_1$$

$$\bar{x}_2 = x_2 - \frac{x_i}{a_i} \cdot a_2$$

.....

$$\bar{x}_i = \frac{x_i}{a_i}$$

.....

$$\bar{x}_n = x_n - \frac{x_i}{a_i} \cdot a_n$$

Daraus läßt sich das folgende Schema für die Umrechnung der Koeffizienten bei einer Basistransformation berechnen:

Basis	\mathbf{a}	\mathbf{x}
\mathbf{b}_1	a_1	x_1
\mathbf{b}_2	a_2	x_2
...
\mathbf{b}_i	a_i	x_i
...		
\mathbf{b}_n	a_n	x_n
\mathbf{b}_1	0	\bar{x}_1
...
...
\mathbf{a}	1	\bar{x}_i

...		
...	0	...
\mathbf{b}_n		x_n

$$\mathbf{a} = a_1 \cdot \mathbf{b}_1 + a_2 \cdot \mathbf{b}_2 + \dots + a_i \cdot \mathbf{b}_i + \dots + a_n \cdot \mathbf{b}_n$$

$$\mathbf{x} = x_1 \cdot \mathbf{b}_1 + x_2 \cdot \mathbf{b}_2 + \dots + x_i \cdot \mathbf{b}_i + \dots + x_n \cdot \mathbf{b}_n$$

$$\mathbf{a} = 0 \cdot \mathbf{b}_1 + 0 \cdot \mathbf{b}_2 + \dots + 1 \cdot \mathbf{a} + \dots + 0 \cdot \mathbf{b}_n$$

$$\mathbf{x} = x_1 \cdot \mathbf{b}_1 + x_2 \cdot \mathbf{b}_2 + \dots + x_i \cdot \mathbf{a} + \dots + x_n \cdot \mathbf{b}_n$$

2.3.2.3 Zulässige Basislösungen

2.3.2.3.1 Der Nullpunkt

Da $\mathbf{b} \geq \mathbf{n}$ ist, erfüllt $\mathbf{x} = \mathbf{n}$ die Nichtnegativitätsbedingungen und die einschränkenden Bedingungen. Das führt zu dem Satz:

Das lineare Ungleichungssystem $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$, $\mathbf{b} \geq \mathbf{n}$, $\mathbf{x} \geq \mathbf{n}$ hat stets eine Lösung, nämlich die triviale Lösung $\mathbf{x} = \mathbf{n}$.

Beweis: Es gilt immer $\mathbf{A} \cdot \mathbf{n} \leq \mathbf{b}$, da $\mathbf{b} \geq \mathbf{n}$. Zu der trivialen Lösung gehört im äquivalenten Gleichungssystem $\mathbf{A} \cdot \mathbf{x} + \mathbf{E} \cdot \mathbf{y} = \mathbf{b}$ die Lösung $\mathbf{y} = \mathbf{b}$. Die triviale Lösung $\mathbf{x} = \mathbf{n}$ bestimmt den Nullpunkt.

Der Nullpunkt ist ein Eckpunkt der Lösungsmenge des linearen Ungleichungssystems: $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$, $\mathbf{b} \geq \mathbf{n}$, $\mathbf{x} \geq \mathbf{n}$.

Für den Beweis dieses Satzes ist eine Reihe von Definitionen wichtig:

1. Ein Punkt „E“ einer konvexen Menge heißt Eckpunkt, wenn er nicht als Linearkombination in folgender Form angegeben werden kann: $\mathbf{x}_E = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)$. Das bedeutet: Ein Eckpunkt kann nie Mittelpunkt einer Strecke sein, die selbst zur Menge gehört.
2. Eine Teilmenge des n-dimensionalen Vektorraums heißt konvex, falls für 2 beliebige Punkte der Menge, die durch \mathbf{x}' und \mathbf{x}'' bestimmt sind, auch alle Punkte \mathbf{x} der Verbindungsstrecke zur Menge gehören. Es gilt $\mathbf{x} = k \cdot \mathbf{x}'' + (1 - k) \cdot \mathbf{x}'$ mit $0 \leq k \leq 1$.
3. Eine konvexe Punktmenge heißt unbeschränkt, wenn es einen Vektor \mathbf{v} gibt, so daß alle Punkte, die durch $k \cdot \mathbf{v}$ bestimmt sind, zur Menge gehören. Gibt es einen solchen Vektor nicht, dann heißt die Menge beschränkt.
4. Eine konvexe Punktmenge heißt konvexes Polyeder, wenn die Menge beschränkt ist und nur endlich viele Eckpunkte hat.

Der Nullpunkt wird durch $\mathbf{x} = \mathbf{n}$ bestimmt. Wäre der Nullpunkt kein Eckpunkt, so müßte er sich in folgender Form darstellen lassen: $\mathbf{x} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)$. Diese Bedingung ist aber nur dann erfüllt, wenn $\mathbf{x}_1 = \mathbf{n}$ und $\mathbf{x}_2 = \mathbf{n}$.

Für jedes lineare Ungleichungssystem ist $\mathbf{x} = \mathbf{n}$ eine zulässige Lösung, die außerdem Eckpunkt der Lösungsmenge ist. Zu $\mathbf{x} = \mathbf{n}$ gehört $\mathbf{y} = \mathbf{b}$ als Basis. Der dazu gehörende Wert der Zielfunktion ist $Z_{\max} = \mathbf{c}^T \cdot \mathbf{n} = 0$. Weitere Lösungen können nach dem Verfahren der elementaren Basistransformation gefunden werden.

2.3.2.3.2 Berechnung zulässiger Basislösungen

Eine neue Basislösung erhält man dann, wenn das Hauptelement durch den kleinsten Quotienten aus den Werten der rechten Seite (RS) und den positiven Werten der Hauptspalte (Pivot-) bestimmt wird und positiv ist.

Beweis: Gesucht sind zulässige Basislösungen. Für den Übergang von der r-ten zulässigen Basislösung zu der (r+1)ten zulässigen Basislösung gilt:

 \mathbf{y}_k	\mathbf{y}_l	\mathbf{b}
\mathbf{y}_1	a_{1k}	0	b_1
\mathbf{y}_2	a_{2k}	0	b_2
...			
\mathbf{y}_l	a_{lk}	1	b_l
	a_{mk}	0	b_m
\mathbf{y}_1	0	$-\frac{a_{1k}}{a_{lk}}$	$b_1 - a_{1k} \cdot \frac{b_l}{a_{lk}}$
\mathbf{y}_2	0	$-\frac{a_{2k}}{a_{lk}}$	$b_2 - a_{2k} \cdot \frac{b_l}{a_{lk}}$
....			
\mathbf{y}_k	1	$\frac{1}{a_{lk}}$	$\frac{b_l}{a_{lk}}$
...			
\mathbf{y}_m	0	$-\frac{a_{mk}}{a_{lk}}$	$b_m - a_{mk} \cdot \frac{b_l}{a_{lk}}$

Die Elemente der rechten Seite müssen sein:

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \\ \vdots \\ b_m \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} b_1 - a_{1k} \cdot \frac{b_l}{a_{lk}} \\ b_2 - a_{2k} \cdot \frac{b_l}{a_{lk}} \\ \vdots \\ \frac{b_l}{a_{lk}} \\ \vdots \\ b_m - a_{mk} \cdot \frac{b_l}{a_{lk}} \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Daraus folgt:

1.

$$b_l \geq a_{1k} \cdot \frac{b_l}{a_{lk}}$$

$$b_l \geq a_{2k} \cdot \frac{b_l}{a_{lk}}$$

...

$$\frac{b_l}{a_{lk}} \geq 0$$

...

$$b_l \geq a_{mk} \cdot \frac{b_l}{a_{lk}}$$

2. $\frac{b_l}{a_{lk}}$ muß positiv sein. Da b_l nicht negativ ist, gilt $a_{lk} \geq 0$. (Bei $a_{lk} = 0$ wird im

folgenden Tableau keine neue Basislösung, sondern wieder die vorhergehende Lösung erhalten).

3. Ferner gilt: $\frac{b_l}{a_{lk}} \leq \frac{b_1}{a_{1k}}, \frac{b_l}{a_{lk}} \leq \frac{b_2}{a_{2k}}, \dots, \frac{b_l}{a_{lk}} \leq \frac{b_m}{a_{mk}}$

Es muß daher sein: $\frac{b_l}{a_{lk}} = \min\left(\frac{b_1}{a_{1k}}, \frac{b_2}{a_{2k}}, \dots, \frac{b_m}{a_{mk}}\right)$

2.3.2.3.3 Eckpunkte

1) Jede zulässige Basislösung, die durch Basistransformation aus der trivialen Lösung gewonnen wird, bestimmt einen **Eckpunkt**.

Beweis

$$\mathbf{A} \cdot \mathbf{x} + \mathbf{E} \cdot \mathbf{y} = \mathbf{b}, \mathbf{b} \geq \mathbf{n}, \mathbf{x} \geq \mathbf{n}, \mathbf{y} \geq \mathbf{n}$$

Für $\mathbf{x} = \mathbf{n}$ ergibt sich die Ausgangsbasis: $B_1 = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_m\}$

Sollen r linear unabhängige Spaltenvektoren in die neue Basis aufgenommen

werden, dann führt das zur neuen Basis $B_2 = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_m\}$. Die

Basistransformation wird durchgeführt nach dem Verfahren der allgemeinen

Basistransformation. Das bedeutet: Zerlegung in Blockmatrizen.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \mathbf{E} = \begin{pmatrix} \mathbf{E}_1 & \mathbf{N}_1 \\ \mathbf{N}_2 & \mathbf{E}_2 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$$

$\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{E}_1, \mathbf{N}_1$ haben r Zeilen. $\mathbf{x}_1, \mathbf{y}_1, \mathbf{b}_1$ haben r Komponenten. Zu \mathbf{A}_{11} existiert die inverse Matrix \mathbf{A}_{11}^{-1} , da die r Spaltenvektoren von \mathbf{A}_{11} linear unabhängig sind (Voraussetzung). Aus $\mathbf{A} \cdot \mathbf{x} + \mathbf{E} \cdot \mathbf{y} = \mathbf{b}$ folgt:

$$\begin{aligned} \mathbf{A}_{11} \cdot \mathbf{x}_1 + \mathbf{A}_{12} \cdot \mathbf{x}_2 + \mathbf{E}_1 \cdot \mathbf{y}_1 + \mathbf{N}_1 \cdot \mathbf{y}_2 &= \mathbf{b}_1 \\ \mathbf{A}_{21} \cdot \mathbf{x}_1 + \mathbf{A}_{22} \cdot \mathbf{x}_2 + \mathbf{E}_2 \cdot \mathbf{y}_2 + \mathbf{N}_2 \cdot \mathbf{y}_1 &= \mathbf{b}_2 \end{aligned}$$

Für $\mathbf{x}_1 \cdot \mathbf{n}$ und $\mathbf{x}_2 \cdot \mathbf{n}$ gilt: $\mathbf{y}_1 = \mathbf{b}_1$, $\mathbf{y}_2 = \mathbf{b}_2$ (Ausgangslösung). Durch Aufnahme von $\mathbf{x}_1^T = (x_1, x_2, \dots, x_r)$ in die Basis ergibt sich: $\mathbf{x}_1 + \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \cdot \mathbf{x}_2 + \mathbf{A}_{11}^{-1} \cdot \mathbf{y}_1 = \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$.

Mit $\mathbf{x}_2 = \mathbf{n}$ und $\mathbf{y}_1 = \mathbf{n}$ gilt: $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 \\ \mathbf{n} \end{pmatrix}$

Ist diese Basislösung ein Eckpunkt?

Falls sie zu keinem Eckpunkt führt, dann müssen 2 verschiedene Vektoren $\mathbf{z}_1, \mathbf{z}_2$

existieren. Dann gilt: $\mathbf{x} = \frac{1}{2} \cdot (\mathbf{z}_1 + \mathbf{z}_2)$ mit $\mathbf{z}_1 \geq \mathbf{n}$, $\mathbf{z}_2 \geq \mathbf{n}$ und $\mathbf{A} \cdot \mathbf{z}_1 \leq \mathbf{b}_1$ bzw. $\mathbf{A} \cdot \mathbf{z}_2 \leq \mathbf{b}_2$.

Die Zerlegung von $\mathbf{z}_1 \geq \mathbf{n}$, $\mathbf{z}_2 \geq \mathbf{n}$ mit $\mathbf{z}_1 = \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{pmatrix}$ und $\mathbf{z}_2 = \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{z}_{22} \end{pmatrix}$ führt zu:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 \\ \mathbf{n} \end{pmatrix} = \frac{1}{2} \cdot \left(\begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{pmatrix} + \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{z}_{22} \end{pmatrix} \right)$$

$$(1) \quad \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 = \frac{1}{2} (\mathbf{z}_{11} + \mathbf{z}_{21})$$

$$\mathbf{n} = \frac{1}{2} (\mathbf{z}_{12} + \mathbf{z}_{22})$$

Daraus folgt, da $\mathbf{z}_1 = \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{pmatrix} \geq \mathbf{n}$ und $\mathbf{z}_2 = \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{z}_{22} \end{pmatrix} \geq \mathbf{n}$: $\mathbf{z}_{12} = \mathbf{n}$ und $\mathbf{z}_{22} = \mathbf{n}$. Nach

Voraussetzung (einschränkende Beziehungen) gilt ferner:

$$a) \quad \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{n} \end{pmatrix} + \begin{pmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$$

$$(2) \quad \mathbf{A}_{11} \cdot \mathbf{z}_{11} + \mathbf{y}_{11} = \mathbf{b}_1$$

$$b) \quad \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{n} \end{pmatrix} + \begin{pmatrix} \mathbf{y}_{21} \\ \mathbf{y}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$$

$$(3) \quad \mathbf{A}_{11} \cdot \mathbf{z}_{21} + \mathbf{y}_{21} = \mathbf{b}_1$$

Einsetzen in (1) ergibt: $\mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 = \frac{1}{2} (\mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 - \mathbf{A}_{11}^{-1} \cdot \mathbf{y}_{11} + \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 - \mathbf{A}_{11}^{-1} \cdot \mathbf{y}_{21})$

$$\mathbf{n} = \frac{1}{2} \cdot \mathbf{A}_{11}^{-1} \cdot (\mathbf{y}_{11} + \mathbf{y}_{21})$$

Damit ergibt sich für (2) und (3): $\mathbf{A}_{11} \cdot \mathbf{z}_{11} = \mathbf{b}_1$, $\mathbf{A}_{11} \cdot \mathbf{z}_{21} = \mathbf{b}_1$, d.h.: $\mathbf{z}_1 = \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{12} \end{pmatrix} = \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{n} \end{pmatrix}$,

$\mathbf{z}_2 = \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{z}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{z}_{21} \\ \mathbf{n} \end{pmatrix}$. Es ist stets $\mathbf{z}_1 = \mathbf{z}_2$ und durch \mathbf{x} wird ein Eckpunkt der Lösungsmenge bestimmt.

2) Jedem Eckpunkt aus der Menge der zulässigen Lösungen eines linearen Gleichungssystems $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$, $\mathbf{b} \geq \mathbf{n}$, $\mathbf{x} \geq \mathbf{n}$ ist eine Basislösung zugeordnet.

3) Die Menge der Basislösungen eines linearen Ungleichungssystems ist endlich.

Beweis: $(\mathbf{A}, \mathbf{E}) = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m)$

Die Basislösung bildet jeweils eine Teilmenge aus diesen Vektoren. Von $(n+m)$ Vektoren können höchstens m Vektoren eine Basis bilden, d.h.: Es gibt maximal

$\binom{n+m}{m}$ verschieden Basislösungen.

4) Die Anzahl der Eckpunkte einer beschränkt konvexen Lösungsmenge des linearen Ungleichungssystems ist endlich:

Beweis: a) Jeder zulässigen Basislösung ist ein Eckpunkt zugeordnet

b) Jedem Eckpunkt ist eine Basislösung zugeordnet

c) Die Anzahl der Basislösungen ist endlich, d.h. auch die Anzahl der Eckpunkte ist endlich.

2.3.2.3.4 Das Maximum

1) Eine lineare Funktion $Z = \mathbf{c}^T \cdot \mathbf{x} = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n$ (, die über einem konvexen Polyeder definiert ist,) erreicht ihr Maximum in mindestens einem Eckpunkt (, sofern eine zulässige Lösung überhaupt existiert).

2) Erreicht die Zielfunktion für die Eckpunkte $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q$ den gleichen optimalen Wert, so ist jede konvexe Linearkombination ebenfalls eine zulässige optimale Lösung.

Beweis: $Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}_1 = \mathbf{c}^T \cdot \mathbf{x}_2 = \dots = \mathbf{c}^T \cdot \mathbf{x}_q$. Für einen beliebigen Punkt der konvexen Menge, die durch $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q$ bestimmt ist, gilt: $\mathbf{x} = \sum k_i \cdot \mathbf{x}_i$ mit $\sum k_i = 1$ und $k_i \geq 0$. Der zugehörige Wert der Zielfunktion ist:

$$Z = \mathbf{c}^T \cdot \mathbf{x} = \mathbf{c}^T \cdot \sum_{i=1}^q k_i \cdot \mathbf{x}_i = \sum_{i=1}^q k_i \cdot \mathbf{c}^T \cdot \mathbf{x}_i = \sum_{i=1}^q k_i \cdot Z_{\max} = Z_{\max} \cdot \sum_{i=1}^q k_i = Z_{\max}$$

2.3.2.4 Bestimmen der optimalen Lösung

Das Bestimmen der optimalen Lösung erfolgt nach dem **Simplexkriterium**:

Ist \mathbf{x} eine zulässige Basislösung und hat die Zielfunktion in der zugehörigen Basisdarstellung keine positiven Koeffizienten, so ist die optimale Lösung erreicht.

BV	\mathbf{x}_1^T	\mathbf{x}_2^T	\mathbf{y}_1^T	\mathbf{y}_2^T	b_i
\mathbf{y}_1	\mathbf{A}_{11}	\mathbf{A}_{12}	\mathbf{E}_1	\mathbf{N}_1	\mathbf{b}_1
\mathbf{y}_2	\mathbf{A}_{21}	\mathbf{A}_{22}	\mathbf{N}_2	\mathbf{E}_2	\mathbf{b}_2
\mathbf{Z}	\mathbf{c}_1^T	\mathbf{c}_2^T	\mathbf{n}^T	\mathbf{n}^T	
\mathbf{x}_1	\mathbf{E}_1	$\mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	\mathbf{A}_{11}^{-1}	\mathbf{N}_1	$\mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$
\mathbf{y}_2	\mathbf{N}_2	$\mathbf{A}_{22} - \mathbf{A}_{21} \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	$-\mathbf{A}_{21} \cdot \mathbf{A}_{11}^{-1}$	\mathbf{E}_2	$\mathbf{b}_2 - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{b}_1$
\mathbf{Z}_{opt}	\mathbf{n}^T	$\mathbf{c}_2^T - \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	$-\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1}$	\mathbf{n}^T	$\mathbf{c}_1^T \mathbf{A}_{11}^{-1} \mathbf{b}_1$

Aus der Tabelle folgt:

$$1. \mathbf{c}_2^T - \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \leq \mathbf{n}^T$$

$$2. -\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \leq \mathbf{n}^T$$

Der Wert der Zielfunktion beträgt dann: $\mathbf{Z}_{opt} = \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$

Beweis: $\mathbf{x}_Z = \begin{pmatrix} \mathbf{x}_{1Z} \\ \mathbf{x}_{2Z} \end{pmatrix}$ sei eine zulässige, beliebige Basislösung.

$$\mathbf{Z}_{opt} = \mathbf{c} \cdot \mathbf{x} = \begin{pmatrix} \mathbf{c}_1^T, \mathbf{c}_2^T \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}_{1Z} \\ \mathbf{x}_{2Z} \end{pmatrix} = \mathbf{c}_1^T \cdot \mathbf{x}_{1Z} + \mathbf{c}_2^T \cdot \mathbf{x}_{2Z}$$

Für eine zulässige Basislösung mit nicht positiven Zielfunktionskoeffizienten gilt:

$$1. \mathbf{c}_2^T - \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \leq \mathbf{n}^T$$

$$2. -\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \leq \mathbf{n}^T$$

Der Wert der Zielfunktion ergibt sich aus:

$$\mathbf{Z}_{opt} = \mathbf{c} \cdot \mathbf{x} = \mathbf{c}_1^T \cdot \mathbf{x}_{1Z} + \mathbf{c}_2^T \cdot \mathbf{x}_{2Z}$$

$$\begin{aligned}
Z_{opt} &\leq \mathbf{c}_1^T \cdot \mathbf{x}_{1Z} + \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \cdot \mathbf{x}_{2Z} \\
Z_{opt} &\leq \mathbf{c}_1^T \cdot \mathbf{A}_{11} \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{x}_{1Z} + \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \cdot \mathbf{x}_{2Z} \\
Z_{opt} &\leq \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot (\mathbf{A}_{11} \cdot \mathbf{x}_{1Z} + \mathbf{A}_{12} \cdot \mathbf{x}_{2Z})
\end{aligned}$$

Da $\mathbf{x}_Z = \begin{pmatrix} \mathbf{x}_{1Z} \\ \mathbf{x}_{2Z} \end{pmatrix}$ eine zulässige Lösung ist, gilt:

$$\mathbf{A} \cdot \mathbf{x}_Z = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}_{1Z} \\ \mathbf{x}_{2Z} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} \cdot \mathbf{x}_{1Z} + \mathbf{A}_{12} \cdot \mathbf{x}_{2Z} \\ \mathbf{A}_{21} \cdot \mathbf{x}_{1Z} + \mathbf{A}_{22} \cdot \mathbf{x}_{2Z} \end{pmatrix} \leq \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$$

Daraus folgt: $\mathbf{A}_{11} \cdot \mathbf{x}_{1Z} + \mathbf{A}_{12} \cdot \mathbf{x}_{2Z} \leq \mathbf{b}_1$

$$Z_{opt} \leq \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot (\mathbf{A}_{11} \cdot \mathbf{x}_{1Z} + \mathbf{A}_{12} \cdot \mathbf{x}_{2Z})$$

$$Z_{opt} \leq \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$$

1. Der Wert der Zielfunktion ist für eine beliebige Basislösung höchstens so groß wie der Wert der Zielfunktion, bei der die Koeffizienten in der zugehörigen Basisdarstellung nicht positiv sind.
2. Die Zielfunktion hat dann den maximalen Wert erreicht, wenn nach dem Verfahren der elementaren Basistransformation alle Koeffizienten der Zielfunktion Null oder negativ sind.

2.3.3 Das Dualitätsproblem

2.3.3.1 Ein einführendes Beispiel

Konkurrenzproblem: Ein Produzent hat 2 Betriebe in den Orten O_1 und O_2 . Von diesen Betrieben sollen 4 Großhändler H_1, H_2, H_3, H_4 beliefert werden. In der folgenden Tabelle sind die Daten des Problems beschrieben:

	Transportkosten O_1 (DM/t)	Transportkosten O_2 (DM/t)	Mindestbedarf (t)
H_1	12	14	600
H_2	7	11	500
H_3	13	6	400
H_4	11	18	800
Lagerbestand (t)	1200	1300	

Der Transport soll so durchgeführt werden, daß die gesamten Transportkosten möglichst gering sind. Vereinbarung: x_{ik} sind die am Ort O_i an den Großhändler H_k gelieferten Mengen ($i = 1, 2; k = 1, 2, 3, 4$).

Mathematisches Modell

$$(1) \quad \begin{aligned} x_{11} &\geq 0, & x_{12} &\geq 0, & x_{13} &\geq 0, & x_{14} &\geq 0, \\ x_{21} &\geq 0, & x_{22} &\geq 0, & x_{23} &\geq 0, & x_{24} &\geq 0 \end{aligned}$$

$$(2) \quad \begin{aligned} x_{11} + x_{21} &\geq 600 \\ x_{12} + x_{22} &\geq 500 \end{aligned}$$

$$\begin{aligned}
x_{13} + x_{23} &\geq 400 \\
x_{14} + x_{24} &\geq 800 \\
x_{11} + x_{12} + x_{13} + x_{14} &\leq 1200 \\
x_{21} + x_{22} + x_{23} + x_{24} &\leq 1300
\end{aligned}$$

$$(3) \quad Z_{\min} = 12x_{11} + 7x_{12} + 13x_{13} + 11x_{14} + 14x_{21} + 11x_{22} + 6x_{23} + 18x_{24}$$

Standardisierung

$$\begin{aligned}
(2) \quad x_{11} + x_{21} &\geq 600 \\
x_{12} + x_{22} &\geq 500 \\
x_{13} + x_{23} &\geq 400 \\
x_{14} + x_{24} &\geq 800 \\
-x_{11} - x_{12} - x_{13} - x_{14} &\geq -1200 \\
-x_{21} - x_{22} - x_{23} - x_{24} &\geq -1300
\end{aligned}$$

Dasselbe Problem aus anderer Sicht betrachtet führt zu dem folgenden Transportproblem: Ein Transportunternehmen übernimmt den Transport des Zements. Er kauft den Zement beim Hersteller und verkauft ihn an die Großhändler. Die Verkaufspreise an H_1, H_2, H_3, H_4 sind y_1, y_2, y_3, y_4 . Die Einkaufspreise in O_1 und O_2 sind y_5, y_6 . Damit der Zement nicht verteuert wird und der Transportunternehmer konkurrenzfähig bleibt, darf die Preisdifferenz von bspw. $(y_1 - y_5)$ höchstens so hoch sein wie die ursprünglichen Transportkosten von O_1 nach H_1 . Entsprechendes gilt für die anderen Preisdifferenzen. Der Transportunternehmer will außerdem einen möglichst hohen Gewinn machen.

Mathematisches Modell

$$(1) \quad y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_4 \geq 0, y_5 \geq 0, y_6 \geq 0$$

$$\begin{aligned}
(2) \quad y_1 - y_5 &\leq 12 \\
y_2 - y_5 &\leq 7 \\
y_3 - y_5 &\leq 13 \\
y_4 - y_5 &\leq 11 \\
y_1 - y_6 &\leq 14 \\
y_2 - y_6 &\leq 11 \\
y_3 - y_6 &\leq 6 \\
y_4 - y_6 &\leq 8
\end{aligned}$$

$$(3) \quad Z_{\max} = 600y_1 + 500y_2 + 400y_3 + 800y_4 - 1200y_5 - 1300y_6$$

Aus dem ursprünglichen Optimierungsproblem wurde ein neues lineares Optimierungsproblem (Transportproblem) gewonnen. Aus dem Minimierungsproblem wurde ein Maximumproblem. In beiden Systemen treten diesselben Zahlen auf. Sie sind nur anders angeordnet. Beide Probleme stehen offenbar in einem engen Zusammenhang.: Das zweite Problem ist das duale Problem (und umgekehrt).

2.3.3.2 Das Dualitätstheorem

2.3.3.2.1 Das duale lineare Optimierungsprobleme

Definition

a) Ist ein lineares Optimierungsproblem in der Form

(1) $\mathbf{x} \geq \mathbf{n}$, (2) $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$, (3) $Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}$ gegeben, dann ist das duale Optimierungsproblem (1) $\mathbf{y}^T \geq \mathbf{n}$, (2) $\mathbf{y}^T \cdot \mathbf{A} \geq \mathbf{c}^T$, (3) $Z_{\min} = \mathbf{y}^T \cdot \mathbf{b}$

b) Ist ein lineares Optimierungsproblem in der Form

(1) $\mathbf{x} \geq \mathbf{n}$, (2) $\mathbf{A} \cdot \mathbf{x} \geq \mathbf{b}$, (3) $Z_{\min} = \mathbf{c}^T \cdot \mathbf{x}$ gegeben, dann ist das duale Optimierungsproblem (1) $\mathbf{y}^T \geq \mathbf{n}$, (2) $\mathbf{y}^T \cdot \mathbf{A} \leq \mathbf{c}^T$, (3) $Z_{\max} = \mathbf{y}^T \cdot \mathbf{b}$

Jedem Maximumproblem ist also genau ein Minimumproblem und jedem Minimumproblem genau ein Maximumproblem als duales Problem zugeordnet.

In den folgenden Tabellen sind zwei duale Probleme gegenübergestellt:

x_1	x_2	...	x_n	\leq	
a_{11}	a_{12}		a_{1n}		b_1
a_{21}	a_{22}		a_{2n}		b_2
...
a_{m1}	a_{m2}		a_{mn}		b_m
c_1	c_2		c_n		Z_{\max}

y_1	y_2 y_m	\geq	
a_{11}	a_{21}		a_{m1}		c_1
a_{12}	a_{22}		a_{m2}		c_2
...
a_{1n}	a_{2n}		a_{mn}		c_n
b_1	b_2		b_m		Z_{\min}

Bsp.: Primales Optimierungsproblem

(1) $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$

(2) $4x_1 - 2x_2 - 6x_3 \leq 17$
 $2x_1 + 7x_2 + 4x_3 \leq 13$
 $8x_1 - 2x_2 + 2x_3 \leq 11$
 $4x_1 + 7x_3 \leq 12$

(3) $Z_{\max} = 5x_1 + 7x_2 + 9x_3$

Duales Optimierungsproblem

$y_1 \geq 0$, $y_2 \geq 0$, $y_3 \geq 0$,
 $y_4 \geq 0$

$4y_1 + 2y_2 + 8y_3 + 4y_4 \geq 5$
 $-2y_1 + 7y_2 - 2y_3 + \quad \geq 7$
 $-6y_1 + 4y_2 + 2y_3 + 7y_4 \geq 9$

$Z_{\min} = 17y_1 + 13y_2 + 11y_3 + 12y_4$

2.3.3.2.2 Das Optimalitätskriterium

Gegeben ist das primale Optimierungsproblem

$$(1) \mathbf{x} \geq \mathbf{n}, (2) \mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}, (3) Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}$$

Das zugehörige duale Optimierungsproblem ist

$$(1) \mathbf{y}^T \geq \mathbf{n}, (2) \mathbf{y}^T \cdot \mathbf{A} \geq \mathbf{c}^T, (3) Z_{\min} = \mathbf{y}^T \cdot \mathbf{b}$$

M_p ist die Menge der zulässigen Lösungen des primären Problems. M_d ist die Menge der zulässigen Lösungen des dualen Problems.

(a) Es gilt der folgende Satz: Ist $\mathbf{x} \in M_p$ und $\mathbf{y} \in M_d$, so gilt stets $\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{b}$

Beweis: Multiplikation von \mathbf{y}^T mit $\mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}$: $\mathbf{y}^T \cdot \mathbf{A} \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{b}$

Multiplikation von $\mathbf{y}^T \cdot \mathbf{A}$ mit \mathbf{x} : $\mathbf{y}^T \cdot \mathbf{A} \cdot \mathbf{x} \geq \mathbf{c}^T \cdot \mathbf{x}$.

Daraus folgt: $\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{A} \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{b}$, $\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{y}^T \cdot \mathbf{b}$

(b) Weiterhin gilt der Satz: Falls das lineare Optimierungsproblem

$$(1) \mathbf{x} \geq \mathbf{n}, (2) \mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}, (3) Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}$$

und das dazu duale Problem

$$(1) \mathbf{y}^T \geq \mathbf{n}, (2) \mathbf{y}^T \cdot \mathbf{A} \geq \mathbf{c}^T, (3) Z_{\min} = \mathbf{y}^T \cdot \mathbf{b}$$

zulässige Lösungen haben und wenn außerdem $\mathbf{c}^T \cdot \mathbf{x}_E \leq \mathbf{y}_E^T \cdot \mathbf{b}$ ist, dann ist \mathbf{x}_E die optimale Lösung der Maximumaufgabe (Optimalitätskriterium) und \mathbf{y}_E die optimale Lösung der Minimumaufgabe.

Beweis: Aus (a) gilt: $\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{y}_E^T \cdot \mathbf{b}$ für $\mathbf{x} \in M_p$, $\mathbf{y}_E^T \cdot \mathbf{b} \geq \mathbf{c}^T \cdot \mathbf{x}_E$ für $\mathbf{y} \in M_d$. Nach

Voraussetzung ist: $\mathbf{c}^T \cdot \mathbf{x}_E \leq \mathbf{y}_E^T \cdot \mathbf{b}$, d.h.:

$$\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{y}_E^T \cdot \mathbf{b} = \mathbf{c}^T \cdot \mathbf{x}_E$$

$$\mathbf{y}^T \cdot \mathbf{b} \geq \mathbf{c}^T \cdot \mathbf{x}_E = \mathbf{y}_E^T \cdot \mathbf{b}$$

Also ist:

$$\mathbf{c}^T \cdot \mathbf{x} \leq \mathbf{c}^T \cdot \mathbf{x}_E$$

$$\mathbf{y}^T \cdot \mathbf{b} \geq \mathbf{y}_E^T \cdot \mathbf{b}$$

Für alle $\mathbf{x} \in M_p$ ist der Wert der Zielfunktion stets kleiner gleich $\mathbf{c}^T \cdot \mathbf{x}_E$, d.h. $\mathbf{c}^T \cdot \mathbf{x}_E$ ist die optimale Lösung des Maximumproblems.

Für alle $\mathbf{y} \in M_d$ ist der Wert der Zielfunktion größer oder gleich $\mathbf{y}_E^T \cdot \mathbf{b}$, d.h. $\mathbf{y}_E^T \cdot \mathbf{b}$ ist die optimale Lösung des Minimumproblems.

\mathbf{x}_E und \mathbf{y}_E sind die optimalen Lösungen.

2.3.3.2.3 Der Dualitätssatz

Die Umkehrung des Optimalitätskriteriums wird als Dualitätssatz bezeichnet, der folgendermaßen lautet:

Existieren für die zwei zueinander dualen Optimierungsprobleme

$$(1) \mathbf{x} \geq \mathbf{n}, (2) \mathbf{A} \cdot \mathbf{x} \leq \mathbf{b}, (3) Z_{\max} = \mathbf{c}^T \cdot \mathbf{x}$$

und

$$(1) \mathbf{y}^T \geq \mathbf{n}, (2) \mathbf{y}^T \cdot \mathbf{A} \geq \mathbf{c}^T, (3) Z_{\min} = \mathbf{y}^T \cdot \mathbf{b}$$

optimale Lösungen, so sind die Werte der Zielfunktion von beiden Systemen gleich.

BV	\mathbf{x}_1^T	\mathbf{x}_2^T	\mathbf{y}_1^T	\mathbf{y}_2^T	\mathbf{b}_j
\mathbf{y}_1	\mathbf{A}_{11}	\mathbf{A}_{12}	\mathbf{E}_1	\mathbf{N}_1	\mathbf{b}_1
\mathbf{y}_2	\mathbf{A}_{21}	\mathbf{A}_{22}	\mathbf{N}_2	\mathbf{E}_2	\mathbf{b}_2
\mathbf{Z}	\mathbf{c}_1^T	\mathbf{c}_2^T	\mathbf{n}^T	\mathbf{n}^T	
\mathbf{x}_1	\mathbf{E}_1	$\mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	\mathbf{A}_{11}^{-1}	\mathbf{N}_1	$\mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$
\mathbf{y}_2	\mathbf{N}_2	$\mathbf{A}_{22} - \mathbf{A}_{21} \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	$-\mathbf{A}_{21} \cdot \mathbf{A}_{11}^{-1}$	\mathbf{E}_2	$\mathbf{b}_2 - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{b}_1$
\mathbf{Z}_{opt}	\mathbf{n}^T	$\mathbf{c}_2^T - \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}$	$-\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1}$	\mathbf{n}^T	$\mathbf{c}_1^T \mathbf{A}_{11}^{-1} \mathbf{b}_1$

Die optimale Lösung des Maximumsproblem ist: $\mathbf{x}_E = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1 \\ \mathbf{n} \end{pmatrix}$ mit

$Z_{\text{opt}} = \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$. Die Tabelle gibt auch die optimale Lösung für das zugehörige, duale Problem an: $\mathbf{y}_E^T = (\mathbf{y}_1^T, \mathbf{y}_2^T) = \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1$. Die dazu gehörenden Werte erscheinen in der letzten Zeile mit negativem Vorzeichen. Es gilt:

$$(1) \mathbf{y}_E^T \geq \mathbf{n}, \text{ da } -\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \leq \mathbf{n}^T$$

$$(2) \mathbf{y}_E^T \cdot \mathbf{A} = (\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1}, \mathbf{n}^T) \cdot \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = (\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{11} + \mathbf{n}^T \cdot \mathbf{A}_{21}, \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} + \mathbf{n}^T \cdot \mathbf{A}_{22})$$

$$= (\mathbf{c}_1^T, \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12}) \geq (\mathbf{c}_1^T, \mathbf{c}_2^T) = \mathbf{c}^T$$

Die vorstehende Beziehungen folgen aus:

$$-\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \leq \mathbf{n}^T$$

$$\mathbf{c}_2^T - \mathbf{c}_1 \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{A}_{12} \leq \mathbf{n}^T$$

Der Vektor \mathbf{y}_E erfüllt die einschränkenden Bedingungen und die Nichtnegativitätsbedingungen, d.h. er ist eine zulässige Lösung des Minimumproblems. Der

zugehörige Wert der Zielfunktion ist: $\mathbf{y}_E^T \cdot \mathbf{b} = (\mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1}, \mathbf{n}^T) \cdot \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{c}_1^T \cdot \mathbf{A}_{11}^{-1} \cdot \mathbf{b}_1.$

Der Wert erscheint in der Tabelle mit negativem Vorzeichen. Er entspricht dem optimalen Wert der Zielfunktion für die Maximumaufgabe. Das bedeutet: \mathbf{y}_E ist nach dem Optimalitätskriterium die optimale Lösung.

2.4 Anwendung des Simplexverfahrens

2.4.1 Das verkürzte Simplex-Tableau

2.4.1.1 Ein einführendes Beispiel

1) Mathematisches Modell

(1) $x_1 \geq 0, x_2 \geq 0$

(2)
$$\begin{aligned} 3x_1 + x_2 &\leq 36 \\ x_1 + x_2 &\leq 16 \\ x_2 &\leq 12 \end{aligned}$$

(3) $Z_{\max} = 4x_1 + 3x_2$

2) Simplex-Verfahren

a) Darstellung mit expliziter Einheitsmatrix

BV	x_1	x_2	y_1	y_2	y_3	b_i	q_i
y_1	3	1	1	0	0	36	
y_2	1	1	0	1	0	16	
y_3	0	1	0	0	1	12	
Z							
x_1							
y_2							
y_3							
Z							
x_1							
x_2							
y_3							
Z_{\max}							

b) Verkürzte Darstellung

(1) $x_1 \geq 0, x_2 \geq 0$

(2) $3x_1 + x_2 \leq 36$

$x_1 + x_2 \leq 16$

$x_2 \leq 12$

(3) $Z_{\max} - 4x_1 - 3x_2 = 0$

	x_1	x_2	RS	
Z_{\max}	-4	-3	0	
Y_1	3	1	36	
Y_2	1	1	16	
Y_3	0	1	12	

2.4.1.2 Allgemeine Darstellung

Bezeichnungen

Schlupfvariable: y_i

Strukturvariable: x_j

Tableau

- Die Zielfunktion wird in dem Tableau als erste Zeile mit dem Zeilenindex $i = 0$ dargestellt
- Die rechte Seite (RS) der Gleichungen erhält den Spaltenindex $j = 0$
- In der Kopfzeile stehen die Nichtbasisvariablen, Sie erhalten nach den getroffenen Vereinbarungen den Wert 0 zugeordnet
- Die Basisvariablen werden in der Kopfspalte ausgeführt

	x_1	x_j	x_n	RS
Z	a_{01}	a_{0j}	a_{0n}	a_{00}
y_1	a_{11}	a_{1j}	a_{1n}	a_{10}
....	
....	
....	
y_i	a_{i1}	a_{ij}	a_{in}	a_{i0}
....	
....	
....	
y_m	a_{m1}	a_{mj}		a_{m0}

2.4.1.3 Steepest-Unit Ascent

Fallbeispiel aus der Produktion: Die Produkte A und B werden auf der Maschinengruppe 1, 2 und 3 hergestellt. Produkt A ist ein Endprodukt, das unmittelbar auf dem Markt abgesetzt werden kann. Produkt B ist ein Vorprodukt, das im Unternehmen weiterverarbeitet wird. Das Vorprodukt B kann selbst gefertigt oder auch von einem Zulieferanten bezogen werden. Die Bearbeitungszeiten je Mengeneinheit der Produkte A und B, die wöchentliche Kapazität der Maschinengruppe ist in der folgenden Tabelle angegeben:

Maschinengruppe	Endprodukt A Bearbeitungszeit (Std/ME)	Vorprodukt B Bearbeitungszeit (Std/ME)	wöchentliche Kapazität (Std)
1	1	2	160
2	1	1	120
3	3	1	300

Der Verkaufspreis des Endprodukts A beträgt 1000.- DM je ME⁷.

Bei bekannten Stückkosten (400DM/ME) ergibt sich zur Deckung der Fixkosten (Deckungsbeitrag) ein Beitrag von 600 DM/ME. Im Produktionsablauf werden wöchentlich 60 ME von B benötigt. Der Einkaufspreis von Zulieferanten für B beträgt 1100.- DM/ME. Die variablen Stückkosten bei Eigenfertigung von B betragen 400 DM/ME.

- 1) Welche Mengen des Produkts A können hergestellt werden?
- 2) Welche Mengen von B sollen selbst gefertigt werden bzw. bezogen werden?

Es soll das Ziel verfolgt werden, den Deckungsbeitrag in diesem Produktionsbereich zu maximieren.

Mathematisches Modell

- (1) $x_1 \geq 0, x_2 \geq 0$
- (2) $x_1 + 2x_2 \leq 160$
 $x_1 + x_2 \leq 120$
 $3x_1 + x_2 \leq 300$
 $x_2 \leq 60$
- (3) $Z_{\max} = 600x_1 - 400x_2 - 1100(60 - x_2)$

Der Deckungsbeitrag setzt sich zusammen aus dem Deckungsbeitrag für A vermindert um die variablen Kosten von B und dem zusätzlichen Fremdbezug von B.

$$Z_{\max} = 600x_1 + 700x_2 - 66000$$

$$Z_{\max} - 600x_1 - 700x_2 = -66000$$

Bemerkungen zur Anwendung des Simplex-Verfahrens

1. Zur Bestimmung der optimalen Lösung des Problems müssen die Eckpunkte des zulässigen Lösungsbereichs untersucht werden.
2. In einem Modell mit n Strukturvariablen werden die Eckpunkte aus den m Restriktionen gebildet
3. Durch geschicktes Tauschen von Nichtbasisvariablen und Basisvariablen werden die Ecken des Lösungsraums so durchlaufen, daß sich mit jedem Schritt eine Verbesserung der Zielfunktion ergibt.
4. Am Vorzeichen der Zielfunktionskoeffizienten erkennt man, ob beim Tausch der Zielfunktionswert vergrößert werden kann.

Lösung

⁷ ME = Mengeneinheit

	x_1	x_2	RS	
Z	-600	-700	-66000	
y_1	1	2	160	
y_2	1	1	120	
y_3	3	1	300	
y_4	0	1	60	

			RS	
Z				

			RS	
Z				

			RS	
Z				

			RS	
Z				

Interpretation der Lösung

Mit $x_1 = 80$ ME des Produkts A und $x_2 = 40$ ME des Produkts B wird ein Deckungsbeitrag von 10000 GE erwirtschaftet.

Iterationsschritte

(1) Höchste Vergrößerung des Zielfunktionswertes.

Der Variablen mit dem absolut größten negativen Koeffizienten in der Zielfunktion wird ein Wert zugewiesen (**Steepest-Unit-Ascent**)

(2) Beim Tausch von Nichtbasisvariablen und Basisvariablen darf keine Restriktion verletzt werden, d.h.: Keine Basisvariable darf negativ werden.

Die engste Restriktion ist gegeben durch: Den kleinsten Quotienten aus der rechten Seite (RS) und dem positiven Koeffizienten der schon festgelegten Pivot-Spalte.

Damit ist auch die Pivot-Zeile bestimmt.

(3) Umrechnung des Tableaus auf den neuen Eckpunkt (identisch mit den Umrechnungen, die bei jedem Schritt des Gauss-Jordan-Algorithmus oder bei der Inversion von Matrizen angewendet werden)

Pivot-Element: $a_{rs}^* = \frac{1}{a_{rs}}$

Pivot-Zeile: $a_{rj}^* = \frac{a_{rj}}{a_{rs}}$

Pivot-Spalte: $a_{is}^* = -\frac{a_{is}}{a_{rs}}$

Übrige Elemente: $a_{ij}^* = a_{ij} - a_{is} \cdot \frac{a_{rj}}{a_{rs}}$

(4) Prüfung, ob eine weitere Verbesserung (negativer Zielfunktionskoeffizient?) möglich ist.

Wenn ja, Wiederholung der Arbeitsschritte **(1)** bis **(3)**.

Wenn nein, ist die optimale Lösung gefunden.

2.4.1.4 Die "Greatest Change"-Version

Bemerkungen

- 1) Bei der "Steepest Unit Ascent"-Version kommt diejenige Basisvariable in die Basis, bei der der steilste Anstieg (je Einheit) zu realisieren ist.
- 2) Zusätzlich ist aber entscheidend: Wie stark wächst die Nichtbasisvariable? Diesen Wert erhält man aus dem kleinsten nichtnegativen Quotienten mit dem die Pivot-Zeile festgelegt wird.

Verfahrensschritte zum Greatest-Change

- (1) Für jede Spalte mit negativem Zielfunktionskoeffizienten wird der jeweils kleinste Quotient von $\frac{a_{i0}}{a_{ij}}$ ermittelt und in der Zeile B eingetragen. Die Zeile r enthält die Nummer der Pivot-Zeile
- (2) Der kleinste Quotient wird mit dem Zielfunktionskoeffizienten multipliziert. Das Ergebnis wird in die Zeile ΔZ eingetragen. Das Produkt bestimmt die Änderung der Zielfunktion.
- (3) Wahl der Pivot-Spalte: Die absolut größte Zahl der Zeile ΔZ bestimmt die Pivot-Spalte.
- (4) Das neue Tableau wird nach der bekannten Verfahrensweise umgerechnet.
- (5) Das Verfahren wird solange fortgesetzt bis die optimale Lösung gefunden wurde.

Hinweis: Bei der „Greatest Change“-Methode gelangt man meistens (nicht immer) mit weniger Iterationen (speziell beim manuellen Lösen kleiner Probleme) zur Optimallösung als mit der „Steepest Unit Ascent“-Version.

Beispiele:

1.

	x_1	x_2	RS	
Z	-600	-700	-66000	-600
Y_1	1	2	160	1
Y_2	1	1	120	1
Y_3	3	1	300	-
Y_4	0	1	60	0
r	3	4		
B	100	60		
ΔZ	-60000	-42000		
	-	1/3	100	

			RS	
Z				

			RS	
Z				

			RS	
Z				

2. Wie lautet die optimale Lösung für das folgende Problem der linearen Planungsrechnung?

(1) $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0$

(2) $15x_1 + 6x_2 + 8x_3 + 11x_4 \leq 120$

$x_1 + x_2 + 4x_3 + 6x_4 \leq 32$

$7x_1 + 3x_2 + 6x_3 + 5x_4 \leq 70$

$26x_1 + 9x_2 + 14x_3 + 22x_4 \leq 210$

(3) $Z_{\max} - 60x_1 - 60x_2 - 70x_3 - 80x_4 = 0$

Operations Research

	x_1	x_2	x_3	x_4	RS	
Z_{\max}						
y_1						
y_2						
y_3						
y_4						
r						
B						
ΔZ						

					RS
Z_{\max}					

2.4.1.5 Sonderfälle

2.4.1.5.1 Entartung

a) Duale Entartung

Mehrere Spalten zeigen einen gleich großen Zielfunktionskoeffizienten.

b) Primale Entartung

Bei der Bestimmung der Pivot-Zeile findet man mehrere Zeilen, die den gleichen Quotienten $\frac{a_{i0}}{a_{is}}$ aufweisen.

Hinweis:

1. Hier besteht die Gefahr in einen Zyklus zu geraten, der immer auf das gleiche Tableau zurückführt.
2. Für EDV-Diagramme verwendet man in solchen Fällen Zufallszahlen. Durch die „zufällige Auswahl“ vermeidet man, daß man auf unbegrenzte Zeit in „Kreisen“ gerät.
3. Geometrisch ist der Fall der primalen Entartung als überbestimmter Punkt im n-dimensionalen Raum zu interpretieren, d.h.: Ein Punkt durch den mehr als n Begrenzungen hindurchgehen.

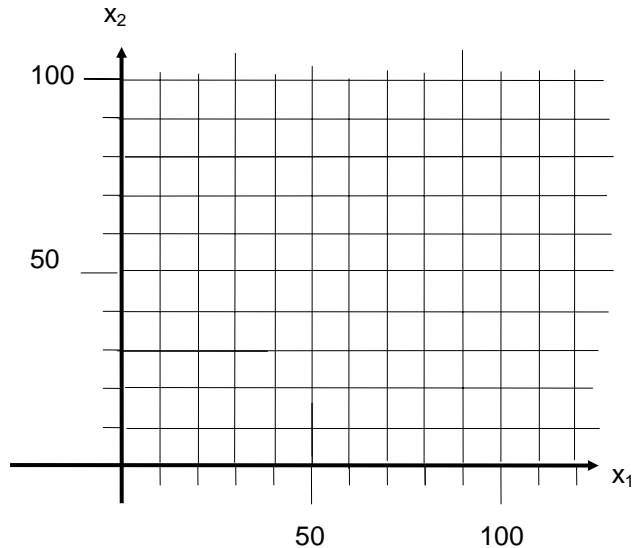
2.4.1.5.2 Unbegrenzte Lösungen

Folgendes Problem ist zu lösen:

	x_1	x_2	RS	
Z	-50	-80	0	
y_1	-3	1	20	
y_2	1	-2	20	
y_3	3	-1	120	

			RS	
Z				
x_2				
y_2				
y_3				

Eine Pivot-Spalte lässt sich für den nächsten Iterationsschritt bestimmen. Jedoch lässt sich keine Pivot-Zeile bestimmen, weil keine nichtnegativen Quotienten $\frac{a_{i0}}{a_{is}}$ auftreten. Es liegt der Sonderfall einer unbegrenzten Lösung vor.



2.4.1.6 Das Simplexverfahren in Java

Der Steepest-Unit-Ascent-Algorithmus ist in dem folgenden Java-Programm implementiert:

```
import java.text.*;

class SimplexException extends Exception
{
    public SimplexException() { super(); }
    public SimplexException(String meldung) { super(meldung); }
}
class SimplexNoSolutionException extends SimplexException
{
    public SimplexNoSolutionException() { super(); }
    public SimplexNoSolutionException(String meldung) { super(meldung); }
}
class SimplexFinishedException extends SimplexException
{
    public SimplexFinishedException() { super(); }
    public SimplexFinishedException(String meldung) { super(meldung); }
}
public class Simplex
{
    private static int m; // Anzahl der Ungleichungen
    private static int n; // Anzahl der Strukturvariablen
    private static int [] bv, nbv;
    // Schlupfv. = -1, Strukturvar. = 0 .. n-1
    private static double[][] a; // Koeffizienten
    private static DecimalFormat df = new DecimalFormat("####0.00");
    private java.util.Random random;
    public Simplex(int m, int n)
    {
        super();
        this.m = m;
        this.n = n;
    }
}
```

```

    bv = new int[m];
    nbv = new int[n];
    a = new double[m + 1][n + 1];
    random = new java.util.Random();
}
public double[][] getCoefficients()
{
    return a;
}
public double solve(double [] x) throws SimplexNoSolutionException
{
    // zu Beginn: Schlupfvariable sind Basisvariable
    for (int i = 0; i < m; i++) bv[i] = -1;
    // und Strukturvariable sind Nicht-Basisvariable
    for (int j = 0; j < n; j++) nbv[j] = j;
    synchronized(a)
    {
        try
        {
            while (true)
            {
                // Pivot-Spalte bestimmen
                int s = pivotColumn();
                // Pivot-Zeile bestimmen
                int r = pivotRow(s);
                // Koeffizienten umrechnen
                calcCoefficients(r,s);
                // r-te Basisvariable tauscht mit s-ter Nicht-Basisvariablen
                swapVariables(r,s);
                for (int i = 0; i <= m; i++)
                {
                    for (int j = 0; j <= n; j++)
                    {
                        String st = df.format(a[i][j]);
                        st = " ".substring(st.length());
                        System.out.print(st);
                        System.out.print(" " + df.format(a[i][j]));
                    }
                    System.out.println();
                }
                System.out.println();
            }
        }
        catch (SimplexFinishedException e) { }
        // Suche die Strukturvariablen und kopiere
        // die Werte (in der 0-ten Spalte) in den Vektor x
        int i = 0;
        while (i < m)
        {
            int j = bv[i];
            if (j >= 0) x[j] = a[++i][0];
            else i++;
        }
        return a[0][0];
    }
}
private int pivotColumn() throws SimplexFinishedException
{
    double min = a[0][1];
    int s = 1;
    for (int j = 2; j <= n; j++)
        if (a[0][j] < min)
        {
            min = a[0][j];
            s = j;
        }
}

```

```

// Wenn alle a[0][j] nicht-negativ sind, ist die Loesung
// gefunden
if (min >= 0.0) throw new SimplexFinishedException();
return s;
}
private int pivotRow(int s) throws SimplexNoSolutionException
{
    double min = 0.0;
    int n = 0;
    int [] r = new int[m];
    for (int i = 1; i <= m; i++)
    {
        double q = a[i][0] / a[i][s];
        if (q > 0.0)
        {
            if (n == 0 || q < min)
            {
                min = q;
                n = 1;
                r[0] = i;
            }
            else if (q == min)
            { r[n++] = i; }
        }
    }
    // primale Entartung, keine Loesung
    if (n == 0) throw new SimplexNoSolutionException();
    // Normalfall
    if (n == 1) return r[0];
    // primale Entartung: mehrere Moeglichkeiten, zufaellig waehlen
    int k = 0;
    do
    {
        k = random.nextInt() % n;
    } while (k < 0);
    return r[k];
}
private void calcCoefficients(int r, int s)
{
    // Pivotelement
    a[r][s] = 1.0 / a[r][s];
    // Pivotspalte
    for (int i = 0; i <= m; i++)
        if (i != r) a[i][s] *= (- a[r][s]);
    // Pivotzeile
    for (int j = 0; j <= n; j++)
        if (j != s) a[r][j] *= a[r][s];
    // andere Elemente
    for (int i = 0; i <= m; i++)
        for (int j = 0; j <= n; j++)
        {
            if (i != r && j != s)
            {
                a[i][j] += (a[i][s] * a[r][j] / a[r][s]);
            }
        }
}
private void swapVariables(int r, int s)
{
    // r-te Basisvariable tauscht mit s-ter Nicht-Basisvariable
    r--;
    s--;
    int k = bv[r];
    bv[r] = nbv[s];
    nbv[s] = k;
}

```



```

public static void main(String args[])
{
    Simplex s;
    // double[][] a;
    double[] x;
    s = new Simplex(3,2);
    // a = new double[3 + 1][2 + 1];
    a = s.getCoefficients();
    // a = new double[3 + 1][2 + 1];
    a[0][0] = 0; a[0][1] = -500; a[0][2] = -800;
    a[1][0] = 24; a[1][1] = 5; a[1][2] = 2;
    a[2][0] = 24; a[2][1] = 1; a[2][2] = 5;
    a[3][0] = 36; a[3][1] = 6; a[3][2] = 6;
    System.out.println();
    for (int i = 0; i <= m; i++)
    {
        for (int j = 0; j <= n; j++)
        {
            String st = df.format(a[i][j]);
            st = " ".substring(st.length());
            System.out.print(st);
            System.out.print(" " + df.format(a[i][j]));
        }
        System.out.println();
    }
    System.out.println();

    x = new double[2];
    try {
        double max;
        max = s.solve(x);
        System.out.println("ZMAX = " + df.format(max));
        for (int j = 0; j < 2; j++)
            System.out.println("x[" + (j + 1) + "] = " + df.format(x[j]));
    }
    catch(Throwable e)
    { System.out.println(e); }
}

```

2.4.1.7 Simplex-Optimierung mit dem Excel-Solver

Für die Lösung ist ein Tableau, z.B. in folgender Form zu erstellen:

	A	B	C	D	E	F	G	H
1	Simplex-Optimierung mit dem Excel-Solver							
2	Fallbeispiel aus der Produktion							
3								
4		x1	x2					
5	Veränderbare Zellen	0	0	-66000				
6	Zielfunktionskoeffizienten	600	700	RS akt	RS max			
7	NB 1: y1	1	2	0	160			
8	NB 2: y2	1	1	0	120			
9	NB 3: y3	3	1	0	300			
10	NB 4: y4	0	1	0	60			
11								
12	In den Bereich B6:C10 sind die Zahlenwerte der Koeffizienten matrix einzugeben,							
13	in Zeile 5 zunächst plausible Startwerte für x1 und x2,							
14	im Bereich E7:E10 die Werte der rechten Seite,							
15	in Spalte D werden die rechten Seiten für aktuelle x1-, x2-Werte berechnet.							
16								
17								
18								
19								
20								
21								
22								

Die Zielzelle enthält eine Formel, die aus der Summe der Produkte von Zielfunktionskoeffizienten und veränderbaren Zellen gebildet wird. Die Zellen im Bereich D7:D10 enthalten die Formeln zur Berechnung des aktuellen Werts der rechten Seite. Die Formel ist jeweils bestimmt durch die Summe der Produkte aus dem jeweiligen Element der Koeffizientenmatrix mit den veränderbaren Zellen.

Beim Aufruf des Solver müssen (im Solver-Fenster) eingetragen werden:

- die Zielzelle (D5)
- Angabe, dass der Zielwert maximiert werden soll
- die veränderbaren Zellen (Bereich B5:C5)
- die Nebenbedingungen

Darüber hinaus können noch einige Optionen gesetzt werden:

Solver-Parameter

Zielzelle: Lösen

Zielwert: ☒ Max ☐ Min ☐ Wert: Schließen

Veränderbare Zellen: Schätzen

Nebenbedingungen:

Hinzufügen Ändern Löschen Optionen... Zurücksetzen Hilfe

Das Ergebnis zeigt der folgende Antwortbericht nach "Lösen" durch den Solver:

Microsoft Excel - Mappe4

Microsoft Excel 9.0 Antwortbericht

1 Microsoft Excel 9.0 Antwortbericht
 2 Tabelle: [Mappe4]Tabelle1
 3 Bericht erstellt am: 09.03.2003 17:29:49

6 Zielzelle (Max)

Zelle	Name	Ausgangswert	Lösungswert
\$D\$5	Veränderbare Zellen	-66000	10000

11 Veränderbare Zellen

Zelle	Name	Ausgangswert	Lösungswert
\$B\$5	Veränderbare Zellen x1	0	80
\$C\$5	Veränderbare Zellen x2	0	40

17 Nebenbedingungen

Zelle	Name	Zellwert	Formel	Status	Differenz
\$D\$7	NB 1: y1 RS akt	160	\$D\$7 <= \$E\$7	Einschränkend	0
\$D\$8	NB 2: y2 RS akt	120	\$D\$8 <= \$E\$8	Einschränkend	0
\$D\$9	NB 3: y3 RS akt	280	\$D\$9 <= \$E\$9	Nicht einschränkend	20
\$D\$10	NB 4: y4 RS akt	40	\$D\$10 <= \$E\$10	Nicht einschränkend	20
\$B\$5	Veränderbare Zellen x1	80	\$B\$5 >= 0	Nicht einschränkend	80
\$C\$5	Veränderbare Zellen x2	40	\$C\$5 >= 0	Nicht einschränkend	40

Antwortbericht 1 / Tabelle1 / Tabelle2 / Tabelle3 /

Bereit

2.4.2 Probleme mit unzulässiger Ausgangsform

2.4.2.1 ">=" -Restriktionen

Fallbeispiel: Die Aufgabenstellung in 2.4.1 Abschnitt 3 wird durch folgenden Zusatz erweitert: Von Produkt A müssen mindestens 20 (ME) hergestellt werden.

Mathematisches Modell: (1) $x_1 \geq 0, x_2 \geq 0$
 (2) $x_1 + 2x_2 \leq 160$
 $x_1 + x_2 \leq 120$
 $3x_1 + x_2 \leq 300$
 $x_2 \leq 60$
 $x_1 \geq 20$
 (3) $Z - 600x_1 - 700x_2 = -66000$

$x_1 \geq 20$?

- 1) Der zulässige Bereich ist dadurch weiter eingeschränkt. $(x_1, x_2) = (0, 0)$ ist nicht mehr im zulässigen Bereich eingeschlossen.
- 2) Für die Anwendung der Simplex-Methode ist unbedingt notwendig: Ermittlung der optimalen Lösung im zulässigen Bereich (Optimierungsphase, Phase 2)
- 3) Das Festlegen des zulässigen Bereichs erfolgt in der Phase 1:

	x_1	x_2	RS	
Z	-600	-700	-66000	
y_1	1	2	160	
y_2	1	1	120	
y_3	3	1	300	
y_4	0	1	60	
y_5	-1	0	-20	

- 1) Das Simplex-Verfahren ist in einem solchen Fall solange anzuwenden wie auf der rechten Seite (RS) negative Werte auftreten. Die Ermittlung der Elemente der Pivot-Zeile erfolgt durch: $a_{rj}^* = \frac{a_{rj}}{a_{rs}}$. Ein negativer Quotient bewirkt bei negativer RS, daß die neue Basisvariable positiv wird.
- 2) Stehen mehrere negative Elemente in der Zeile mit negativer RS, kann man ein beliebiges Element wählen.

Es ergibt sich damit folgendes Tableau zur Lösung des Fallbeispiels:

	y_5	x_2	RS	
Z				
y_1				
y_2				
y_3				
y_4				
x_1				

Phase 2: Bestimmen der optimalen Lösung

			RS	
Z				

			RS	
Z				

			RS	
Z				

2.4.2.2 Die duale Simplex-Methode

2.4.2.2.1 Das zweidimensionale Verschnittproblem

Aus Tafeln der Abmessung $2m \cdot 1m$ sind Tafeln vorgegebener Länge und Breite bei minimalem Verschnitt zu gewinnen. Zu liefern sind:

102 Tafeln $0.5 \cdot 0.6 \text{ m}^2$ (Tafelart I), 50 Tafeln $0.8 \cdot 0.4 \text{ m}^2$ (Tafelart II), 100 Tafeln $0.3 \cdot 1.0 \text{ m}^2$ (Tafelart III)

Gesucht ist der Verschnittplan mit minimalem Verschnitt.

Festlegung der Zuschnittvarianten: Aus produktionstechnischen Gründen sollen nur die folgenden 6 Zuschnittvarianten (V_1 bis V_6) betrachtet werden.

V_1 : 0.20m^2

I	I	I	
I	I	I	

V_2 : 0.16m^2

III	III	III	III	II
				II

V_3 : 0.20m^2

I				
	II	II	III	III
I				

V_4 : 0.40m^2

II	II	II	II	II

V_5 : 0.16m^2

I	I	
		II
I	I	II

V_6 : 0.20m^2

III	III	III	III	III	III

Man kann folgende Kombinationstafel aufstellen:

i	Variante V_i	Ges. Stückzahl	Tafelart	Verschnitt
1	V_1	x_1	6I	$0.20m^2$
2	V_2	x_2	2II 4III	$0.16m^2$
3	V_3	x_3	2I 2II 2III	$0.16m^2$
4	V_4	x_4	5II	$0.40m^2$
5	V_5	x_5	4I 2II	$0.16m^2$
6	V_6	x_6	6III	$0.20m^2$

Mathematisches Modell:

$$(1) x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0$$

$$(2) 6x_1 + 2x_3 + 4x_5 \geq 102$$

$$2x_2 + 2x_3 + 5x_4 + 2x_5 \geq 50$$

$$4x_2 + 2x_3 + 6x_6 \geq 100$$

$$(3) Z_{\min} = 0.2x_1 + 0.16x_2 + 0.16x_3 + 0.4x_4 + 0.16x_5 + 0.2x_6$$

Die Forderung den Verschnitt minimal zu halten, kann man - etwas vereinfacht - auch so formulieren: Die Anzahl der benötigten Tafeln soll möglichst gering sein:

Dies führt zu der Zielfunktion:

$$Z_{\min} = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

$$Z_{\max} = -Z_{\min}$$

$$Z_{\max} = -x_1 - x_2 - x_3 - x_4 - x_5 - x_6$$

Normalform: (1) $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0$

$$(2) -6x_1 - 2x_3 - 4x_5 \leq -102$$

$$-2x_2 - 2x_3 - 5x_4 - 2x_5 \leq -50$$

$$-4x_2 - 2x_3 - 6x_6 \leq -100$$

$$(3) Z_{\max} + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 0$$

Ausgangstableau:

	x_1	x_2	x_3	x_4	x_5	x_6	RS	
Z								
y_1								
y_2								
y_3								

Anwendung der dualen Simplex-Methode:

- Pivot-Zeile: Absolut größtes Element der rechten Seite (RS)

- Pivot-Spalte: Absolut kleinster Quotient aus Zielfunktionskoeffizient und den Koeffizienten der Pivot-Zeile.

Tableaus der optimalen Lösung

							RS	
Z								

							RS	
Z								

							RS	
Z								

							RS
Z							

							RS
Z							

							RS
Z							

Die optimale Lösung zeigt drei Alternativen. Nur die 1. Alternative erfüllt die Forderung nach Ganzzahligkeit. Die optimale Lösung ist: $x_1 = 17$, $x_2 = 25$ d.h.: 17 Tafeln werden nach Variante 1 zugeschnitten, 25 Tafeln werden nach Variante 2 zugeschnitten. Der minimale Verschnitt beträgt: 7.4m^2 .

2.4.2.2.2 Das eindimensionale Verschnittproblem

Aus Stäben von 10m Länge werden Stäbe von 3m, 4m und 5m Länge geschnitten. Verlangt werden 240 Stäbe zu 3m Länge, 150 Stäbe zu 4m Länge, 200 Stäbe zu 5m Länge.

1) Wieviel Stäbe zu 10m werden benötigt und wie sind sie zu zerschneiden, wenn der Abfall minimiert werden soll?

Plan mit den möglichen Schnittkombinationen

	5m	4m	3m	Abfall
x_1	2	0	0	0
x_2	1	1	0	1
x_3	1	0	1	2
x_4	0	2	0	2
x_5	0	1	2	0
x_6	0	0	3	1

Mathematisches Modell:

(1) $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0$

(2) $2x_1 + x_2 + x_3 \geq 200$

$x_2 + 2x_4 + x_5 \geq 150$

$x_3 + 2x_5 + 3x_6 \geq 240$

(3) $Z_{\min} = x_2 + 2x_3 + 2x_4 + x_6$

Anwendung des Simplex-Verfahrens zur Ermittlung der optimalen Lösung

	x_1	x_2	x_3	x_4	x_5	x_6	RS	
Z								
y_1								
y_2								
y_3								

							RS	
Z								

							RS	
Z								

							RS	
Z								

Optimale Lösung:

Wieviel Stäbe werden benötigt, wenn man möglichst wenig Stäbe zu 10 m Länge zerschneidet?

Mathematisches Modell

(1) $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0$

(2) $2x_1 + x_2 + x_3 \geq 200$

$x_2 + 2x_4 + x_5 \geq 150$

$x_3 + 2x_5 + 3x_6 \geq 240$

(3) $Z_{\min} = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$

Ausgangstableau

							RS	
Z								

Anwendung des Simplex-Verfahrens zur Ermittlung der optimalen Lösung

							RS	
Z								

							RS	
Z								

							RS	
Z								

							RS	
Z								

Alternative zur vorliegenden optimalen Lösung

							RS	
Z								

2.4.2.3 Gleichungen als Restriktionen

1) Einführung

Im Rahmen des Simplex-Verfahrens können auch „=“-Restriktionen berücksichtigt werden. Eine Gleichung, die nur Strukturvariable enthält ist im Ausgangstableau bestimmt nicht erfüllt. Ausnahme ist: Die rechte Seite (RS) der Gleichung ist 0. Es ist zweckmäßig auch hier Schlupfvariable einzuführen. Dann ist die Gleichung formal immer erfüllt. Diese Schlupfvariable muß selbstverständlich in der Lösung den Wert 0 annehmen, d.h.: Sie darf nicht mehr in der Basis sein und nicht mehr in die Basis gelangen. Sie ist gesperrt für Basisvariable.

2) Verfahrensweise

- Falls gesperrte Basisvariable vorhanden sind, ist als Pivot-Zeile eine Zeile mit solchen Variablen zu wählen.
- Bestimmung der Pivot-Spalte aus den Spalten nicht gesperrter Basis-Variablen, die in der Pivot-Zeile Nichtnullelemente enthalten.
- Berechnung des Simplex-Tableaus nach den bekannten Regeln, die für Phase 1 und Phase 2 gelten. Die Phase der Austreibung der gesperrten Variable ist somit Phase 0.

3) Bsp.: Erweiterung der Aufgabe

Genau 110 Einheiten der Produkte A und B sollen hergestellt werden.

Mathematisches Modell: (1) $x_1 \geq 0$, $x_2 \geq 0$

$$(2) \quad x_1 + 2x_2 \leq 160$$

$$x_1 + x_2 \leq 120$$

$$3x_1 + x_2 \leq 300$$

$$x_2 \leq 60$$

$$x_1 + x_2 = 110$$

$$(3) \quad Z_{\max} = 600x_1 + 700x_2 - 66000$$

Lösungsschritte

Z			RS	

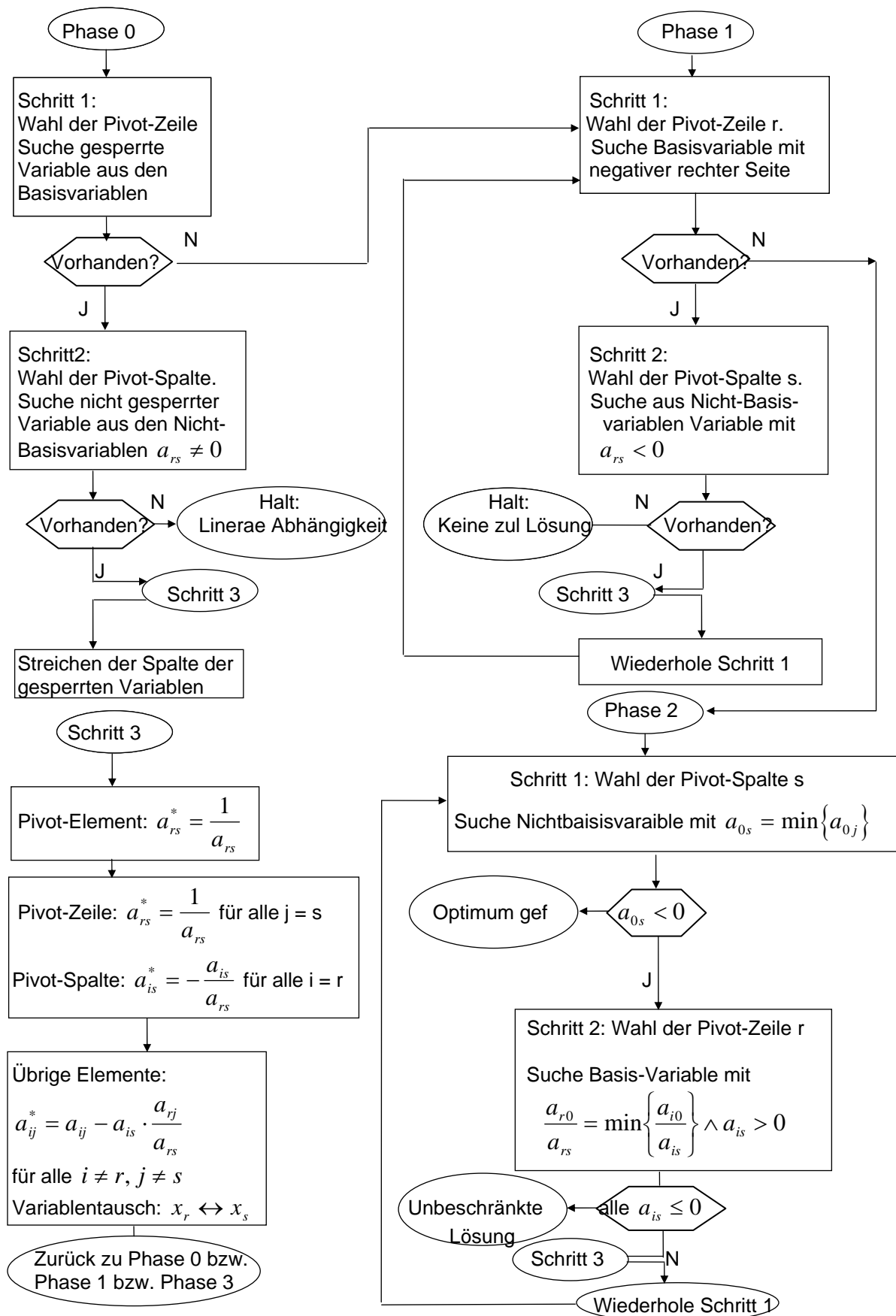
Z			RS	

Z		RS	

Z		RS	

Z		RS	

2.4.2.4 Zusammenfassung: Ablauf des Simplex-Verfahrens



2.4.3 Unter- und Obergrenzen von Variablen

2.4.3.1 Vereinbarungen

Untergrenzen (Lower Bounds) haben die Form: $x_j \geq l_j$

Obergrenzen (Upper Bounds) haben die Form: $x_j \leq u_j$

Es ist möglich, solche Grenzbedingungen wie gewöhnliche Restriktionen zu behandeln. Jedoch ist eine einfachere Behandlung möglich.

2.4.3.2 Untergrenzen von einzelnen Variablen $x_j \geq l_j$

Falls man Untergrenzen nicht als Restriktion in das Simplex-Tableau aufnehmen will, führt man eine Koordinatenverschiebung durch. Es gilt: $x_j - l_j \geq 0$. Diese Grenzbedingung wird durch die Nichtnegativitätsbedingung einer Variablen $x_j' \geq 0$ erfüllt. Für sie gilt dann: $x_j' = x_j - l_j$. Dadurch ergibt sich:

$$y_i + \sum_{j=1}^n a_{ij} \cdot x_j = b_i$$

$$y_i + \sum_{j=1}^n a_{ij} \cdot (x_j' + l_j) = b_i$$

$$y_i + \sum_{j=1}^n a_{ij} x_j' = b_i - \sum_{j=1}^n a_{ij} \cdot l_j$$

Die letzte Formel zeigt: Nur Elemente der rechten Seite (RS) sind zu ändern.

Bsp.: Zur Produktion von 2 Artikeln stehen 2 Maschinengattungen A und B zur Verfügung. Montagearbeiter sind ebenfalls nur in geringer Anzahl vorhanden. Die speziellen Daten sind in der folgenden Tabelle zusammengefaßt.:

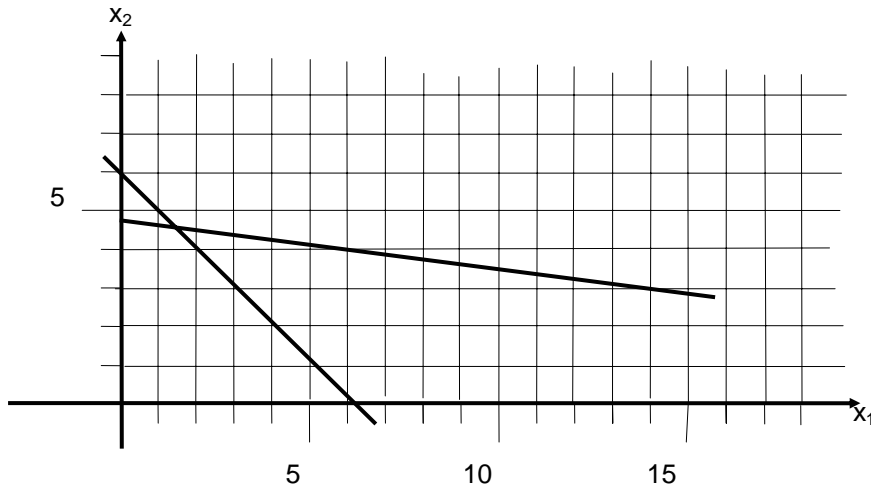
	Artikel 1 (ZE/ME)	Artikel 2 (ZE/ME)	Kapazität (ZE)
Maschine A	5	2	24
Maschine B	1	5	24
Montagegruppe	6	6	36

Artikel 1 kann mit einem Gewinn je Stück von 500 (GE), Artikel 2 kann mit einem Gewinn je Stück von 800 (GE) verkauft werden. Durch Lieferverpflichtungen sind mindestens 1 ME von Artikel 1 und 3 ME von Artikel 2 bereitzustellen. Gesucht sind die Mengen x_1 und x_2 der Artikel 1 und 2, die angefertigt werden müssen, um den Gewinn zu maximieren.

Mathematisches Modell: (1) $x_1 \geq 0, x_2 \geq 0$
 (2) $5x_1 + 2x_2 \leq 24$
 $x_1 + 5x_2 \leq 24$

$$\begin{aligned}
 6x_1 + 6x_2 &\leq 36 \\
 x_1 &\geq 1 \\
 x_2 &\geq 3 \\
 (3) \quad Z_{\max} - 500x_1 - 800x_2 &= 0
 \end{aligned}$$

Grafische Lösung:



Lösung nach dem Simplex-Verfahren (mit Berücksichtigung unterer Grenzen):

	x_1	x_2	RS
untere Grenzen	1	3	-
Z_{\max}	-500	-800	0
y_1	5	2	24
y_2	1	5	24
y_3	6	6	36

$$RS = b_i - \sum_{j=1}^n a_{ij} \cdot l_j$$

	x_1'	x_2'	RS	
Z_{\max}				
y_1				
y_2				
y_3				

Optimierung:

			RS	
Z_{\max}				

			RS
Z_{\max}			

Optimale Lösungswerte: $x_1' = 0.5$, $x_2' = 1.5$. Nach der Optimierung werden die gefundenen Lösungswerte wieder in den ursprünglichen Koordinaten angegeben:

Aus $x_1' = x_1 - 1$, $x_2' = x_2 - 3$ folgt: $x_1 = 1.5$ und $x_2 = 4.5$

2.4.3.3 Obergrenzen von einzelnen Variablen

1) Grenzbedingung für Nichtbasisvariable ($x_j \leq u_j$)

Transformation: $\bar{x}_j = u_j - x_j$

$$y_i + \sum_{j=1}^n a_{ij} \cdot x_j = b_i$$

$$y_i + \sum_{j=1}^n a_{ij} \cdot (u_j - \bar{x}_j) = b_i$$

$$y_i - \sum_{j=1}^n a_{ij} \cdot \bar{x}_j = b_i - \sum_{j=1}^n a_{ij} \cdot u_j$$

Die Koeffizienten in der Spalte x_j ändern ihre Vorzeichen

2) Grenzbedingung für Basisvariable ($y_i \leq u_i^*$)

Transformation: $\bar{y}_i = u_i^* - y_i$ ($y_i = u_i^* - \bar{y}_i$)

$$y_i + \sum_{j=1}^n a_{ij} \cdot x_j = b_i$$

$$-y_i + u_i^* + \sum_{j=1}^n a_{ij} \cdot x_j = b_i$$

$$\bar{y}_i + \sum_{j=1}^n a_{ij} \cdot x_j = u_i^* - b_i$$

Die Transformation ist hier nicht vor, sondern während der Rechnung zu vollziehen. Eine Transformation ist immer erforderlich, wenn eine Variable ihre obere Grenze erreicht.

3) Upper-Bounding-Technique

Unter diesem Begriff werden die speziellen Rechenregeln für die Behandlung oberer Grenzen zusammengefaßt.

Phase 0: Keine Änderung durch die „Upper-Bounding-Technique“

Phase 1: Vor Auswahl der Pivot-Zeile und Pivot-Spalte werden diejenigen Basisvariablen durch ihre korrespondierenden Variablen ersetzt (z.B. y_i durch \bar{y}_i), deren Wert die obere Grenze übersteigt

Phase 2: 1) Keine Änderung bei der Auswahl der Pivot-Spalte

2) Wahl der Pivot-Zeile

a) Bestimme den kleinsten nichtnegativen Quotienten Q_1 der Elemente von

$$RS \text{ und der Koeffizienten der Pivot-Spalte: } Q_1 = \min_{i=1..m, a_{is} > 0} \left\{ \frac{a_{i0}}{a_{is}} \right\}$$

b) Berücksichtigung korrespondierender Variablen zu vorhandenen

Basisvariablen (z.B. \bar{y}_i zu y_i) durch Bestimmen des Quotienten nach der

$$\text{Formel: } Q_2 = \min_{i=1..m, a_{is} < 0} \left\{ \frac{a_{i0} - u_i^*}{a_{is}} \right\}$$

c) Berücksichtigung der Obergrenze der Nichtbasisvariablen der Pivot-Spalte durch Bestimmen von Q_3 : $Q_3 = u_s$

Fallunterscheidung: 1) $Q_1 < Q_2 \wedge Q_3$

2) $Q_2 < Q_1 \wedge Q_3$

a) Die Basisvariable der entsprechenden Zeile wird durch die korrespondierende Variable ersetzt (z.B. y_i durch \bar{y}_i).

b) Dann folgt eine gewöhnliche Simplex-Iteration

3) $Q_3 < Q_1 \wedge Q_2$

Die Nichtbasisvariable der Spalte s wird durch die korrespondierende Variable ersetzt.

Bsp.: 1.3 Millionen sollen möglichst vorteilhaft angelegt werden. Es stehen dafür 7 Anlagenprojekte zur Verfügung. Die Investitionssumme für jedes Projekt ist durch den in der nachfolgenden Tabelle angegebenen Höchstwert begrenzt. Ebenfalls ist hier die jährliche Nettorendite angegeben. Alle Projekte sind mit dem gleichen allerdings vernachlässigbaren Verlustrisiko behaftet. Das in den Projekten nicht angelegte Kapital lässt sich nicht zinsbringend verwenden.

Projekt	Investitionssumme	Jährliche Rendite
1	200	18000.-DM 9%
2	100	8000.-DM 8%
3	500	35000.-DM 7%
4	300	20000.-DM 20/3%
5	400	26000.-DM 13/2%
6	500	31000.-DM 31/5%
7	100	6000.-DM 6%

Stelle den optimalen Investitionsplan auf!

Mathematisches Modell:

$$(1) x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0, x_6 \geq 0, x_7 \geq 0$$

$$(2) x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 \leq 1300$$

$$(3) Z_{\max} - 9x_1 - 8x_2 - 7x_3 - 20/3x_4 - 13/2x_5 - 31/5x_6 - 6x_7 = 0$$

Lösung mit der „Upper-Bounding Technique“:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
Obere Grenzen	200	100	500	300	400	500	100	RS
Z_{\max}	-9	-8	-7	-20/3	-13/2	-31/5	-6	0
y	1	1	1	1	1	1	1	1300

Obere Grenzen	200	100	500	300	400	500	100	RS
Z_{\max}								
y								

Obere Grenzen	200	100	500	300	400	500	100	RS
Z_{\max}								
y								

Obere Grenzen	200	100	500	300	400	500	100	RS
Z_{\max}								
y								

Obere Grenzen	200	100	500	300	400	500	100	RS
Z_{\max}								
y								

Optimale Lösung: $x_1 = 200.000$.-DM
 $x_2 = 100.000$.-DM
 $x_3 = 500.000$.-DM
 $x_4 = 300.000$.-DM
 $x_5 = 200.000$.-DM
Rendite: 94.000.-DM

Aufgaben

1. Löse mit Hilfe der Simplex-Methode

(1) $x_1 \geq 0, x_2 \geq 0$

(2) $5 \leq x_1 \leq 7$
 $x_2 \geq 3$
 $2x_1 + 4x_2 \geq 6$

(3) $Z_{\min} = x_1$

Standardisierung: $-2x_1 - 4x_2 \leq -6$

$$Z_{\min} - x_1 = 0$$

$$Z_{\max} + x_1 = 0$$

Untere Grenze: $x_1' = x_1 - 5, x_2' = x_2 - 3$

	x_1	x_2	
untere Grenze	5	3	RS
Z_{\max}			
Y_1			

	x_1'	x_2'	
obere Grenze	2	∞	RS
Z_{\max}			
y_1			

Lösung: $x_1 = 5$, $x_2 = 3$, $Z_{\min} = 5$

2. Löse mit der Simplex-Methode

(1) $x_1 \geq 0$, $x_2 \geq 0$

(2) $x_1 + x_2 \leq 8$
 $x_1 \leq 6$

(3) $Z_{\max} = 4x_1 + x_2$

	x_1	x_2	RS
obere Grenze	6	∞	
Z_{\max}	-4	-1	0
y_1	1	1	8

	\bar{x}_1	x_2	RS
obere Grenze	6	∞	
Z_{\max}	4	-1	24
y_1	-1	1	2

	\bar{x}_1	y_1	RS
obere Grenze	6	∞	
Z_{\max}	3	1	26
x_2	-1	1	2

Lösung: $x_1 = 6$, $x_2 = 2$, $Z_{\max} = 26$

3. Zweipersonen-Nullsummenspiele

3.1 Grundlagen

3.1.1 Matrixspiele

Zweipersonen-Nullsummenspiele sind ein wichtiger Untersuchungsgegenstand der Spieltheorie. Die Spieltheorie ist ein relativ junger Zweig der Mathematik. Ihre grundlegenden Ideen stammen von dem Franzosen Emil Borel, die ersten fundierten mathematischen Aussagen wurden 1926 von John von Neumann ausgearbeitet. Die Spieltheorie bildet die Basis für Optimierungsrechnungen von Konkurrenzmodellen. Spiele beschreiben generell Konfliktsituationen. Das Ziel einer Konfliktanalyse besteht im Aufdecken des optimalen (vernünftigen, gerechten) Ausgangs. Alle anderen Aspekte, die einen Konflikt charakterisieren, haben sich diesem Ziel unterzuordnen. Die exakte Beschreibung der Konfliktsituationen durch mathematische Modelle, die Bereitstellung von Verfahren zur Berechnung für die Ermittlung der optimalen Entscheidungen ist Aufgabe der Spieltheorie.

Überwiegend beschäftigt sich die Spieltheorie mit **Zweipersonen-Nullsummenspielen**, d.h.: In einem Zweipersonen-Nullsummenspiel besteht die Spielermenge aus 2 Personen. Jede dieser beiden Personen erwartet vom Ausgang des Spiels einen Gewinn. Dieser Gewinn soll möglichst groß sein, falls jeder der beteiligten Spieler die für ihn zweckmäßigste Strategie gewählt hat. In Nullsummenspielen ist die Summe der erwarteten Gewinne ausgeglichen:

$$\sum_{i=1}^n F_i = 0$$

In Zweipersonen-Nullsummenspielen gilt:

$$\sum_{i=1}^2 F_i = 0$$

bzw.

$$F_1 = -F_2$$

Neben der Gewinnfunktion sind die Strategien

$$X = \{x_1, \dots, x_m\}$$

bzw.

$$Y = \{y_1, \dots, y_n\}$$

der beiden Spieler wichtig. Die Gewinnfunktion ist von der Wahl einer bestimmten Strategie der vorliegenden Strategien bestimmt.

$$F(x_i, y_j)$$

	y_1	y_2	y_3	...	y_j	...	y_{n-1}	y_n
x_1	F_{11}	F_{12}	F_{13}	...	F_{1j}	...	F_{1n-1}	F_{1n}
x_2	F_{21}	F_{22}	F_{23}	...	F_{2j}	...	F_{2n-1}	F_{2n}
...
x_i	F_{i1}	F_{ij}
...
x_m	F_{m1}	F_{mj}	F_{mn}

Bsp.:

Spieler 1 hat 3 Strategien zur Verfügung: x_1, x_2, x_3

Spieler 2 hat 2 Strategien zur Verfügung: y_1, y_2

Die Spielregeln setzen fest, daß Zahlungen in der folgenden Weise mit den eingeschlagenen Strategien verbunden sind:

gewählte Strategie	Folge
x_1 bzw. y_1	1 zahlt 2: 2[GE]
x_1 bzw. y_2	2 zahlt 1: 1[GE]
x_2 bzw. y_1	1 " 2: 1[GE]
x_2 " y_2	2 " 1: 3[GE]
x_3 " y_1	2 " 1: 1[GE]
x_3 " y_2	2 " 1: 2[GE]

Weitaus übersichtlicher ist die folgende Matrix:

		2	
		y_1	y_2
	x_1	-2	1
1	x_2	-1	3
	x_3	1	2

Eine solche Tabelle (Matrix) bezeichnet man als Gewinnmatrix des Spielers 1.

F_{11}	F_{1j}	F_{1n}
...
F_{i1}	F_{ij}	F_{in}
...
F_{m1}	F_{mj}	F_{mn}

Spiele, für die eine solche Matrix aufgestellt werden kann, heißen **Matrixspiele**.

3.1.2 Wert eines Spiels und optimale Strategien

Optimale Strategien sind für Gleichgewichtsbedingungen gegeben.

Jeder Spieler muß sich mit seiner Strategiemenge und der seiner Mitspieler auseinandersetzen. Als Ergebnis dieser Auseinandersetzung wird eine Situation erarbeitet, die jeder Spieler akzeptiert und von der kein Spieler abweichen will (Gleichgewichtssituation). Weicht der Spieler davon ab, hat er keinen größeren Gewinn zu erwarten. Deshalb wird eine solche Situation von allen Spielern als optimal anerkannt.

Beschreibt man die Gleichgewichtssituation durch die Strategie (x^*, y^*) , dann gilt folgende Gleichgewichtsdefinition:

$$F_1(x, y^*) \leq F_1(x^*, y^*)$$

$$F_2(x^*, y) \leq F_2(x^*, y^*)$$

Aus $F_1(x, y) = -F_2(x, y)$ und aus der letzten Ungleichung folgt:

$$-F_1(x^*, y) \leq -F_1(x^*, y^*) \text{ bzw. } F_1(x^*, y^*) \leq F_1(x^*, y)$$

Damit läßt sich die Bedingung für einen Gleichgewichtspunkt (x^*, y^*) formulieren:

$$F(x, y^*) \leq F(x^*, y^*) \leq F(x^*, y)$$

Diese Bedingung ist gleichwertig mit

$$F(x^*, y^*) = \max_{x \in X} F(x, y^*) = \min_{y \in Y} F(x^*, y)$$

bzw.

$$\max_{x \in X} F(x, y^*) \geq F(x, y^*) \quad \text{für alle } x \in X$$

$$\min_{y \in Y} F(x^*, y) \leq F(x^*, y) \quad \text{für alle } y \in Y$$

Die Gewinnfunktion $F(x, y)$ muß also in der Gleichgewichtssituation (x^*, y^*) bzgl. x ein Maximum und bzgl. y ein Minimum erreichen.

Die Gleichgewichtspunkte der Zweipersonen-Nullsummenspiele bezeichnet man als **Sattelpunkte**.

Minimax-Prinzip

Für die Existenz des Sattelpunktes und damit für den Wert des Spiels gilt:

$$\max_{x \in X} \min_{y \in Y} F(x, y) = \min_{y \in Y} \max_{x \in X} F(x, y)$$

3.1.3 Ermittlung des Sattelpunkts für Matrixspiele

Der Wert des Spiels ist gemäß dem **Minimax**-Prinzip bestimmt durch:

$$\max_{i=1}^m \min_{j=1}^n F_{ij} = \min_{j=1}^n \max_{i=1}^m F_{ij}$$

Bei der Bestimmung des Werts vom Spiel geht man zweckmäßigerweise so vor:

$$\max_{i=1}^m \min_{j=1}^n F_{ij}$$

In jeder Zeile i ($i=1,\dots,m$) ermittelt man den minimalen Funktionswert der Gewinnfunktion F_{ij} . Aus der so erhaltenen Menge bestimmt man das Maximum.

$$\min_{j=1}^n \max_{i=1}^m F_{ij}$$

In jeder Spalte j ($j = 1,\dots,n$) ermittelt man den maximalen Funktionswert und anschließend aus der erhaltenen Menge das Minimum.

Bsp.:

1. Bestimme , welche der folgenden durch die jeweilige Spielmatrix festgelegte Spiele einen Sattelpunkt besitzen:

a)

0	2
-1	4

besitzt einen Sattelpunkt

b)

5	0
0	2

besitzt keinen Sattelpunkt

c)

2	1
4	0

besitzt einen Sattelpunkt

d)

1	-1
-1	1

besitzt keinen Sattelpunkt

2. 2 Spieler zeigen jeder gleichzeitig einen oder 2 Finger. Spieler 1 zahlt an Spieler 2 einen Betrag, der der Gesamtzahl gezeigter Finger entspricht. Spieler 2 zahlt an Spieler 1 einen Betrag, der gleich dem Produkt der Zahlen der gezeigten Finger ist. Stelle die Spielmatrix auf (Die Werte entsprechen dem Nettogewinn von Spieler 2). Bestimme Wert und optimale Strategie!

1	1
1	0

Das Spiel besitzt einen Sattelpunkt.

Der Wert des Spiels ist 1.

Die optimale Strategie ist: Spieler 2 muß einen Finger zeigen, Spieler 1 muß einen Finger zeigen.

3. Löse das folgende durch die Gewinnmatrix festgelegte Spiel:

-2	2	-2
-1	3	-1
1	2	1
1	3	1

Das Spiel hat einen Sattelpunkt. Der Wert des Spiels beträgt 1

4. Knobelspiel "Papier(P) - Schere(Sch) - Stein(St)"

Beide Spieler wählen unabhängig voneinander eines der Symbole "Papier", "Schere" oder "Stein". Danach vergleicht man das Gewählte. Der Sieger erhält vom Verlierer 1 [GE] nach der Regel:

- Papier verliert gegenüber Schere
- Schere verliert gegenüber Stein
- Stein verliert gegenüber Papier

Die Gewinnmatrix des 1. Spielers hat dann folgendes Aussehen:

	$y_1(P)$	$y_2(SCH)$	$y_3(St)$	$\min_{j=1}^3 F_{ij}$
$x_1(P)$	0	-1	1	-1
$x_2(Sch)$	1	0	-1	-1
$x_3(St)$	-1	1	0	-1
$\max_{i=1}^3 F_{ij}$	1	1	1	

$$\max_{i=1}^3 \min_{j=1}^3 F_{ij} = -1$$

$$\min_{j=1}^3 \max_{i=1}^3 F_{ij} = 1$$

Das Minimax-Kriterium ist nicht erfüllt. Es existiert kein Sattelpunkt und Wert des Spiels.

Wählen die beiden Spieler ihre optimalen Strategien, so ist der Gewinn gleich dem Wert des Sattelpunkts. Weicht einer der beiden Spieler von seiner optimalen Strategie ab, so kann er dabei nur verlieren und niemals seinen Gewinn vergrößern.

	y_1	...	y_j	...	y_n	$\min_{j=1}^n F_{ij}$
x_1	F_{11}		F_{1j}		F_{1n}	F_1
x_2	F_{21}		F_{2j}		F_{2n}	F_2
...
x_i	F_{i1}		F_{ij}		F_{in}	F_i
...						...
x_m	F_{m1}	...	F_{mi}	...	F_{mn}	F_m
$\max_{i=1}^m F_{ij}$	E_1	...	E_j	...	E_n	$F_i = E_j$

$$\min (E_1, E_2, \dots, E_j, \dots, E_n) = E_j$$

$$\max (F_1, F_2, \dots, F_i, \dots, F_m) = F_i$$

Ein Zweipersonen-Nullsummenspiel mit Sattelpunkt verliert seinen Reiz, sobald es einmal vollständig analysiert wurde.

Interessanterweise besitzt auch das Schachspiel einen solchen Sattelpunkt, vom theoretischen Standpunkt ist das Schachspiel in diesem Sinne trivial. Es dürfte jedoch kaum gelingen, eine vollständige Matrix zu erstellen und daraus die (vollständige) Lösung des Schachspiels abzulesen. Weitaus interessanter im Rahmen der Spieltheorie sind Spiele ohne Sattelpunkt, z.B. das in den vorstehenden Beispielen angegebene Knobelspiel.

3.1.4 Gemischte Strategien

Ein Spiel, wie das Knobelspiel "Papier - Schere - Stein" wird in der Regel nicht nur einmal sondern in mehreren Partien hintereinander gespielt. In solchen Fällen ist zweckmäßiges Spielverhalten: "Häufiges Wechseln der Strategien nach den Gesetzen der Wahrscheinlichkeitsrechnung".

Definition

Gegeben ist das Matrixspiel mit den Strategiemengen $X = \{x_1, \dots, x_m\}$ und $Y = \{y_1, \dots, y_n\}$. Unter gemischter Strategie des Spieler 1 versteht man den m-dimensionalen Vektor

$\mathbf{p}^T = (p_1, p_2, \dots, p_m)$ mit $p_i \geq 0$ für alle $i = 1, \dots, m$ und

$$\sum_{i=1}^m p_i = 1$$

Eine gemischte Strategie des Spielers 2 ist dann der n- dimensionale Vektor

$\mathbf{q}^T = (q_1, q_2, \dots, q_n)$ mit $q_j \geq 0$ für alle $j = 1, \dots, n$ und

$$\sum_{j=1}^n q_j = 1$$

Mit Hilfe der gemischten Strategien läßt sich der Spielbegriff erweitern. Gegeben ist ein (endliches) Zweipersonen- Nullsummenspiel $X = \{x_1, x_2, \dots, x_m\}$ und $Y = \{y_1, \dots, y_n\}$ sind die **reinen Strategiemengen** der Spieler, \mathbf{F} ist die dazugehörige Gewinnmatrix. Eine reine Strategie ist ein Spezialfall der gemischten Strategie. Die ausgewählte reine Strategie hat die Wahrscheinlichkeit 1, alle anderen Strategien den Wert 0. Die Menge der **gemischten Strategien** sind

$$P = \{\mathbf{p} \mid \mathbf{p} \geq 0 \text{ und } \sum_{i=1}^m p_i = 1\}$$

und

$$Q = \{\mathbf{q} \mid \mathbf{q} \geq 0 \text{ und } \sum_{j=1}^n q_j = 1\}$$

Für jede gemischte Strategie $\mathbf{p} \in P$ und $\mathbf{q} \in Q$ läßt sich der Erwartungswert E (für Spieler 1) nach

$$E(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^m \sum_{j=1}^n p_i \cdot F_{ij} \cdot q_j = \mathbf{p}^T \mathbf{F} \mathbf{q}$$

bestimmen.

Das Spiel $G = (P, Q, E)$ heißt gemischte Erweiterung von G

$$\begin{aligned}
 E(\mathbf{p}, \mathbf{q}) &= \sum_{i=1}^m \sum_{j=1}^n p_i \cdot F_{ij} \cdot q_j = \\
 &= \sum_{i=1}^m \sum_{j=1}^n p_i \cdot F(x_i, y_j) \cdot q_j = \sum_{i=1}^m p_i \sum_{j=1}^n F(x_i, y_j) \cdot q_j = \sum_{j=1}^n q_j \sum_{i=1}^m p_i \cdot F(x_i, y_j) = \sum_{i=1}^m p_i \cdot E(x_i, \mathbf{q}) = \\
 &= \sum_{j=1}^n q_j \cdot E(\mathbf{p}, y_j)
 \end{aligned}$$

3.1.5 Bestimmen des Spielwerts bei gemischter Erweiterung

Ziel des 1. Spielers: $E(\mathbf{p}, \mathbf{q})$ soll möglichst groß sein.

Ziel des 2. Spielers: $E(\mathbf{p}, \mathbf{q})$ soll möglichst klein sein

Strategie des 1. Spielers: Eine Strategie $\mathbf{p} \in P$ so zu wählen, daß

$\min_{p \in P, q \in Q} E(\mathbf{p}, \mathbf{q})$ möglichst groß wird

$$v_1 = \max_{p \in P} \min_{q \in Q} E(\mathbf{p}, \mathbf{q})$$

Strategie des 2. Spielers: Eine Strategie $\mathbf{q} \in Q$ so zu wählen, daß

$\max_{p \in P, q \in Q} E(\mathbf{p}, \mathbf{q})$ möglichst klein wird.

$$v_2 = \min_{q \in Q} \max_{p \in P} E(\mathbf{p}, \mathbf{q})$$

Hauptsatz zur Theorie der Matrixspiele

Für jedes Matrixspiel besitzt die zugehörige gemischte Erweiterung $G = (P, Q, E)$ einen Wert und für beide Spieler existieren optimale Strategien. Den Wert $v(G)$ des Spiels G erhält man aus $v_1 = v_2$

$$v(G) = \max_{p \in P} \min_{q \in Q} E(\mathbf{p}, \mathbf{q}) = \min_{q \in Q} \max_{p \in P} E(\mathbf{p}, \mathbf{q})$$

Zu jedem endlichen Zweipersonen-Nullsummenspiel existiert ein Paar gemischte Strategien

$\mathbf{p}^* \in P$ und $\mathbf{q}^* \in Q$, so daß $(\mathbf{p}^*, \mathbf{q}^*)$ Sattelpunkt von G ist:

$$E(\mathbf{p}, \mathbf{q}^*) \leq E(\mathbf{p}^*, \mathbf{q}^*) \leq E(\mathbf{p}^*, \mathbf{q})$$

3.2 Lösung mit dem Simplex-Verfahren

3.2.1 Grundlagen

Ausgangspunkt ist die Sattelpunktrelation

$$v(G) = \max_{p \in P} \min_{q \in Q} E(p, q) = \min_{q \in Q} \max_{p \in P} E(p, q)$$

Aus der Sattelpunktrelation folgt:

$$v(G) = \max_{p \in P} \min_{q \in Q} E(p, q) = \max_{p \in P} E(p, q^*) = \max_{i=1}^m E(x_i, q^*)$$

bzw.

$$v(G) = \min_{q \in Q} \max_{p \in P} E(p, q) = \min_{q \in Q} E(p^*, q) = \min_{j=1}^n E(p^*, y_j)$$

$$\text{Mit } E(x_i, q^*) = \sum_{j=1}^n F_{ij} \cdot q_j^* \text{ und } E(p^*, y_j) = \sum_{i=1}^m p_i^* \cdot F_{ij} \text{ folgt:}$$

$$v(G) = \max E(x_i, q^*) = \max_{i=1}^m \sum_{j=1}^n F_{ij} \cdot q_j^* \geq \sum_{j=1}^n F_{ij} \cdot q_j^* \quad (i=1, \dots, m)$$

$$v(G) = \min E(p^*, y_j) = \min_{j=1}^n \sum_{i=1}^m p_i^* \cdot F_{ij} \leq \sum_{i=1}^m p_i^* \cdot F_{ij} \quad (j=1, \dots, n)$$

Daraus ergibt sich die Sattelpunktrelation in folgender Form:

$$\sum_{j=1}^n F_{ij} \cdot q_j^* \leq v(G) \leq \sum_{i=1}^m p_i^* \cdot F_{ij} \quad (i=1, \dots, m; j=1, \dots, n)$$

Aus dieser Sattelpunktrelation lässt sich ein zugehöriges lineares Programm ableiten, mit dessen Hilfe man die optimalen Strategien p^* und q^* ermitteln kann.

Die optimalen Strategien des Spielers 1 $p^{*T} = (p_1^*, \dots, p_m^*) \in P$ und die des Spielers 2 $q^{*T} = (q_1^*, \dots, q_n^*) \in Q$ bestimmt man aus dem folgenden System von Ungleichungen:

a)

$$\sum_{i=1}^m p_i^* \cdot F_{ij} \geq v(G) \quad (i=1, \dots, m; j=1, \dots, n)$$

$$p_i^* \geq 0, \quad \sum_{i=1}^m p_i^* = 1$$

b)

$$\sum_{j=1}^n F_{ij} \cdot q_j^* \leq v(G) \quad (i = 1, \dots, m; j = 1, \dots, n)$$

$$q_j^* \geq 0, \quad \sum_{j=1}^n q_j^* = 1$$

Ist $v(G) > 0$ (das Spiel besitzt eine positive Gewinnmatrix), so kann man beide Ungleichungssysteme durch den Wert des Spiels dividieren, ohne daß sich die Ungleichungen verändern.

a)

$$\sum_{i=1}^m \frac{p_i^*}{v(G)} \cdot F_{ij} \geq 1 \quad (j = 1, \dots, n)$$

$$u_i = \frac{p_i^*}{v(G)} \geq 0, \quad \sum_{i=1}^m \frac{p_i^*}{v(G)} = \frac{1}{v(G)} \quad (i = 1, \dots, m)$$

b)

$$\sum_{j=1}^n F_{ij} \cdot \frac{q_j^*}{v(G)} \leq 1 \quad (i = 1, \dots, m)$$

$$w_j = \frac{q_j^*}{v(G)} \geq 0, \quad \sum_{j=1}^n \frac{q_j^*}{v(G)} = \frac{1}{v(G)} \quad (j = 1, \dots, n)$$

Das Ziel des Spielers 1 besteht in der Maximierung des Spielwerts. Das Ziel des Spielers 2 besteht in der Minimierung des Spielwerts. Es ergeben sich die folgenden linearen Optimierungsprogramme.

a)

$$(1) \quad u_i \geq 0 \quad (i = 1, \dots, m)$$

$$(2) \quad \sum_{i=1}^m u_i \cdot F_{ij} \geq 1 \quad (j = 1, \dots, n)$$

$$(3) \quad Z_{\min} = \sum_{i=1}^m u_i = \frac{1}{v(G)}$$

b)

$$(1) \quad w_j \geq 0 \quad (j = 1, \dots, n)$$

$$(2) \quad \sum_{j=1}^n F_{ij} \cdot w_j \leq 1 \quad (i = 1, \dots, m)$$

$$(3) \quad Z_{\max} = \sum_{j=1}^n w_j = \frac{1}{v(G)}$$

Die beiden Optimierungsprogramme sind offenbar dual zueinander. Sie können mit dem Simplex-Verfahren gelöst werden.

3.2.2 Beispiele

1. Bestimme die optimalen Strategien zum Knobelspiel "Papier - Schere - Stein".

- Das Spiel besitzt kein Sattelpunkt

- Die Gewinnmatrix des Spiels hat folgendes Aussehen:

0	-1	1
1	0	-1
-1	1	0

Der Wert des Spiels $v(G)$ muß größer als 0 sein. Dies ist sicher dann der Fall, wenn die Gewinnmatrix nur positive Elemente enthält. Es ist daher notwendig, ein zu diesem Spiel (G) äquivalentes Spiel (G_1) zu konstruieren.

Nach der Äquivalenzrelation gilt: $F_{ij}' = a \cdot F_{ij} + b$

Ist $a = 1$ und $b = 2$, dann erhält man die Gewinnmatrix:

2	1	3
3	2	1
1	3	2

Es kann daraus folgendes lineares Programm aufgestellt werden:

$$\begin{array}{ll}
 (1) \quad u_i \geq 0 & (i=1,2,3) \\
 (2) \quad \begin{array}{l} 2u_1 + 3u_2 + u_3 \geq 1 \\ u_1 + 2u_2 + 3u_3 \geq 1 \\ 3u_1 + u_2 + 2u_3 \geq 1 \end{array} \\
 (3) \quad Z_{\min} = u_1 + u_2 + u_3
 \end{array}
 \quad
 \begin{array}{ll}
 (1) \quad w_j \geq 0 \\
 (2) \quad \begin{array}{l} 2w_1 + w_2 + 3w_3 \leq 1 \\ 3w_1 + 2w_2 + w_3 \leq 1 \\ w_1 + 3w_2 + 2w_3 \leq 1 \end{array} \\
 (3) \quad Z_{\max} = w_1 + w_2 + w_3
 \end{array}$$

Die Lösung kann mit dem Simplex-Verfahren erfolgen:

	w_1	w_2	w_3	RS	
Z_{\max}	-1	-1	-1	0	
u_1	2	1	3	1	
u_2	3	2	1	1	
u_3	1	3	2	1	

$$w_1 = w_2 = w_3 = 1/6$$

$$u_1 = u_2 = u_3 = 1/6$$

Der Wert des Spiels beträgt: $\frac{1}{v(G_1)} = \frac{1}{2}$

Die optimalen Strategien $\mathbf{p}^* \in P$, $\mathbf{q}^* \in Q$ folgen aus den Beziehungen

$$\frac{p_i^*}{v(G_1)} = u_i \quad \text{bzw.} \quad \frac{q_j^*}{v(G_1)} = w_j$$

Man erhält:

$$p_i^* = 2 \cdot \frac{1}{6} = 1/3 \quad (i = 1, 2, 3)$$

$$q_j^* = 2 \cdot \frac{1}{6} = 1/3 \quad (j = 1, 2, 3)$$

Die optimalen Strategien sind

$$\mathbf{p}^* = (1/3, 1/3, 1/3) \quad \text{bzw.} \quad \mathbf{q}^* = (1/3, 1/3, 1/3)$$

Wegen der Äquivalenz der Spiele G und G_1 sind sie auch optimal in G .

Wert des Spiels (optimaler Gewinn):

$$E'(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^m \sum_{j=1}^n p_i \cdot F'_{ij} \cdot q_j$$

$$F'_{ij} = a \cdot F_{ij} + b$$

$$E'(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^m \sum_{j=1}^n p_i \cdot (a \cdot F_{ij} + b) \cdot q_j = a \cdot \sum_{i=1}^m \sum_{j=1}^n p_i \cdot F_{ij} + b \cdot \sum_{i=1}^m \sum_{j=1}^n p_i \cdot q_j = a \cdot E(\mathbf{p}, \mathbf{q}) + b$$

$$\text{d.h.: } v(G_1) = a \cdot v(G) + b$$

Mit $a = 1$ und $b = 2$:

$$v(G) = \frac{v(G_1) - b}{a} = 0$$

Das Ergebnis war zu erwarten, weil das Knobelspiel "Papier - Schere - Stein" ein symmetrisches Spiel ist. Jedes symmetrische Matrixspiel hat den Wert $v(G) = 0$ (faïres Spiel).

2. Skin-Spiel

Spieler 1 hat die Karten Pik As (PA), Karo As (KA) und Karo Zwei (KZ).

Spieler 2 hat die Karten Pik As (PA), Karo As (KA) und Pik Zwei (PZ).

Beide Spieler legen gleichzeitig eine Karte auf den Tisch. Haben die beiden Karten die gleiche Farbe gewinnt Spieler 1, sonst Spieler 2. Ein As hat den Wert 1, eine Zwei den Wert 2. Die Höhe des Gewinns richtet sich nach der Karte, die der Gewinner aufgedeckt hat.

Lösung:

Ermitteln der Gewinnmatrix

	PA=y ₁	KA=y ₂	PZ=y ₃	$\min_{j=1}^3 F_{ij}$
PA=x ₁	1	-1	1	-1
KA=x ₂	-1	1	-2	-2
KZ=x ₃	-1	2	-2	-2
$\max_{i=1}^3 F_{ij}$	1	2	1	

$$\max_{i=1}^3 \min_{j=1}^3 F_{ij} = -1 \text{ bzw. } \min_{j=1}^3 \max_{i=1}^3 F_{ij} = 1$$

Es existiert kein Sattelpunkt-

- optimale Strategie (gemischte Erweiterung)

Notwendig ist die Konstruktion eines äquivalenten Spiels:

$$F'_{ij} = a \cdot F_{ij} + b$$

Mit $a = 1$ und $b = 3$ erhält man die folgende Gewinnmatrix eines äquivalenten Spiels:

4	2	4
2	4	1
2	5	1

- Aufstellen des linearen Programms

$$(1) \quad u_i \geq 0$$

$$(2) \quad 4u_1 + 2u_2 + 2u_3 \geq 1$$

$$2u_1 + 4u_2 + 5u_3 \geq 1$$

$$4u_1 + u_2 + u_3 \geq 1$$

$$(3) \quad Z_{\min} = u_1 + u_2 + u_3$$

$$(1) \quad w_j \geq 0$$

$$(2) \quad 4w_1 + 2w_2 + 4w_3 \leq 1$$

$$2w_1 + 4w_2 + 1w_3 \leq 1$$

$$2w_1 + 5w_2 + w_3 \leq 1$$

$$(3) \quad Z_{\max} = w_1 + w_2 + w_3$$

- Lösung mit Hilfe des Simplex-Verfahrens

	w ₁	w ₂	w ₃	RS	
Z _{max}	-1	-1	-1		
u ₁				1	
u ₂				1	
u ₃				1	

Der Wert des Spiels beträgt $\frac{1}{v(G_1)} = 1/3$

$$v(G_1) = 3$$

- Optimale Strategien

$$\frac{p_i^*}{v(G_1)} = u_i \quad \text{bzw.} \quad \frac{q_j^*}{v(G_1)} = w_j$$

$$p_1 = 2/3, p_2 = 0, p_3 = 1/3$$

$$q_1 = 0, q_2 = 1/2, q_3 = 1/2$$

- Bemerkungen zur Lösung

Einzelne Strategien können von anderen dominiert werden, d.h.:

Keine Strategie ist bei keiner Gegenstrategie des Gegenspielers vorteilhaft gegenüber einer anderen Strategie.

Wendet man diese Aussage auf das vorliegende Beispiel an, so lassen sich folgende Vereinfachungen erzielen:

Spieler 1.

Die Strategie x_2 ist z.B. in keinem Fall vorteilhaft zur Strategie x_3 . Spieler 1 wird diese Strategie nicht auswählen.

Spieler 2.

Die Strategie y_1 ist in keinem Fall vorteilhaft zur Strategie y_3 . Spieler 2 wird Strategie y_1 nicht einschlagen.

Damit vereinfacht sich die Gewinnmatrix wesentlich:

-1	1
2	-2

bzw.

2	4
5	1

Das mathematische Modell für die das Spiel beschreibende lineare Planungsrechnung ist jetzt wesentlich einfacher:

$$(1) \ u_i \geq 0 \quad (i = 1, 2) \qquad (1) \ w_j \geq 0 \quad (j = 1, 2)$$

$$(2) \ 2u_1 + 5u_2 \geq 1 \qquad (2) \ 2w_1 + 4w_2 \leq 1$$

$$4u_1 + u_2 \geq 1 \qquad 5w_1 + w_2 \leq 1$$

$$(3) \ Z_{\min} = u_1 + u_2 \qquad (3) \ Z_{\max} = w_1 + w_2$$

Es ist sogar eine grafische Lösung möglich.

4. Graphen und Graphenalgorithmen

4.1 Theorie und praktische Bedeutung

4.1.1 Einführung

Viele Objekte und Vorgänge in verschiedenen Bereichen besitzen den Charakter eines Systems, d.h.: Sie setzen sich aus einer Anzahl von Bestandteilen, Elementen zusammen, die in gewisser Weise in Beziehung stehen. Sollen an einem solchen System Untersuchungen durchgeführt werden, dann ist es oft zweckmäßig, den Gegenstand der Betrachtungen durch ein graphisches Schema (Modell) zu veranschaulichen. Dabei stehen grundsätzlich immer 2 Elemente untereinander in Beziehung, d.h.: Die Theorie des graphischen Modells ist ein Teil der Mengenlehre, die binäre Relationen einer abzählbaren Menge mit sich selbst behandelt.

Bsp.: Es ist $K = \{A, B, C, D\}$ eine endliche Menge. Es ist leicht die Menge aller geordneten Paare von K zu bilden:

$K \times K =$

$\{(A,A),(A,B),(A,C),(A,D),(B,A),(B,B),(B,C),(B,D),(C,A),(C,B),(C,C),(C,D),(D,A),(D,B),(D,C),(D,D)\}$

Gegenüber der Mengenlehre ist die Graphentheorie nicht autonom.

Die Graphentheorie besitzt ein eigenes, sehr weites und spezifisches Vokabular. Sie umfaßt viele Anwendungsmöglichkeiten in der Physik, aus dem Fernmeldewesen und dem Operations Research. Im OR sind es vor allem Organisations- bzw. Verkehrs- und Transportprobleme, die mit Hilfe von Graphenalgorithmen untersucht und gelöst werden.

Generell dienen Graphenalgorithmen in der Praxis zum Lösen von kombinatorischen Problemen. Dabei geht man folgendermaßen vor:

1. Modelliere das Problem als Graph
2. Formuliere die Zielfunktion als Eigenschaft des Graphen
3. Löse das Problem mit Hilfe eines Graphenalgorithmus

4.1.2 Anwendungsgebiete aus dem Bereich des OR

Reihenfolgeprobleme

Hier umfaßt der Organisationsprozeß eine bestimmte Anzahl von Arbeitsgängen A, B, C, ..., die auf mehrere Arten hintereinander ausgeführt werden können, aber bestimmten technische Voraussetzungen erfüllen müssen. Für den Produktionsablauf muß eine im Hinblick auf eine gegebene Nutzenfunktion optimale Lösung gefunden werden.

Netzplantechnik

Unter Netzplantechnik (NPT) versteht man alle Verfahren zur Analyse, Beschreibung, Planung, Steuerung und Überwachung von Abläufen. Die Anwendungsbereiche sind vielfältig und umfassen:

- Forschungs- und Entwicklungsprojekte
- Fertigungsabläufe
- Wartungsarbeiten
- Bau- und Konstruktionsvorhaben im Schiffs-, Flugzeug-, Brückenbau
- Produkteinführungsvorhaben
- Vorhaben zur Entwicklung von Absatzstrategien.

Bei diesen Anwendungsfällen handelt es sich um Projekte. Der Begriff Projekt legt fest, daß eine bestimmte Leistungserstellung einmalig in ganz bestimmter Art und Weise durchgeführt wird (engl. Ausdruck „one-time-through“). Projekte sind in einzelne Teilabschnitte und Aktivitäten zerlegbar, die ablauftechnisch, zeitlich, kapazitäts- und kostenmäßig voneinander abhängen. Bei der Durchführung eines Projekts hat man für die einzelnen Schritte bestimmte Zeitspannen einzuhalten. Man will sich über den zeitlichen Spielraum informieren, über den man zu Beginn einer jeden Operation verfügt. Außerdem will man den Zeitpunkt wissen, ob das vollständige Programm nicht mehr zeitgerecht abgeschlossen werden kann. Es wird ein sogenannter „kritischer Weg“ bestimmt, auf den man alle Kraft konzentrieren muß, um die gegebenen Termine einzuhalten. Dieses Problem führt man auf das Auffinden eines Wegs mit einem maximalen Wert in einem Graphen zurück.

Die traditionellen Planungsverfahren (GANTT-Charts, Balkendiagramme) waren für komplexe Ablaufdiagramme, wie sie die bei Projekten vorliegen, nur bedingt geeignet.

Wegen der nur bedingten Eignung der Gantt-Diagramme entstanden 1957/58 unabhängig voneinander auf der Basis der Graphentheorie die 3 grundlegenden Verfahren der Netzplantechnik (NPT):

CPM (Critical Path Method)

Sie wurde 1957 in den USA beim Chemiekonzern DuPont de Nemours für Umstellungs- und Installationsarbeiten entwickelt

PERT (Program Evaluation and Review Technique)

Sie wurde 1958 beim Special Force Office der US - Navy zur Entwicklung der Polaris-Rakete benutzt.

MPM (Metra Potential Method)

Sie wurde 1958 in Frankreich für den Bau von Atomkraftwerken entwickelt.

4.1.3 Definition und Darstellung von Graphen

1 Ein einführendes Beispiel

Es ist $K = \{A, B, C, D\}$ eine endliche Menge.

Es ist leicht die Menge aller geordneten Paare von K zu bilden:

$$K \times K = \{(A,A), (A,B), (A,C), (A,D), (B,A), (B,B), (B,C), (B,D), (C,A), (C,B), (C,C), (C,D), (D,A), (D,B), (D,C), (D,D)\}$$

Die Menge dieser Paare kann auf verschiedene Arten dargestellt werden.

Koordinatendarstellung

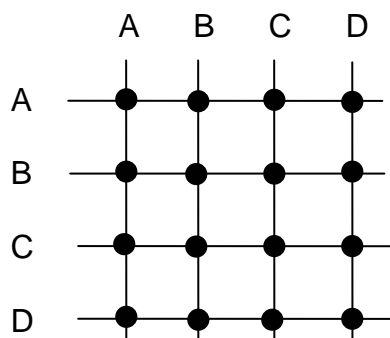


Abb.:

Darstellung durch Punkte (Kreise) und Kanten (**ungerichteter Graph**)

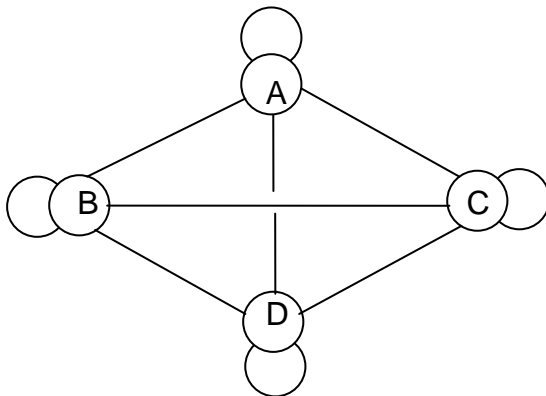


Abb.:

Eine Kante (A,A) nennt man **Schlinge**

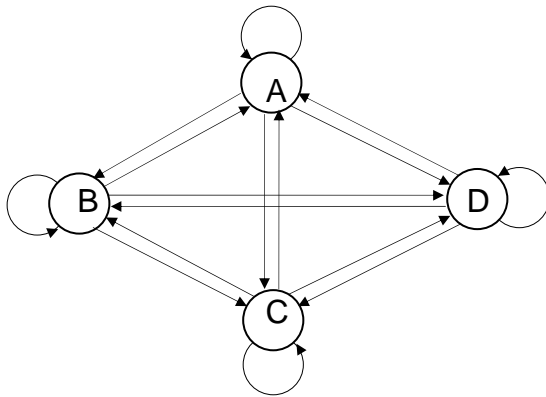
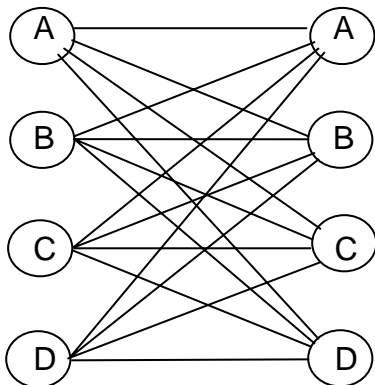
Darstellung durch Punkte (Kreise) und Pfeile (**gerichteter Graph**)

Abb.:

Einen Pfeil (A,A) nennt man eine **Schlinge**. Zwei Pfeile mit identischem Anfangs- und Endknoten nennt man parallel. Analog lassen sich parallele Kanten definieren. Ein Graphen ohne parallele Kanten bzw. Pfeile und ohne Schlingen bezeichnet man als schlichte Graphen.

Darstellung durch paarweise geordnete Paare



Abb

Darstellung mit Hilfe einer Matrix

	A	B	C	D
A	1	1	1	1
B	1	1	1	1
C	1	1	1	1
D	1	1	1	1

Einige der geordneten Paare aus der Produktmenge $K \times K$ sollen eine bestimmte Eigenschaft haben, während die anderen sie nicht besitzen.

Eine solche Untermenge von $K \times K$ ist :

$$G = \{(A,B), (A,D), (B,B), (B,C), (B,D), (C,C), (D,A), (D,B), (D,C), (D,D)\}$$

Üblicherweise wird diese Untermenge (Teilgraph) so dargestellt:

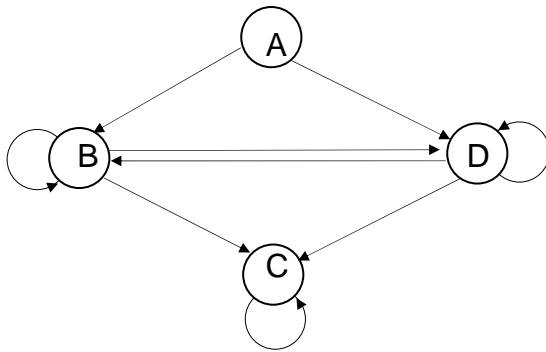


Abb.:

Betrachtet man hier die Paare z.B. (A,B) bzw. (A,D), so kann man feststellen: Von A erreicht man, den Pfeilen folgend, direkt B oder D. B und D heißt auch die "**Inzidenzabbildung**" von A und {B,D} das volle Bild von A.

Verwendet man das Symbol Γ zur Darstellung des vollen Bilds, dann kann man das vorliegende Beispiel (vgl. Abb. :) so beschreiben:

$$\Gamma(A) = (B, D)$$

$$\Gamma(B) = (B, C, D)$$

$$\Gamma(C) = C$$

$$\Gamma(D) = (A, B, C, D)$$

2. Bezeichnungen, Definitionen

Gegeben ist eine endliche (nicht leere) Menge K^8 . Ist G eine Untermenge der Produktmenge $K \times K$, so nennt man ein Element der Menge K einen Knoten von G . Die Elemente der Knotenmenge K können auf dem Papier durch Punkte (Kreise) markiert werden.

Ein Element von G selbst ist eine (gerichtete) Kante. Im vorstehenden Bsp⁹ sind (A,B), (A,D), (B,B), (B,C), (B,D), (C,C), (D,A), (D,B), (D,C), (D,D) (gerichtete) Kanten. Ein Graph wird durch die Menge seiner Knoten K und die seiner Inzidenzabbildungen beschrieben: $G=(K,\Gamma)$

Ein Graph kann aber auch folgendermaßen beschrieben werden: $G = (K,E)$. E ist die Menge der Kanten (gerichtet, ungerichtet, gewichtet). In gewichteten Graphen werden jeder Kante ganze Zahlen (Gewichte, z.B. zur Darstellung von Entfernungen oder Kosten) zugewiesen. Gewichtete gerichtete Graphen werden auch Netzwerke genannt.

Falls die Anzahl der Knoten in einem Graphen " n " ist, dann liegt die Anzahl der Kanten zwischen 0 und $\frac{n \cdot (n-1)}{2}$ im ungerichteten Graphen. Ein gerichteter Graph

kann bis zu $n \cdot (n-1)$ Pfeile besitzen.

Eingangsgrad: Zahl der ankommenden Kanten.

⁸ Anstatt K schreibt man häufig auch V (vom englischen Wort Vertex abgeleitet)

⁹ vgl. 5.1.1

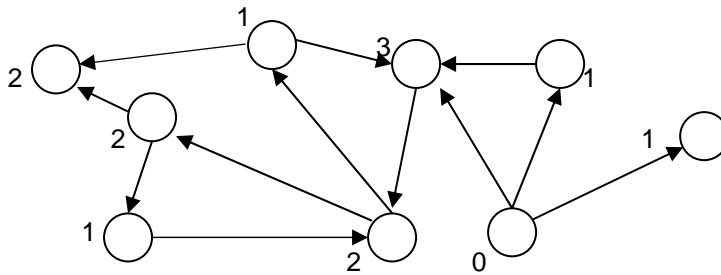


Abb.: Eingangsgrad

Ausgangsgrad: Zahl der abgehenden Kanten

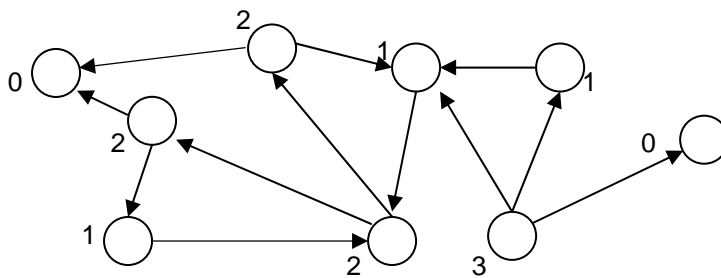


Abb.: Ausgangsgrad

Bei ungerichteten Graphen ist der Ausgangsgrad gleich dem Eingangsgrad. Man spricht dann nur von Grad.

Ein **Pfad** vom Knoten K_1 zum Knoten K_k ist eine Folge von Knoten K_1, K_2, \dots, K_k , wobei $(K_1, K_2), \dots, (K_{k-1}, K_k)$ Kanten sind. Die Länge des Pfads ist die Anzahl der Kanten im Pfad. Auch Pfade können gerichtet oder ungerichtet sein.

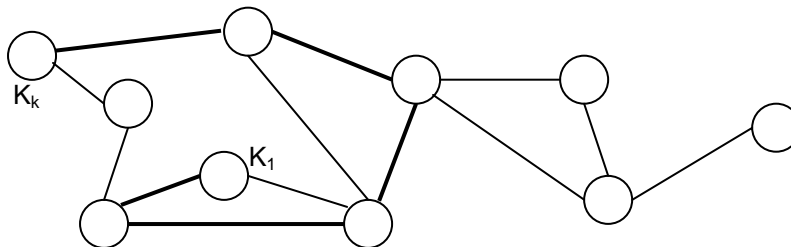


Abb.:

Ein Graph ist **zusammenhängend**, wenn von jedem Knoten zu jedem anderen Knoten im Graph ein Weg (Pfad) existiert.

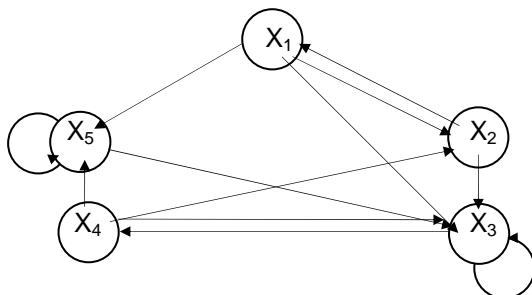


Abb.:

Dieser Graph ist streng zusammenhängend. Man kann sehen, daß es zwischen je 2 Knoten mindestens einen Weg gibt. Dies trifft auf den folgenden Grafen nicht zu:

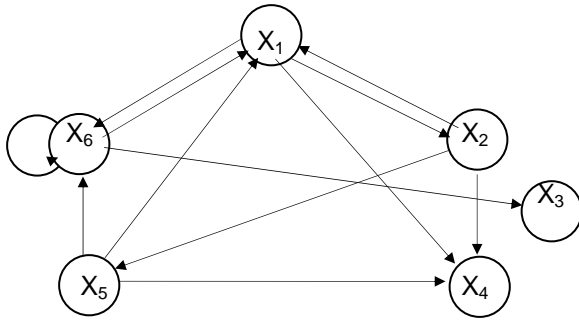
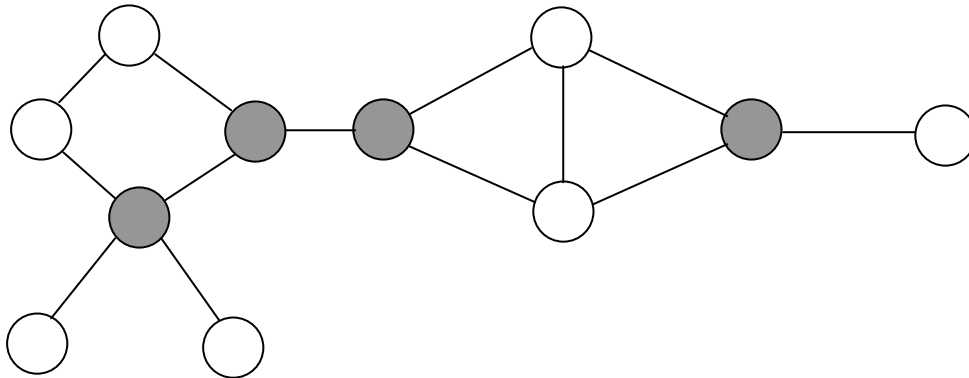


Abb.:

Hier gibt es bspw. keinen Weg von X_4 nach X_1 .

Ein Graph, der nicht zusammenhängend ist, setzt sich aus zusammenhängenden Komponenten zusammen.

Ein Knoten in einem zusammenhängenden Netzwerk heißt **Artikulationspunkt**, wenn durch sein Entfernen der Graph zerfällt, z.B.



Artikulationspunkte sind dunkel eingefärbt.

Abb.:

Erreichbarkeit: Der Knoten B ist in dem folgenden Graphen erreichbar vom Knoten G, wenn es einen Pfad von G nach B gibt.

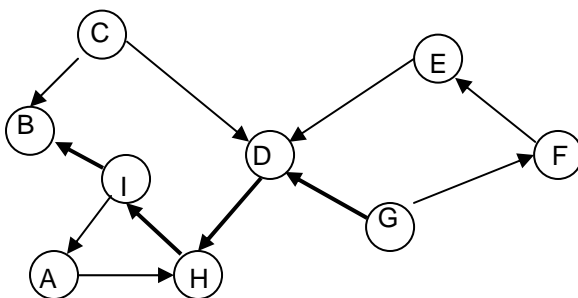


Abb.:

Ein Graph heißt **Zyklus**, wenn sein erster und letzter Knoten derselbe ist.

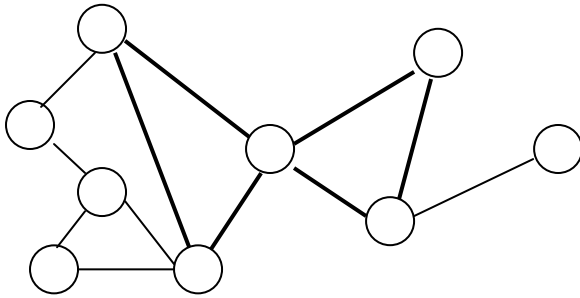


Abb.: Zyklus (manchmal auch geschlossener Pfad genannt)

Ein Zyklus ist ein **einfacher Zyklus**, wenn jeder Knoten (außer dem ersten und dem letzten) nur einmal vorkommt.

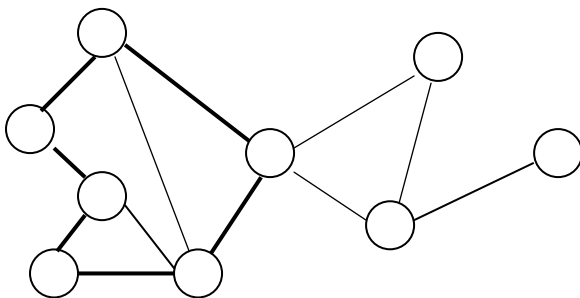


Abb.: Einfacher Zyklus (manchmal auch geschlossener Pfad genannt)

Ein gerichteter Graph heißt **azyklisch**, wenn er keine Zyklen enthält. Ein azyklischer Graph kann in Schichten eingeteilt werden (Stratifikation).

Bäume sind Graphen, die keine Zyklen enthalten. Graphen, die keine Zyklen enthalten heißen Wald. Zusammenhängender Graphen, die keine Zyklen enthalten, heißen Bäume. Wenn ein gerichteter Graph ein Baum ist und genau einen Knoten mit Eingangsgrad 0 hat, heißt er Baum mit Wurzel.

Ein **spannender Baum** (Spannbaum) eines ungerichteten Graphen ist ein Teilgraph des Graphen, und ist ein Baum der alle seine Knoten enthält.

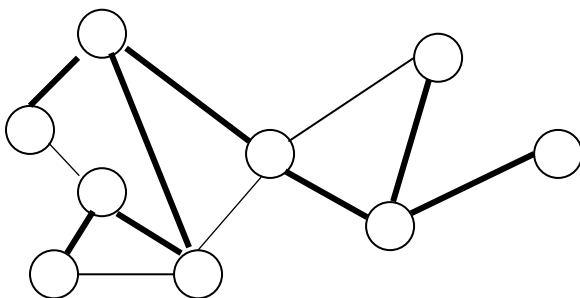
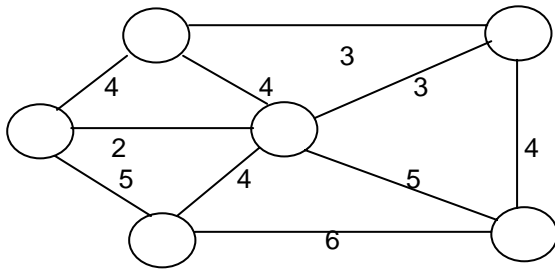


Abb.

Das Entfernen einer beliebigen Kante in einem Spannbaum bewirkt, daß ein nicht zusammenhängender Graph entsteht.

Einen spannenden Baum mit minimaler Summe der Kantenbewegungen bezeichnet man als minimalen spannenden Baum. Zu dem folgenden Graphen



gehört der folgende minimale spannende Baum

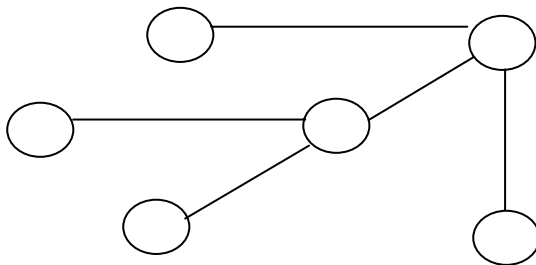


Abb.: Minimaler spannender Baum

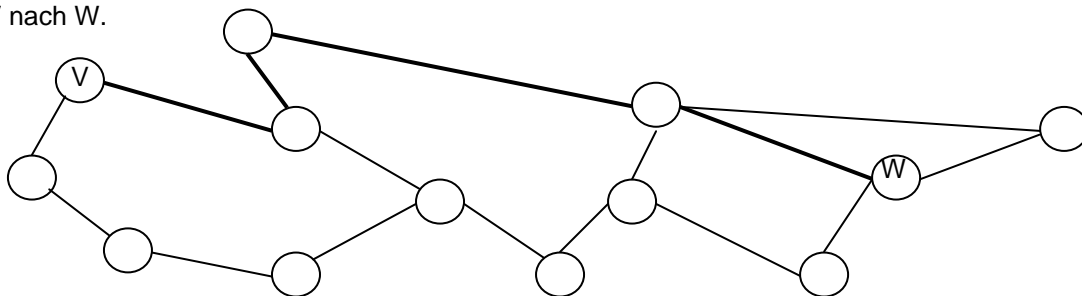
3. Grundtypen bei der Zielfindung

Bei der Zielfindung kann man folgende Grundtypen unterscheiden:

- Kanten auswählen

-- Shortest Path

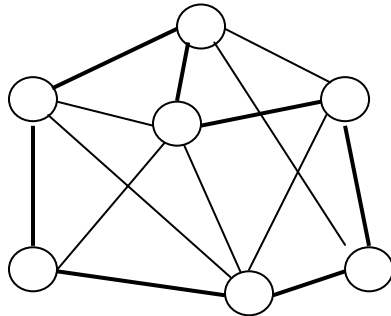
Gegeben ist ein Graph, zwei Knoten V und W. Gesucht ist der kürzeste Weg (Anzahl Kanten) von V nach W.



Verallgemeinerung: Pfadlänge nicht durch Kantenzahl bestimmt, sondern durch die Summe der Kantenmarkierungen (, die alle positiv sein sollen).

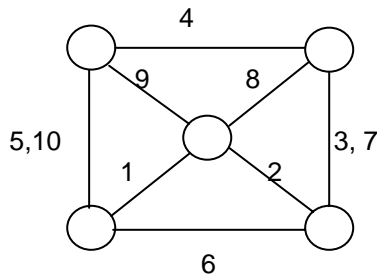
-- Hamiltonsche Pfade

Möglicher Ausgangspunkt: Gegeben ist eine Landkarte mit Orten und Verbindungen. Gesucht ist ein Rundgang einmal durch jeden Ort.



Verschärfung: jede Verbindung ist mit Kosten gewichtet. Gesucht ist der billigste Rundgang. Ein Hamiltonscher Pfad ist ein einfacher Zyklus, der jeden Knoten eines Graphen enthält. Ein Algorithmus für das Finden eines Hamiltonschen Graphen ist relativ einfach (modifizierte Tiefensuche) aber sehr aufwendig. Bis heute ist kein Algorithmus bekannt, der eine Lösung in polynomialer Zeit findet.

-- Chinese Postman



Die an den Kanten angegebenen Zahlen geben den jeweiligen Schritt beim Durchlaufen der Kanten aus. Startpunkt ist links unten.

Ausgangspunkt dieses Problems ist das sog. Königsberger Brückenproblem, das Leonard Euler 1736 gelöst hat. Euler interpretierte dabei die Brücken über den Fluß Pregel in Königsberg als Kanten und Ufer bzw. Inseln als Knoten.

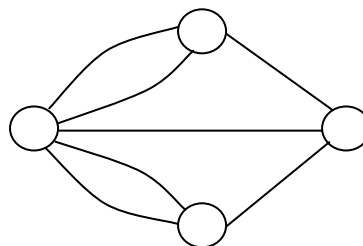
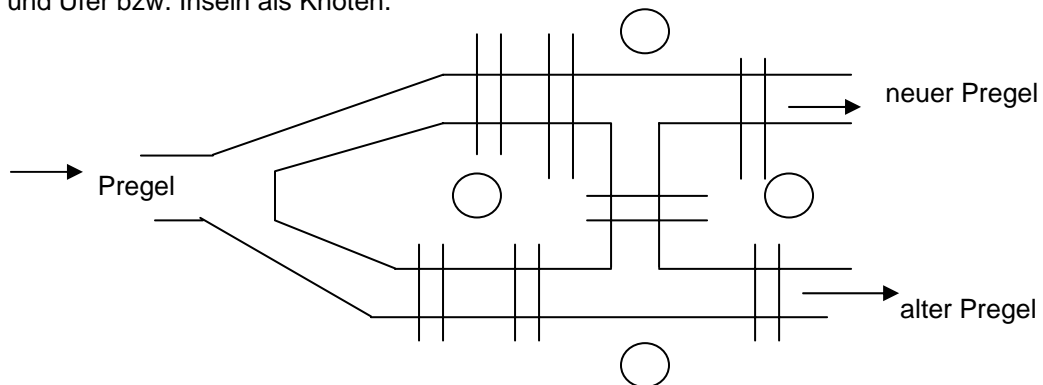
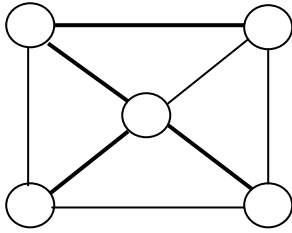


Abb.: Königsberger Brückenproblem mit Darstellung als Graph

Neuformulierung des Problems: Gibt es einen Zyklus im Graphen, der alle Kanten genau einmal enthält (Eulerkreis).

-- Spanning Tree



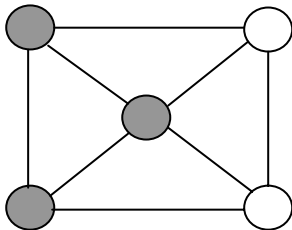
Problem: Gegeben ist eine Landkarte als Graph

- Es soll ein Wasserleitungsnetz aufgebaut werden, dass alle Ortschaften versorgt
- die Gesamtkosten sind proportional zur Summe der Länge aller vorkommenden Leitungen
- Welche Ortschaften sollen miteinander verbunden werden? (Teilgraph von G , so dass die Kosten minimal werden)

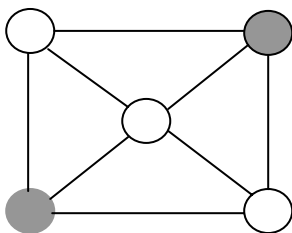
Ähnliche Probleme gibt es beim Entwurf von Rechnernetzen.

- Knoten auswählen

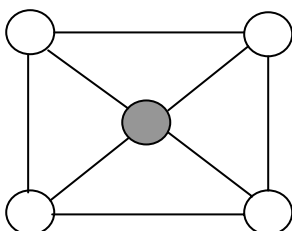
-- größte Clique finden



-- Independent Set



-- Zentrum finden



- Kanten Werte zuordnen

-- Maximum Flow

In einem Graphen versucht man unter Einhaltung von Kantenkapazitäten möglichst viele Einheiten von einer Quelle zu einer Senkte fließen zu lassen.

-- Minimum Cost Flow

Hier wird versucht bestimmte Nachfragen im Knoten durch Angebote in anderen Knoten durch einen Fluß zu befriedigen und die dabei entstehenden Kosten minimal zu halten.

- Einfärben von Graphen

Einfärben von Knoten derart, dass keine zwei benachbarten Knoten dieselbe Farbe haben.

4.1.4 Darstellung in Rechnerprogrammen

1. Der abstrakte Datentyp (**ADT**) für gewichtete Graphen

Ein gewichteter Graph besteht aus Knoten und gewichteten Kanten. Der **ADT** beschreibt die Operationen, die einem solchen gewichteten Graphen Datenwerte hinzufügen oder löschen. Für jeden Knoten K_i definiert der **ADT** alle benachbarten Knoten, die mit K_i durch eine Kante $E(K_i, K_j)$ verbunden sind.

ADT Graph

Daten

Sie umfassen eine Menge von Knoten $\{K_i\}$ und Kanten $\{E_i\}$. Eine Kante ist ein Paar (K_i, K_j) , das anzeigt: Es gibt eine Verbindung vom Knoten K_i zum Knoten K_j . Verbunden ist mit jeder Kante die Angabe eines Gewichts. Es bestimmt den Aufwand, um entlang der Kante vom Knoten K_i nach dem Knoten K_j zu kommen.

Operationen

Konstruktor

Eingabe: keine

Verarbeitung: Erzeugt den Graphen als Menge von Knoten und Kanten

Einfuegen_Knoten

Eingabe: Ein neuer Knoten

Vorbedingung: keine

Verarbeitung: Füge den Knoten in die Menge der Knoten ein

Ausgabe: keine

Nachbedingung: Die Knotenliste nimmt zu

Einfügen_Kante

Eingabe: Ein Knotenpaar K_i und K_j und ein Gewicht

Vorbedingung: K_i und K_j sind Teil der Knotenmenge

Verarbeitung: Füge die Kante (K_i, K_j) mit dem gewicht in die Menge der Kanten ein.

Ausgabe: keine

Nachbedingung: Die Kantenliste nimmt zu

Loesche_Knoten

Eingabe: Eine Referenz für den Knoten K_1

Vorbedingung: Der Eingabewert muß in der Knotenmenge vorliegen

Verarbeitung: Lösche den Knoten aus der Knotenliste und lösche alle Kanten der Form (K, K_1) bzw. (K_1, K) , die eine Verbindung mit Knoten K_1 besitzen

Loesche_Kante

Eingabe: Ein Knotenpaar K_i und K_j

Vorbedingung: Der Eingabewert muß in der Kantenliste vorliegen

Verarbeitung: Falls (K_i, K_j) existiert, lösche diese Kante aus der

Kantenliste

Ausgabe: keine

Nachbedingung: Die Kantenmenge wird modifiziert

Hole_Nachbarn:

Eingabe: Ein Knoten K

Vorbedingung: keine

Verarbeitung: Bestimme alle Knoten K_n , so daß (K, K_n) eine Kante ist

Ausgabe: Liste mit solchen Kanten

Nachbedingung: keine

Hole_Gewichte

Eingabe: Ein Knotenpaar K_i und K_j

Vorbedingung: Der Eingabewert muß zur Knotenmenge gehören

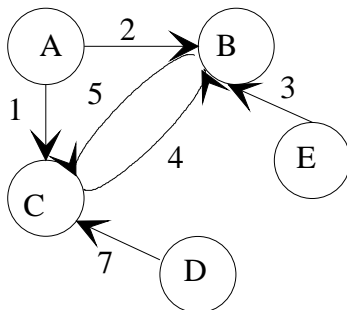
Verarbeitung: Beschaffe das Gewicht der Kante (K_i, K_j) , falls es existiert

Ausgabe: Gib das Gewicht dieser Kante aus (bzw. Null, falls die Kante nicht existiert)

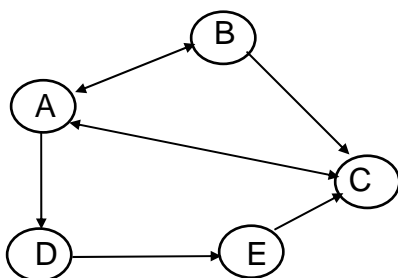
Nachbedingung: keine

2. Abbildung der Graphen

Es gibt zahlreiche Möglichkeiten zur Abbildung von Knoten und Graphen in einem Rechnerprogramm. Eine einfache Abbildung speichert die Knoten in einer sequentiellen Liste. Die Kanten werden in einer Matrix beschrieben (Adjazenzmatrix), in der Zeile i bzw. Spalte j den Knoten K_i und K_j zugeordnet sind. Jeder Eintrag in der Matrix gibt das Gewicht der Kante $E_{ij} = (K_i, K_j)$ oder den Wert 0 an, falls die Kante nicht existiert. In ungewichteten, gerichteten Graphen hat der Eintrag der (booleschen) Wert 0 oder 1, je nachdem, ob die Kante zwischen den Knoten existiert oder nicht, z.B.:



0	2	1	0	0
0	0	5	0	0
0	4	0	0	0
0	0	7	0	0
0	3	0	0	0

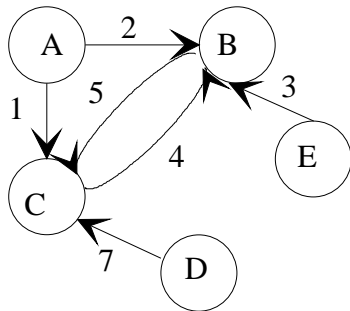


```

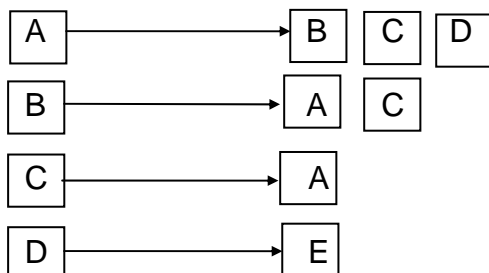
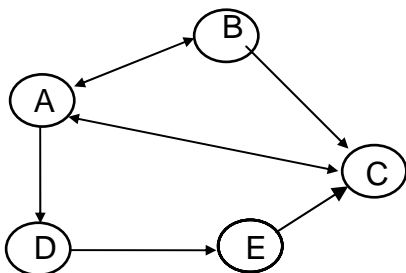
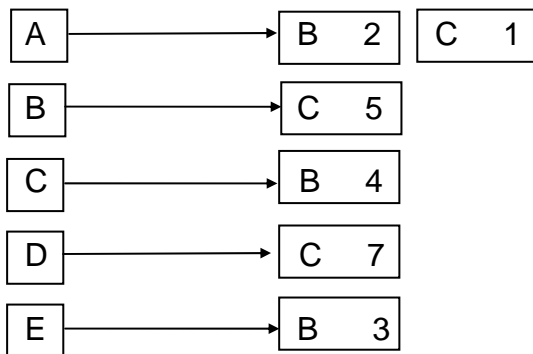
0  1  1  1  0
1  0  1  0  0
1  0  0  0  0
0  0  0  0  1
0  0  1  0  0
    
```

Abb.: Einfache Darstellung eines Graphen in einer Adjazenzmatrix

In der Darstellungsform „Adjazenzstruktur“ werden für jeden Knoten alle mit ihm verbundenen Knoten in eine **Adjazenzliste** für diese Knoten aufgelistet. Das lässt sich leicht über verkettete Listen realisieren. In einem gewichteten Graph kann zu jedem Listenelement ein Feld für das Gewicht hinzugefügt werden, z.B.:



Knoten: Liste der Nachbarn:



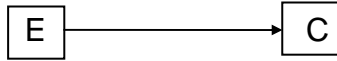


Abb.: Darstellung eines Graphen in einer Adjazenzliste

Zur Implementierung besteht die Möglichkeit, Knotennamen in eine sequentiell gespeicherte Liste bzw. in eine Hash-Tabelle bzw. in einen biären Baum einzulesen und jedem Knoten eine ganze Zahl zuzuweisen. Dies wird benutzt, um auf knotenindizierte Felder von der Art der Adjazenzmatrix zuzugreifen.

3. Typische Graphemalgorithmen

- Feststellen bestimmter Eigenschaften (z.B. zusammenhängend) über Knoten und Kanten
- Finden von explizit¹⁰ definierten Pfaden
- Finden von Pfaden in implizit¹¹ definierten Pfaden, die zusätzlichen Randbedingungen genügen (z.B. Optimalität bzgl. eines Bewertungskriteriums).
- Finden von Untergraphen, die bestimmten Randbedingungen erfüllen (z.B. minimal spannender Baum).

4. Lösungsstrategien

Für die Lösung der Graphenprobleme stattet man die Algorithmen mit verschiedenen Strategien aus:

- Greedy (sukzessive bestimmung der Lösungsvariablen)
- Divide and Conquer (Aufteilen, Lösen, Lösungen vereinigen)
- Dynamic Programming (Berechne Folgen von Teillösungen)
- Enumeration (Erzeuge alle Permutationen und überprüfe sie)
- Backtracking (Teillösungen werden systematisch erweitert)
- Branch and Bound (Erweitere Teillösungen an der vielversprechenden Stelle)

¹⁰ explizit: im Graphen als Kanten repräsentiert

¹¹ implizit: nicht direkt repräsentiert, aber erreichbar

4.2 Grundlegende Algorithmen

4.2.1 Die Ermittlung kürzester bzw. längster Wege nach dem Algorithmus von Ford

1. Welches Problem löst dieser Algorithmus?

Er löst das Problem, einen Weg mit minimalem (oder maximalem) Wert zwischen 2 Knoten eines Graphen zu finden. Anstatt von minimalen Wegen spricht man häufig auch von kürzesten Wegen. Unter der Länge eines Wegs wird in diesem Fall die $\sum w(e)$ der den Pfeilen des Weges zugeordneten Bewertungen (Gewichte) $w(e)$ verstanden. Es wird nach dem Weg zwischen den Knoten X_0 und X_n gesucht, für den die Gesamtbewertung möglichst klein ist. Ein Streckenwert $w(e)$ kann bspw. Kosten, Zeit oder Länge symbolisieren.

2. Beschreibung des Algorithmus

a) Bestimmen eines Wegs mit minimalem Wert

X_0 ist der Ausgangsknoten. Von ihm ausgehend sind die Weglängen zu anderen Knoten auszurechnen, X_0 erhält am Anfang den Wert $l_0 = 0$, alle anderen Knoten X_i den Wert $l_i = \infty$ zugeordnet.

Ist l_i der Wert von X_i , so sucht man eine gerichtete Kante mit der Eigenschaft:

$$l_i - l_j > w(X_i, X_j)$$

$w(X_i, X_j)$: Wert der gerichteten Kante

Es ist eine Strecke (X_i, X_j) zu bestimmen, für die $l_j - l_i > w(X_i, X_j)$ zutrifft. l_i ist danach durch $l'_i = l_i + w(X_i, X_j) < l_j$ zu ersetzen. Dabei ist $l_j > 0$, falls $j \neq 0$ ist. Der Vorgang ist solange zu wiederholen, bis es durch keine direkte gerichtete Kante mehr möglich ist, die l_i zu verringern.

Man hat danach die minimalen Werte, die von X_0 ausgehen und zu anderen Knoten führen, erhalten.

Bsp.: In einer Stadt wird die kürzeste Verbindung zwischen 2 Kreuzungen gesucht. Die mittlere Fahrdauer auf den möglichen Verbindungen zwischen den einzelnen Kreuzungen ist bekannt.

Wie kann man einen minimalen Weg (z.B. zwischen "S" und "Z") errechnen?

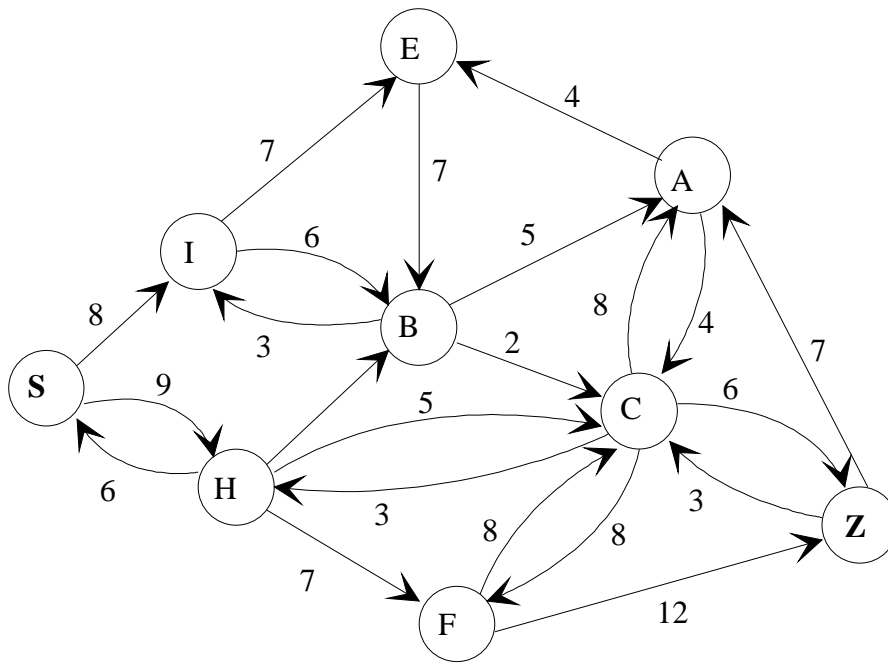


Abb.:

Zum Knoten „S“ schreibt man 0, zu allen anderen ∞ . Man erhält dann fortlaufend:

$$\begin{aligned}
 l_I &= \infty, l_I' = 0 + 8 = 8 < \infty \\
 l_E &= \infty, l_E' = 8 + 7 = 15 < \infty \\
 l_B &= \infty, l_B' = 15 + 7 = 22 < \infty \\
 l_A &= \infty, l_A' = 22 + 5 = 27 < \infty \\
 l_H &= \infty, l_H' = 0 + 9 = 9 < \infty \\
 l_C &= \infty, l_C' = 22 + 2 = 24 < \infty \\
 l_F &= \infty, l_F' = 9 + 7 = 16 < \infty \\
 l_Z &= \infty, l_Z' = 16 + 12 = 28 < \infty
 \end{aligned}$$

$$\begin{aligned}
 l_I &= 8 \\
 l_E &= 15 \\
 l_B &= 22, l_B' = 8 + 6 = 14 < 22 \text{ oder } l_B' = 9 + 5 = 14 < 22 \text{ (In B wird 22 durch 14 ersetzt)} \\
 l_A &= 27, l_A' = 14 + 5 = 19 < 27 \\
 l_H &= 9, l_H' = 8 + 10 = 18 > 9 \text{ (H bleibt unverändert)} \\
 l_C &= 24, l_C' = 14 + 2 = 16 < 24, l_C' = 9 + 5 = 14 < 24 \\
 l_F &= 16 \\
 l_Z &= 28, l_Z' = 14 + 6 = 20 < 28
 \end{aligned}$$

Wenn man auf diese Weise fortfährt, dann stellt man fest, daß in keinem Knoten der Wert weiter reduziert werden kann. Man hat damit den optimalen Weg gefunden.

b) Bestimmen eines Wegs mit maximalem Wert

- Am Anfang ist $l_0 = 0$ und $l_i = 0$ zu setzen.
- Bestimmen einer gerichteten Strecke (X_i, X_j) für die $l_j - l_i < w(X_i, X_j)$ zutrifft. Dann ist l_j durch $l'_j = l_i + w(X_i, X_j) > l_j$ zu ersetzen.
- Fortsetzen dieser Verfahrensweise bis keine weitere Vergrößerung der l_j mehr möglich ist.

c) Zusammenfassung

In der Praxis wird das Verfahren von Ford, falls der Graph zyklensfrei ist, folgendermaßen abgeändert:

1. Suche des Weges mit dem maximalen Wert

$$l_0 = 0$$

.....

$$l_j = \max_{(X_i, X_j) \in E_j^-} (l_i + w(X_i, X_j))$$

E_j^- : Menge, der in X_j einmündenden Strecken des Graphen.

2. Suche des Weges mit minimalem Wert

$$l_0 = 0$$

.....

$$l_j = \min_{(X_i, X_j) \in E_j^-} (l_i + w(X_i, X_j))$$

Unter allen in X_j einmündenden Strecken (X_i, X_j) ist diejenige mit dem kleinsten Wert $l_i + w(X_i, X_j)$ zu nehmen.

Das dem Knoten X_n zugeordnete l_n ist gleich dem Wert des optimalen Weges zwischen X_0 und X_n .

3. PERT Fordsches Verfahren

Gegeben ist der folgende zyklenfreie Graph:

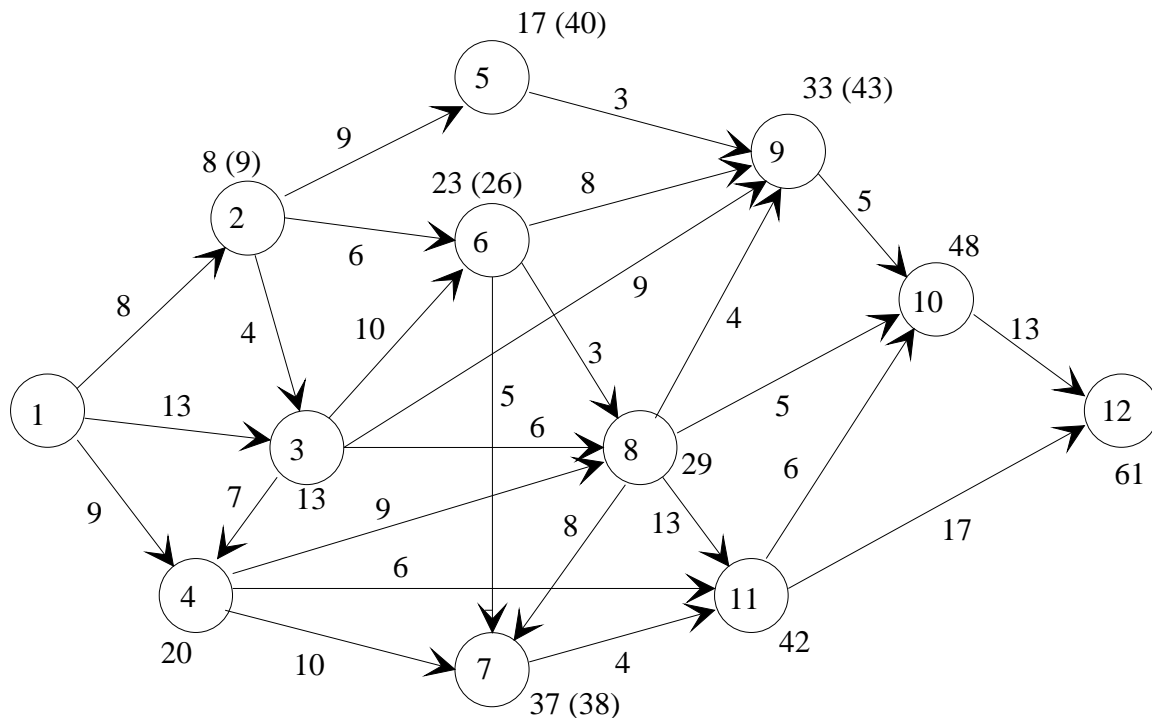


Abb.:

Der vorliegende Graph könnte einen Netzplan beschreiben, z.B. ein Arbeitsprojekt mit einer bestimmten Anzahl von Tätigkeiten. Die Knoten symbolisieren die Ereignisse. Je nachdem, ob ein Knoten ein Anfangs- oder Endknoten eines Pfeils ist, ist als Ereignis der Beginn oder der Abschluß der dem Pfeil entsprechenden Tätigkeit zu deuten.

Den Zeitpunkt, zu dem das Ereignis E_{12} eintreten kann, erhält man folgendermaßen:

Vom Anfangsereignis E_1 sucht man den Weg nach E_{12} mit der maximalen Zeitsumme. Diese Summe ist der gesuchte Zeitpunkt. Da sie maximal ist, besteht Gewißheit, daß für sämtliche Tätigkeiten genügend Zeit zur Durchführung bereit steht.

Allgemein läßt sich so der Zeitpunkt des frühestens Eintreffens von einem Ereignis E_i als Summe aller Streckenwerte des Weges mit dem maximalen Wert zwischen E_1 und E_i bestimmen (frühester Eintrittstermin).

Bestimmen der frühesten Eintrittstermine nach dem Fordschen Verfahren:

Knoten 1 erhält die Zeit 0 zugeordnet.

Zum Knoten 2 führt ein einziger Pfeil, E_2 erhält den frühesten Termin 8 zugeordnet (Diese Zahl wird neben dem Knoten eingetragen).

Zum Knoten 3 führen 2 Pfeile (1,3) und (2,3). Es ist zu vergleichen: $8 + 4 = 12$ und 13. 13 wird eingetragen.

Zum Knoten 4 führen 2 Pfeile (1,4) und (3,4). Es ist zu vergleichen: 9 und $(13 + 7) = 20$. 20 wird neben dem Knoten 4 eingetragen.

Auf diese Weise kann man schließlich am Knoten „12“ 61 eintragen.

Bestimmen der spätesten Eintrittstermine:

Zum spätesten Eintrittstermin ist der Beginn der Tätigkeiten einzuleiten, damit das Gesamtprojekt nicht verzögert wird. Zwischen Ereignis E_i und E_{12} existiert ein Weg mit der maximalen Zeitsumme. Dieser stellt die erforderliche optimale Zeit für die Ausführung aller Tätigkeiten zwischen E_i und E_{12} dar. Der gesuchte späteste Eintrittstermin ergibt sich aus der Differenz:

(Minimale Zeit für die Erledigung aller Tätigkeiten zwischen E_i und E_{12}) - Endtermin von E_{12} (= 61))

Knoten 12 wird mit 61 markiert

Vom Knoten 11 gehen 2 Pfeile (11,12) und (11,10) aus. Von den Werten $61 - 17 = 44$ und $48 - 6 = 42$ ist als spätester Wert von E_{11} der kleinere Wert 42 zu nehmen.

Vom Knoten 9 geht ein einzelner Pfeil aus, der späteste Termin ist also $48 - 5 = 43$.

Von Knoten 7 geht nur der Pfeil (7,11) aus, der späteste Termin für E_7 ist $42 - 4 = 38$.

Vom Knoten 8 gehen 4 Pfeile aus (8,9), (8,10), (8,11), (8,7). Von den 4 Werten ($43 - 4 = 39$, $48 - 5 = 43$, $42 - 13 = 29$ und $38 - 8 = 30$) wird 29 als der kleinste Wert dem Knoten 8 zugeordnet.

Auf diese Weise erreicht man schließlich den Knoten 1.

Je Ereignis wurden 2 Termine erhalten:

- frühester Eintrittstermin t_i des Ereignisses E_i (bezogen auf $t_i = 0$)
- spätester Eintrittstermin t_i^* des Ereignisses E_i , der nicht zu einer Verzögerung des Projektabschlusses führt.

Die Differenz der beiden Termine wird als gesamte **Pufferzeit** bezeichnet. Für bestimmte Ereignisse fallen die Termine zusammen. Derartige Ereignisse heißen **kritisch**.

4.2.2 Netzplantechnik

4.2.2.1 Grundlagen

Ein Netzplan wird durch 3 Strukturelemente beschrieben:

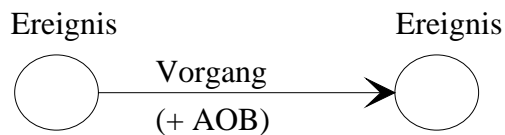
1. Als **Ereignisse** bezeichnet man die einzelnen Zustände im Projektablauf, die durch den Beginn oder den Abschluß bestimmter Aktivitäten gekennzeichnet sind.
2. Als **Vorgänge** bezeichnet man die einzelnen Aktivitäten (**Tätigkeiten**). Sie beanspruchen Zeit und Kapazität und verursachen Kosten. Der Begriff ist sehr weit zu fassen (personelle und maschinelle Arbeiten, Beratungen, Lieferzeiten, Wartezeiten).
3. Als **Anordnungsbeziehungen** bezeichnet man die strukturelle und zeitliche Aufeinanderfolge zweier Vorgänge.

Ereignisse, Vorgänge und Anordnungsbeziehungen (AOB) können auf verschiedene Weise im Netzplan dargestellt werden.

In **Vorgangspfeil-Netzplänen** (VPN) werden Vorgänge durch Pfeile und Ereignisse durch Knoten dargestellt. Die Pfeile werden mit der Vorgangsdauer bewertet.

Eine solche Struktur liegt z.B. **CPM** (Vorgangspfeil-Netzplan) und **PERT** (Ereignisknoten-Netzplan) zugrunde. CPM ist vorgangsorientiert, PERT ereignisorientiert.

Vorgangspfeil-Netzplan (kantenorientierter Netzplan)



Ereignisknoten-Netzplan



Abb.:

Vorgangsknoten-Netzplan (knotenorientierter Netzplan)

Vorgänge werden beschrieben und durch Knoten dargestellt. Kanten beschreiben die Anordnungsbeziehungen (mit Zeitwerten) der Vorgänge. Ereignisse werden nicht dargestellt

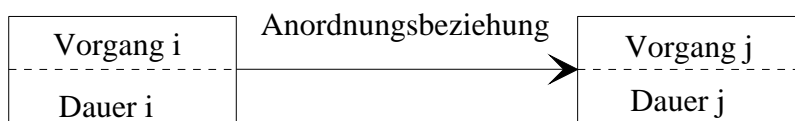
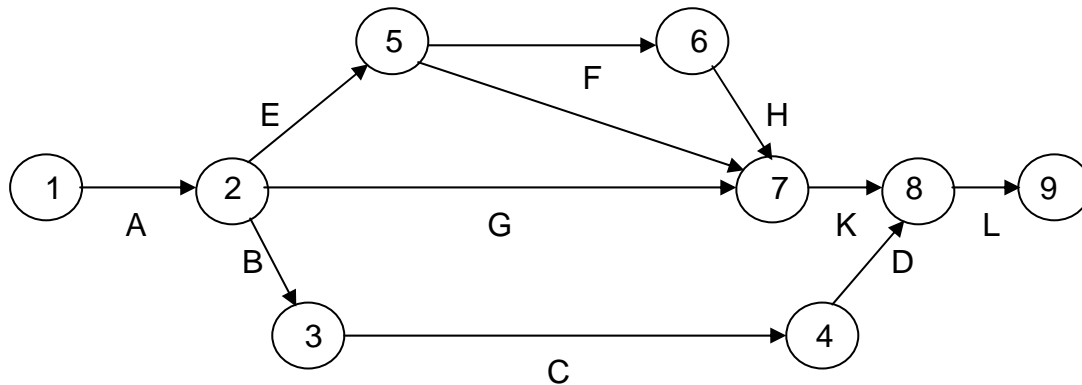


Abb.:

4.2.2.2 CPM

1 Grundbegriffe

Einführendes Beispiel: „Netzplan einer Motorüberholung“



- A ... Motor ausbauen
- B ... Zylinderkopf zerlegen
- C ... Ventile einschleifen
- D ... Kopf zusammenbauen
- E ... Kurbelwelle und Kolben ausbauen
- F ... Kurbelwelle schleifen
- G ... Zylinder schleifen
- H ... Pleuellager erneuern
- I ... Kolben ersetzen
- K ... Motorblock zusammenbauen
- L ... Motor zusammenbauen und einbauen

Abb.:

Vorgänge werden durch Zuordnungspfeile dargestellt, die Ereignisse verbinden.

Die Länge der Zuordnungspfeile ist unabhängig von der Dauer der Vorgänge.

Jedem Vorgang entspricht genau ein Pfeil, und umgekehrt entspricht jedem Pfeil genau ein Vorgang.

Jeder Vorgang beginnt mit einem Anfangsereignis und endet mit einem Endereignis.

Die durch einen Pfeil oder mehrere aufeinanderfolgende Pfeile hergestellte Verbindung zweier Ereignisse heißt Weg.

Ein Ereignis kann erst stattfinden, wenn alle zu ihm führenden Vorgänge beendet sind.

Der früheste Zeitpunkt FZ

Das ist der Zeitpunkt, in dem Ereignis frühestens stattfinden kann.

In das vorliegende Bsp. sind die frühestmöglichen Zeitpunkte in die linken Teil der bei den Knoten befindlichen Kästchen eingetragen:

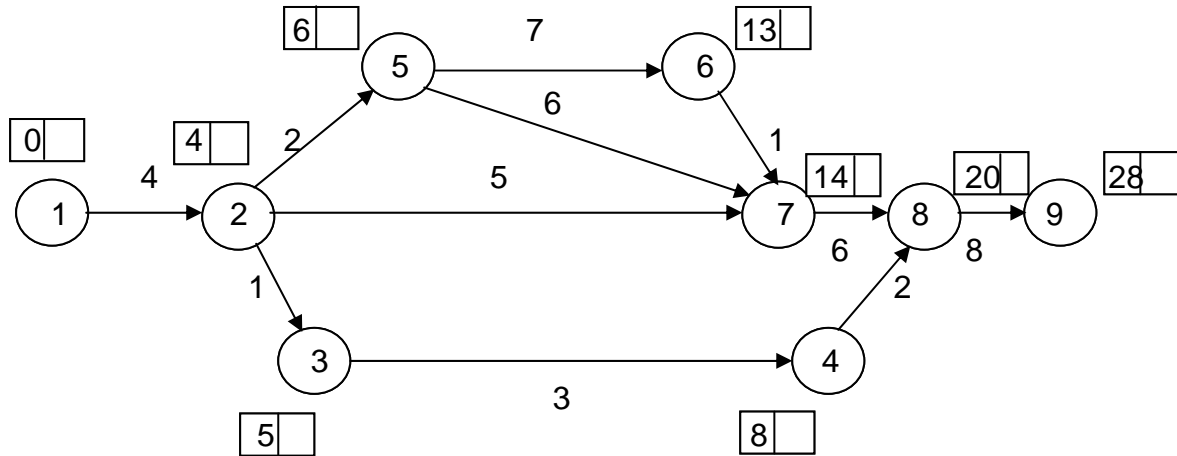


Abb.:

Der **späteste Zeitpunkt SZ**

Zu diesem Zeitpunkt darf ein Ereignis spätestens stattfinden, ohne daß das gesamte Projekt verzögert wird.

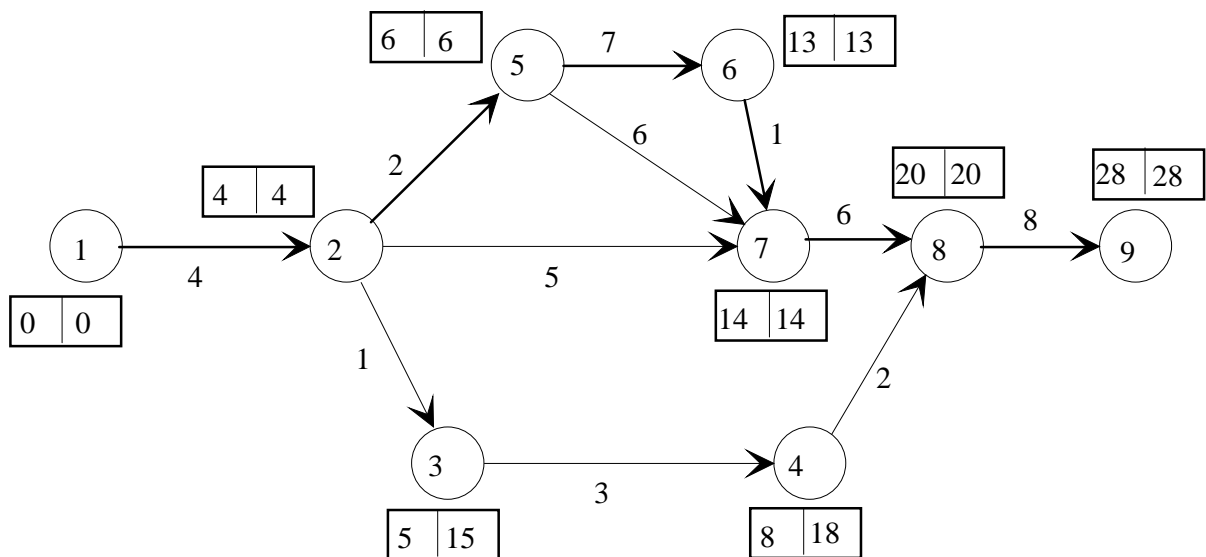


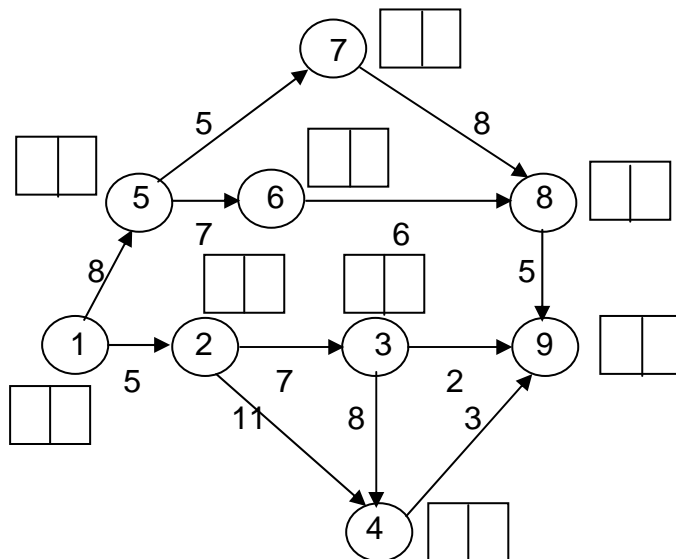
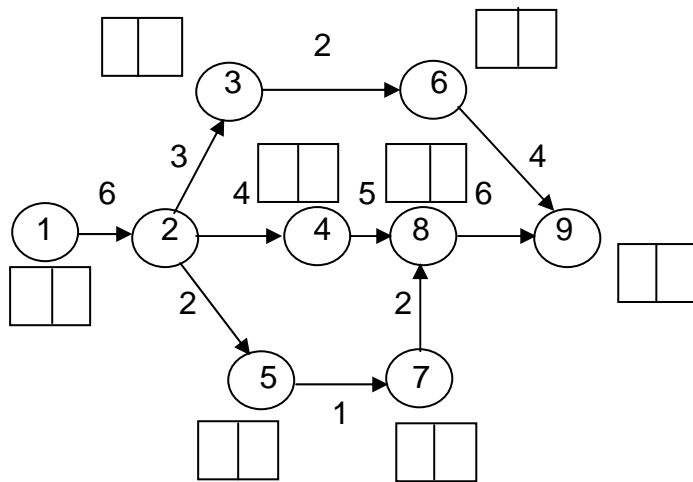
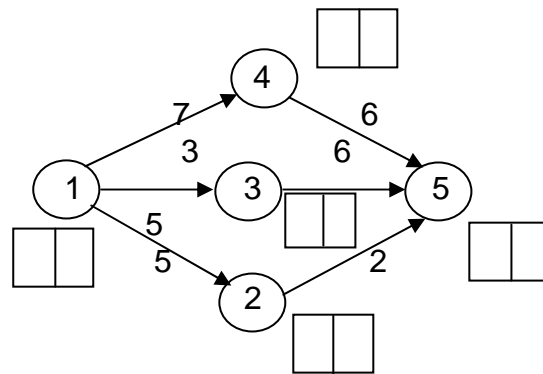
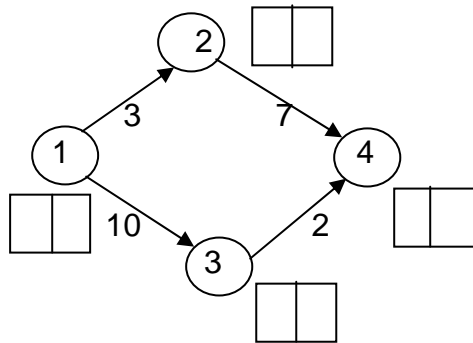
Abb.:

Definition "**Kritischer Weg**"

Vorgänge, die sich nicht verschieben lassen, ohne daß das Projektende verzögert wird, heißen kritische Vorgänge.

Die Folge aller kritischen Vorgänge heißt **kritischer Weg**.

Aufgaben: In den folgenden Netzplänen sind die kritischen Wege und die frühesten bzw. spätesten Zeitpunkte zu ermitteln.



2. Aufstellen eines CPM-Netzplans

1) Stelle den in folgenden Balkendiagramm abgebildeten Vorgang A in einem CPM-Netzplan dar!

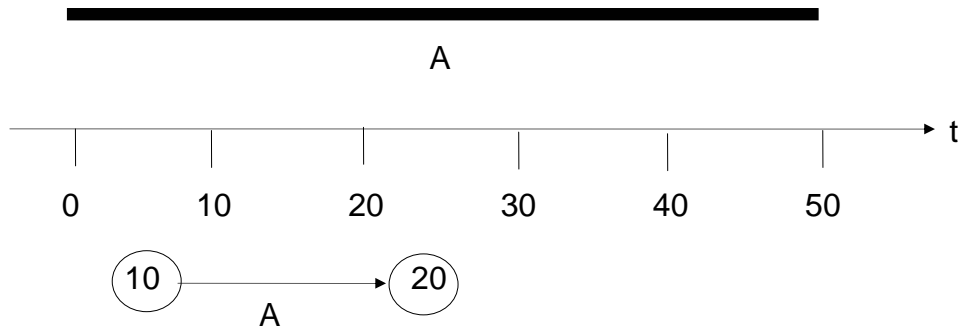


Abb.:

2) Der Vorgang B darf erst nach Beendigung des Vorgangs A begonnen werden.

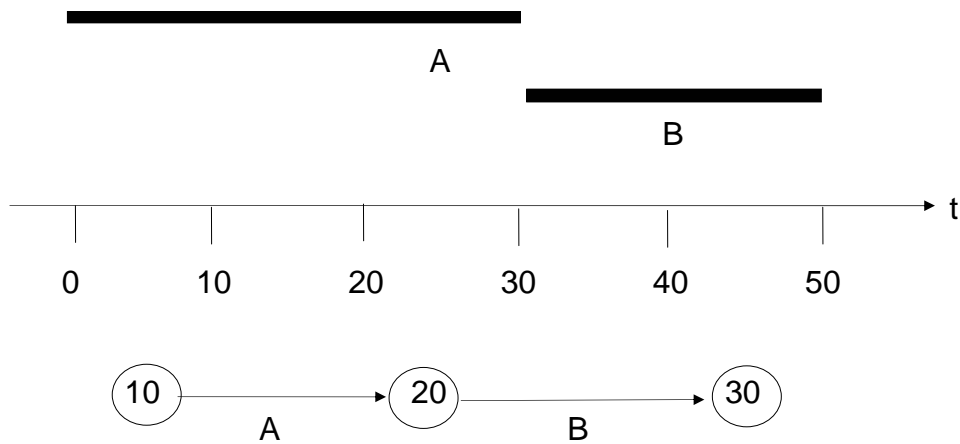


Abb.:

3) Die Vorgänge B und C dürfen erst nach Beendigung von A begonnen werden

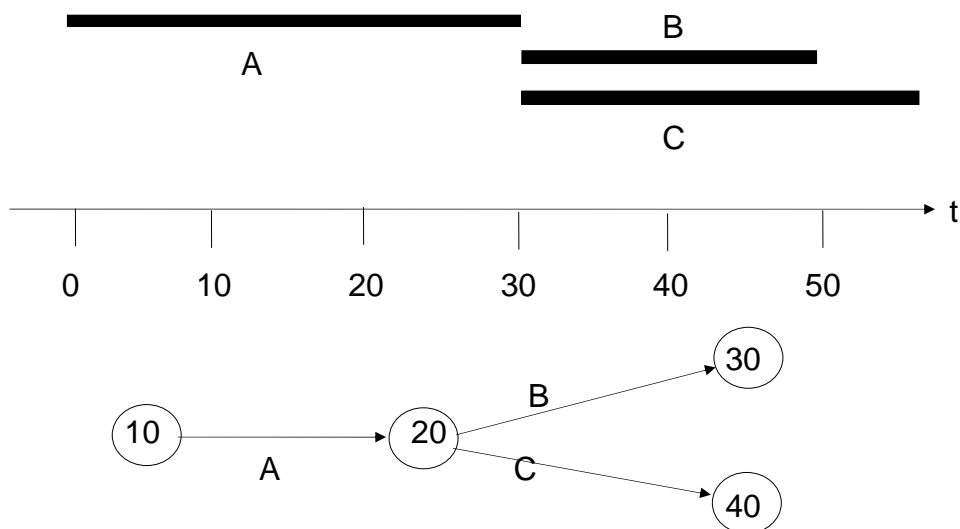


Abb.:

4) Der Vorgang B beginnt während des Vorgangs A

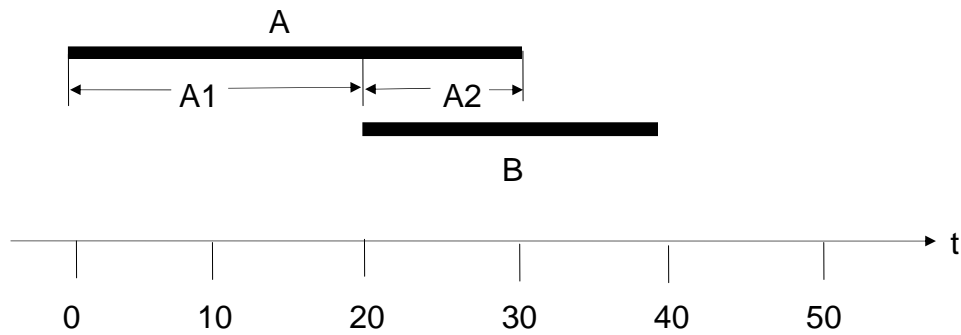
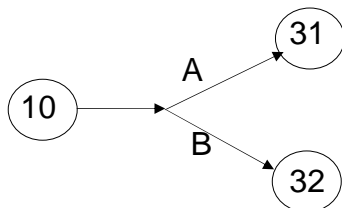
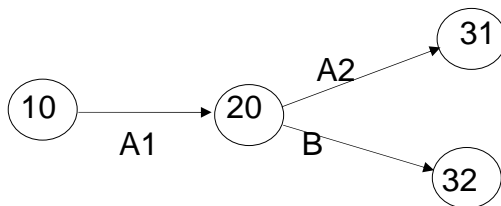


Abb.:

Ist die folgende Darstellung erlaubt?



Nein. Es wurde vereinbart, daß jeder Vorgang mit einem Ereignis beginnt und mit einem Ereignis schließt. Wie kann der im vorstehenden Balkendiagramm vorgegebene Sachverhalt dann dargestellt werden?



5) Wie würde der folgende in einem Balkendiagramm wiedergegebene Sachverhalt in einem Netzplan dargestellt werden?

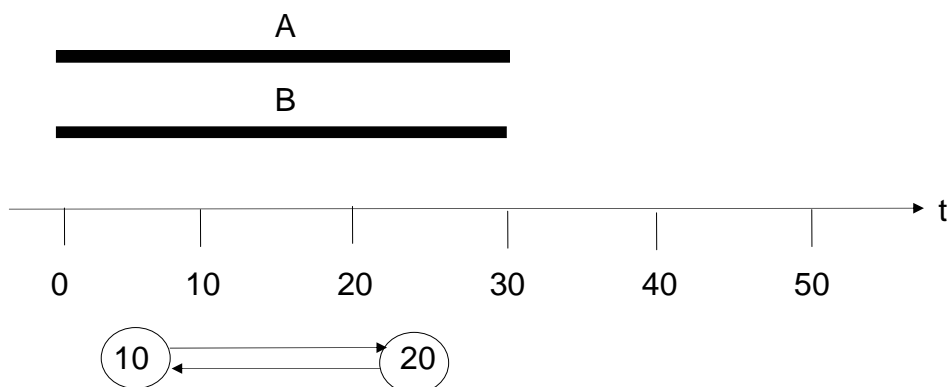


Abb.:

Eine solche Darstellung ist in Netzplänen nicht zugelassen, denn es gilt folgende Vereinbarung: „Knoten dürfen nur durch einen Vorgang verbunden werden“. Eine eindeutige Darstellung ist nur über ein neues Grundelement, den **Scheinvorgang** erreichbar.

Scheinvorgänge zeigen Anordnungsbeziehungen auf. Sie werden durch gestrichelte Pfeile dargestellt.

Das gegebene Balkendiagramm kann demnach durch folgende Netzpläne beschrieben werden:

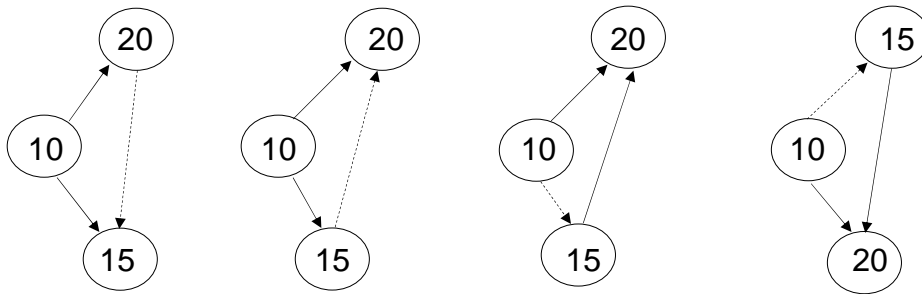


Abb.:

6) Welche Netzpläne entsprechen dem folgenden Balkendiagramm?

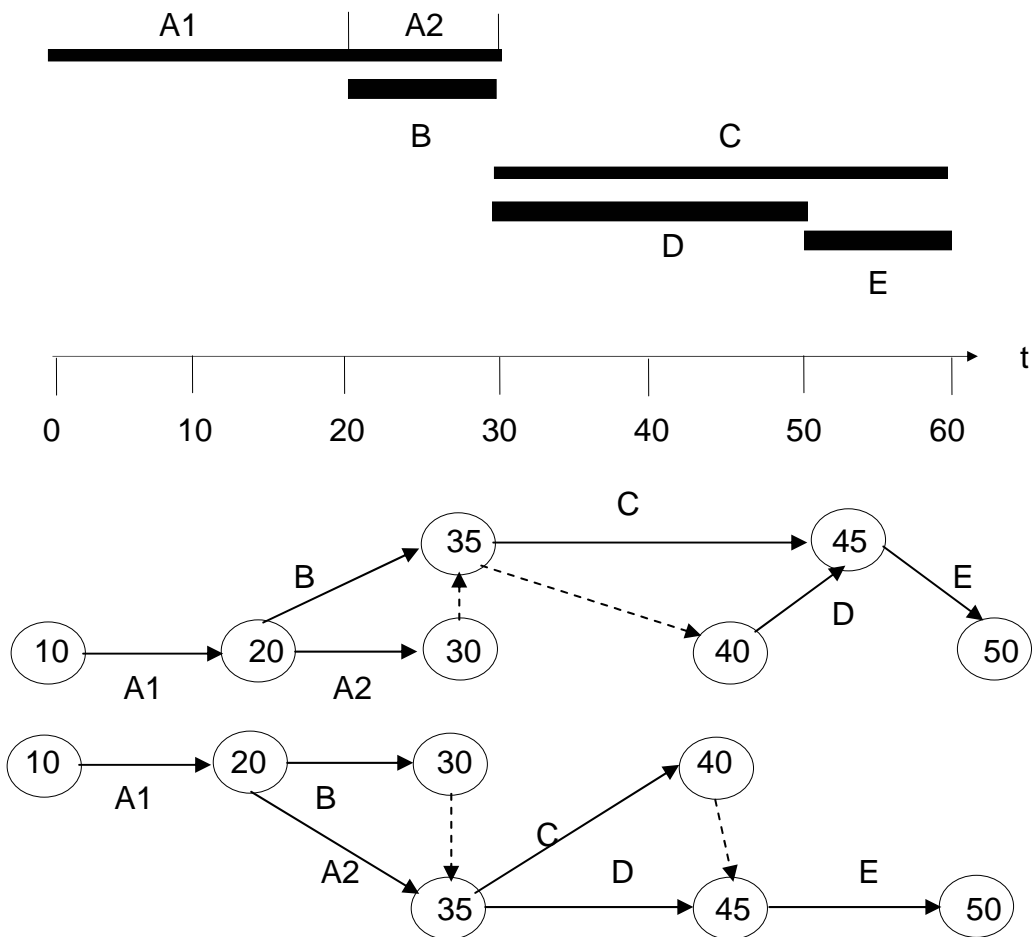


Abb.:

7) In dem folgenden Netzplan sollen die Bedingungen "Vorgang E darf erst nach Beendigung von B und Vorgang H darf erst nach Beendigung von D stattfinden" eingebaut werden?

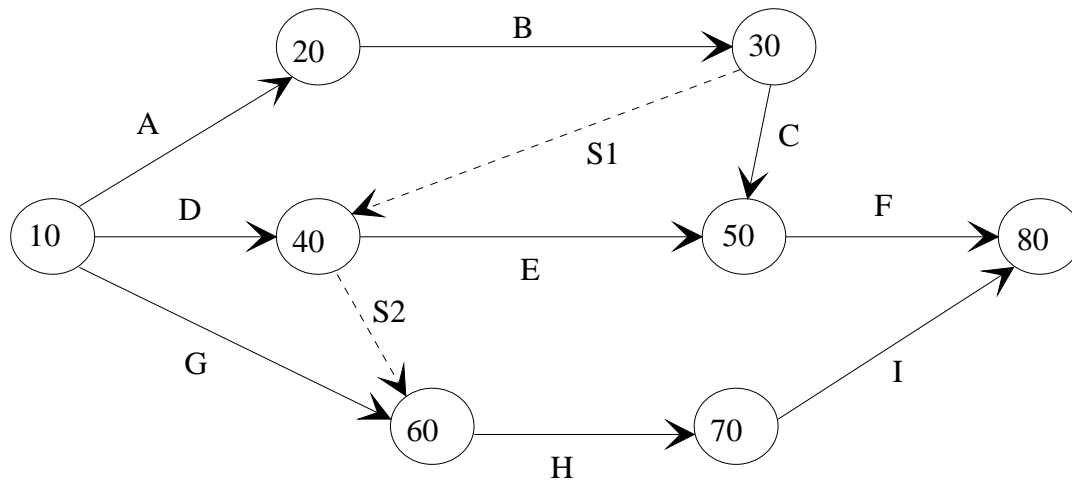


Abb.:

S1 und S2 erfüllen die geforderten Bedingungen. Dieser Netzplan ist aber keine richtige Darstellung des angegebenen Prozesses. Nach diesem Netzplan darf der Vorgang H erst nach Beendigung des Vorgangs B stattfinden. Die Ereignisse Ende S1 und Anfang S2 sind daher zu trennen. Das kann so erfolgen:

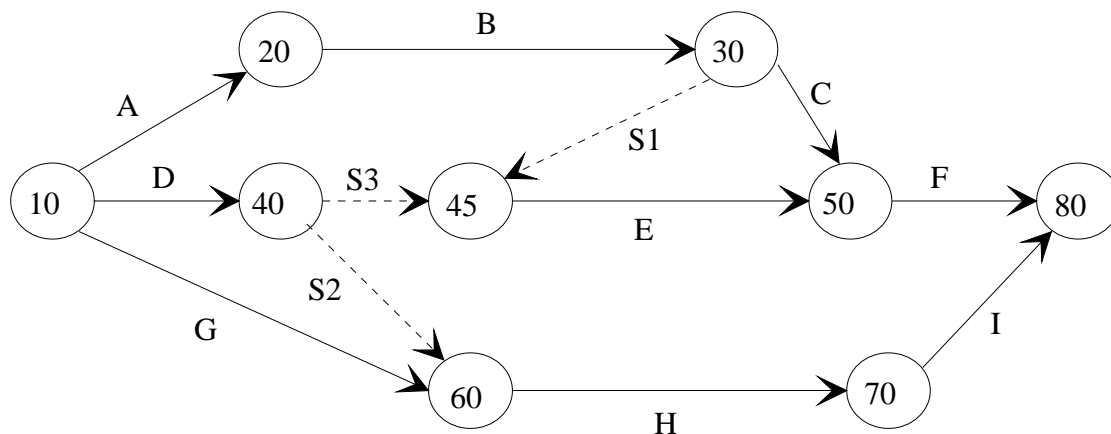


Abb.:

3. Berechnung

1) Berechnung der Vorgangszeiten

Bei CPM stehen die Vorgänge im Mittelpunkt der Betrachtung. Die Berechnung soll eine Tabelle liefern, in der alle wichtigen Zeitwerte der Vorgänge erscheinen.

Nr. Anfangs-knoten	Nr. Endknoten	Vorgangs-dauer	Frühester Anfangs-Zeitpunkt FZ_i	Frühester End-Zeitpunkt FZ_j	Spätester Anfangs-Zeitpunkt SZ_i	Spätester End-Zeitpunkt SZ_j
1	2	4	0	4	0	4
2	3	1	4	5	4	15
2	5	2	4	6	4	6
2	7	5	4	14	4	14
3	4	3	5	8	15	18
4	8	2	8	20	18	20
5	6	7	6	13	6	13
5	7	6	6	14	6	14
6	7	1	13	14	13	14
7	8	6	14	20	14	20
7	8	6	14	20	14	20
8	9	8	20	28	20	28

2) Berechnung der Pufferzeiten der Vorgänge

Gesamtpufferzeit GP

Das ist die Zeitspanne, um die man die Länge eines Vorgangs verschieben kann, ohne den Endtermin zu beeinflussen

$$GP_j = SZ_j - FZ_i - D_{ij}$$

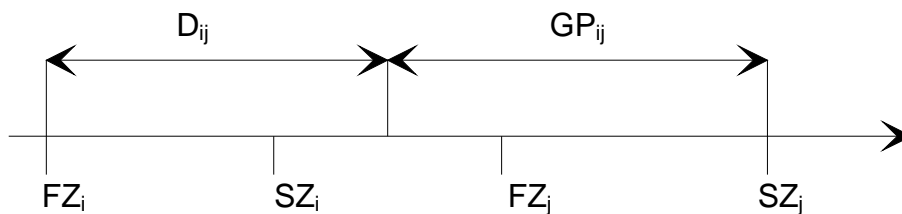


Abb.:

Die Gesamtpufferzeit GP kann aufgeteilt werden in:

- freie Pufferzeit FP
- freie Rückwärtspufferzeit FRP
- unabhängige Pufferzeit UP

Freie Pufferzeit FP

Das ist die Zeitspanne, um die ein Vorgang gegenüber seiner frühesten Lage verschoben werden kann, ohne die früheste Lage anderer Vorgänge zu beeinflussen:

$$FP_{ij} = FZ_j - FZ_i - D_{ij}$$

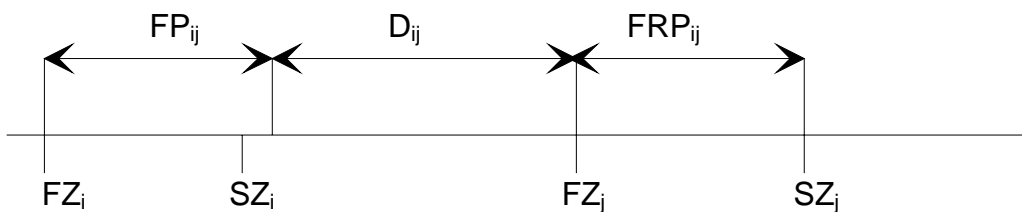


Abb.:

Freie Rückwärtspufferzeit FRP (bedingt verfügbare Pufferzeit)

Das ist die Zeitspanne, um die ein Vorgang gegenüber seiner spätesten Lage verschoben werden kann, ohne die späteste Lage anderer Vorgänge zu beeinflussen:

$$FRP_{ij} = SZ_j - FZ_j \quad (FRP_{ij} = GP_{ij} - FP_{ij})$$

Unabhängige Pufferzeit

Das ist die Zeitspanne, um die ein Vorgang verschoben werden kann, wenn sich seine Vorgänge in spätestster Lage und seine Nachfolger in frühester Lage befinden

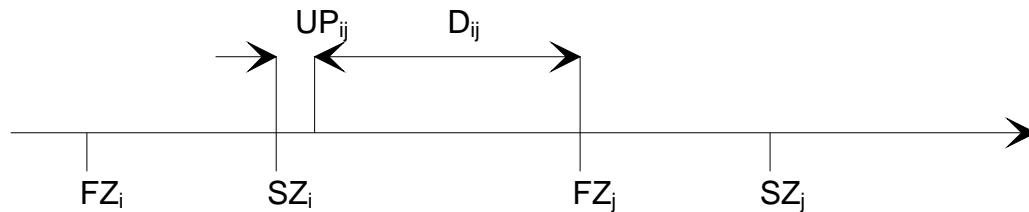


Abb.:

Für $UP_{ij} \geq 0$ gilt: $UP_{ij} = FZ_j - SZ_i - D_{ij}$

Ist $UP_{ij} < 0$, dann setzt man $UP_{ij} = 0$.

Für alle Vorgänge, deren Anfangsknoten am kritischen Weg liegen, gilt $UP = FP$.

Aufgaben:

1. Berechne die Pufferzeiten zum einführenden Beispiel

Nr. i	Nr. j	D _{ij}	FAZ	FEZ	SAZ	SEZ	GP _{ij}	FP _{ij}	FRP _{ij}	UP _{ij}
1	2	4	0	4	0	4	0	0	0	0
2	3	1	4	5	4	15	10	0	8	0
2	5	2	4	6	4	6	0	0	0	0
2	7	5	4	14	4	14	5	5	0	5
3	4	3	5	8	15	18	10	0	10	0
4	8	2	8	20	18	20	10	10	0	0
5	6	7	6	13	6	13	0	0	0	0
5	7	6	6	14	6	14	2	2	0	2
6	7	1	13	14	13	14	0	0	0	0
7	8	6	14	20	14	20	0	0	0	0
8	9	8	20	28	20	28	0	0	0	0

2. Gegeben ist der folgende CPM-Netzplan.

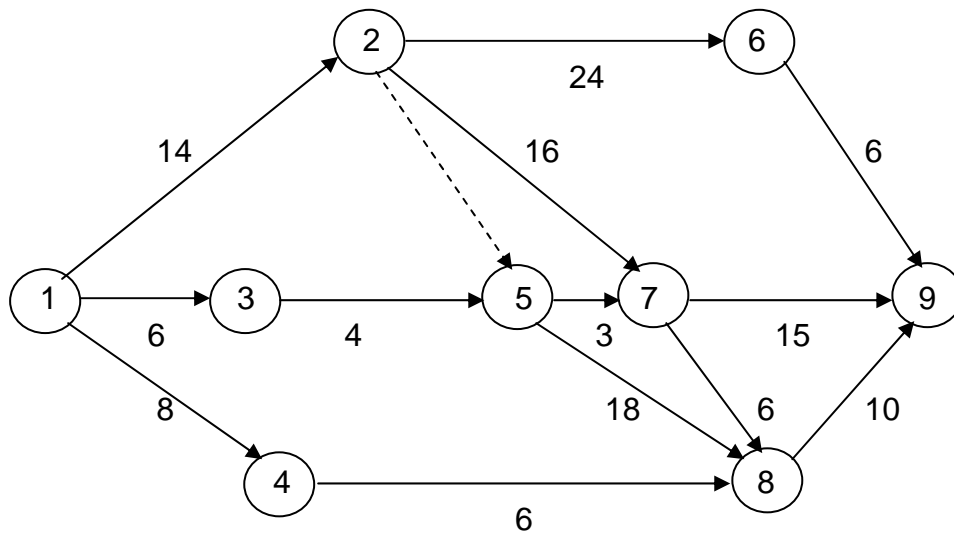


Abb.

Berechne die frühesten Zeitpunkte Fz_i bzw. FZ_j , die spätesten Zeitpunkte Sz_i und SZ_j , die Gesamtpufferzeit Gp_{ij} , die freie Pufferzeit Fp_{ij} und die unabhängige Pufferzeit Up_{ij} . Die errechnete Werte trage in die folgende Tabelle zur Zeitanalyse ein.

Nr. i	Nr. j	D_{ij}	FAZ	FEZ	SAZ	SEZ	GP_{ij}	FP_{ij}	FRP_{ij}	UP_{ij}
1	2	14								
1	3	6						0	10	0
1	4	8								
2	5	-								
2	6	24								
2	7	16								
3	5	4								
4	8	6								
5	7	3								
5	8	18								
6	9	6								
7	8	6								
7	9	14								
8	9	10								

4.2.2.3 MPM

1 Grundbegriffe

Grundelemente eines knotenorientierten Netzplans sind Vorgänge. Ein Vorgang wird folgendermaßen dargestellt:

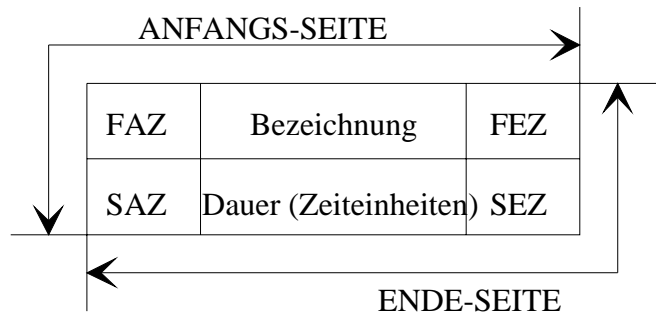


Abb.:

Die zur Fertigstellung des Vorgangs benötigte Zeit wird in das Feld mit der Bezeichnung "DAUER" eingetragen. Einem Vorgang sind 4 Zeitwerte zugeordnet:

FAZ ... frühester Anfangszeitpunkt (frühester Anfang)

SAZ ... spätester Anfangszeitpunkt (spätester Anfang)

FEZ ... frühester Endzeitpunkt (frühestes Ende)

SEZ ... spätester Endzeitpunkt (spätestes Ende)

Ein Vorgang muß im Intervall zwischen frühestem und spätestem Anfang beginnen und im Intervall zwischen frühestem und spätestem Ende fertig werden.

$$FEZ = FAZ + DAUER$$

$$SEZ = SAZ + DAUER$$

Ein Vorgang beansprucht Zeit und Hilfsmittel. Eine Anordnungsbeziehung vom Ende eines Vorgangs zum Anfang seines Nachfolgers heißt **Normalfolge (NF)**.

Bsp.: "Motorüberholung"

Vorgangnr.	Bezeichnung	Dauer (in 1/4 Std.)
1	Start	0
2	Motor ausbauen	4
3	Kurbelwelle und Kolben ausbauen	2
4	Zylinderkopf zerlegen	1
5	Kurbelwelle schleifen	7
6	Zylinder schleifen	6
7	Kolben ersetzen	5
8	Ventile schleifen	3
9	Block zusammensetzen	6
10	Kopf zusammenbauen	2
11	Motor zusammenbauen	8
12	Ende	0

Netzdarstellung:

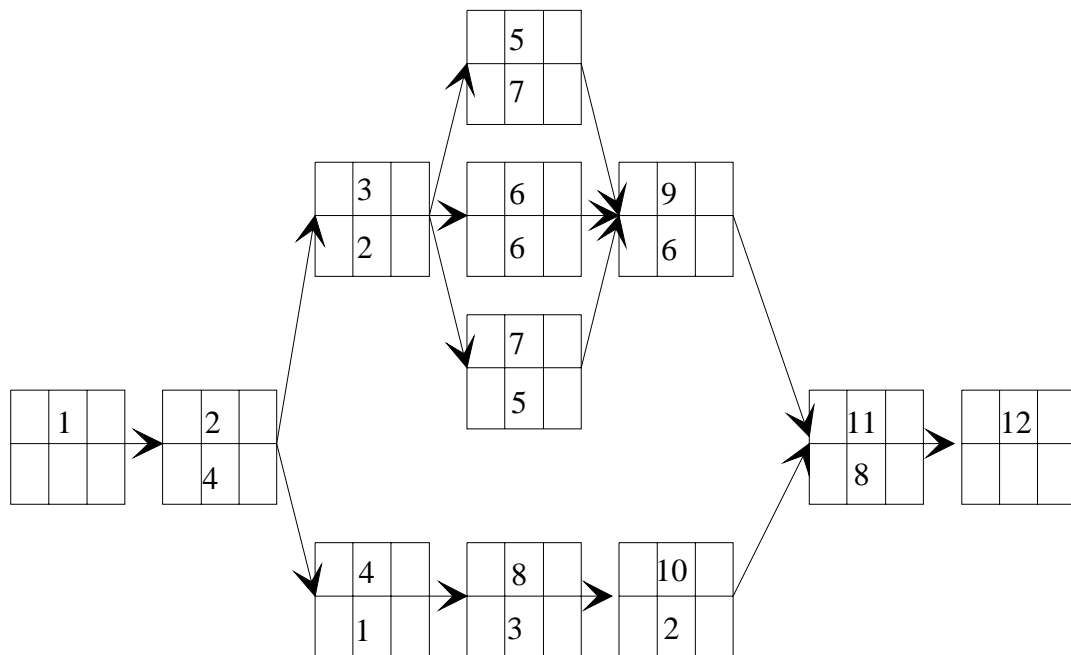


Abb.:

Mit Hilfe der vorgegebenen Netzplanstruktur und der Vorgangsliste werden berechnet:

1. FAZ
2. SAZ

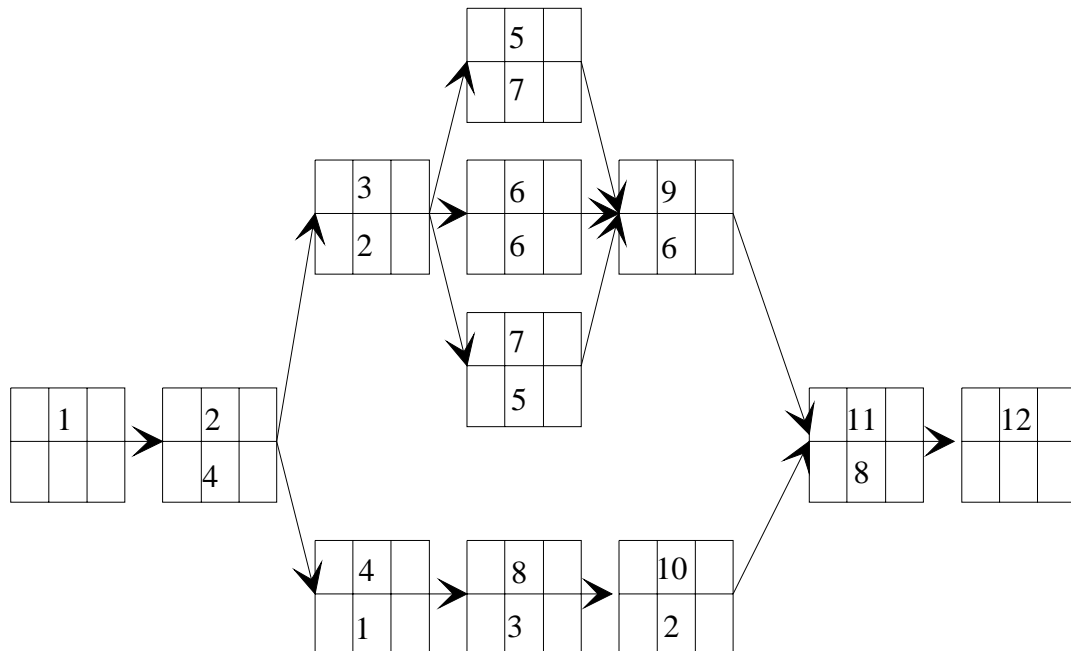


Abb.:

Zwei Arten von Vorgängen sind erkennbar

1. FAZ = SAZ und FEZ = SEZ

Vorgänge mit dieser Eigenschaft müssen genau zum angegebenen Zeitpunkt beginnen und auch enden. Ihre Ausführung ist für die Einhaltung des Endtermins wichtig (kritische Vorgänge)

2. FAZ < SAZ und FEZ < SEZ

Diese Vorgänge lassen sich im Intervall zwischen FAZ und SEZ verschieben und beeinflussen den Endtermin nicht, solange sie in diesem Intervall verbleiben. Vorgänge mit dieser Eigenschaft müssen genau zum angegebenen Zeitpunkt beginnen und auch enden. Ihre Ausführung ist für die Einhaltung des Endtermins wichtig (kritische Vorgänge). Sie werden durch den sog. kritischen Weg, der vom Start zum Ende führen muß, verbunden.

2. Anordnungsbeziehungen

1) Aufstellen eines Netzplans

Vorarbeiten

Vor dem Aufstellen eines Netzplans ist auszuführen:

1. Unterteilung des Gesamtvorhabens in einzelne Abschnitte
2. Ermittlung der Dauer (in Zeiteinheiten [ZE]) je Einzelvorgang.
3. Ermittlung der Abhängigkeiten je Einzelvorgang

Ablauf von Vorgängen

a) Einzelner Vorgang

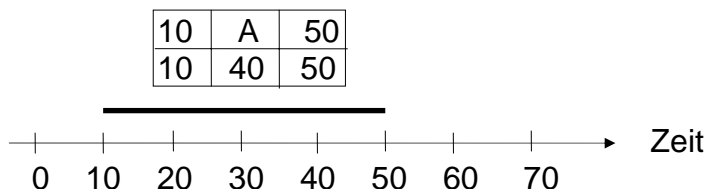


Abb.:

b) Normalfolgen

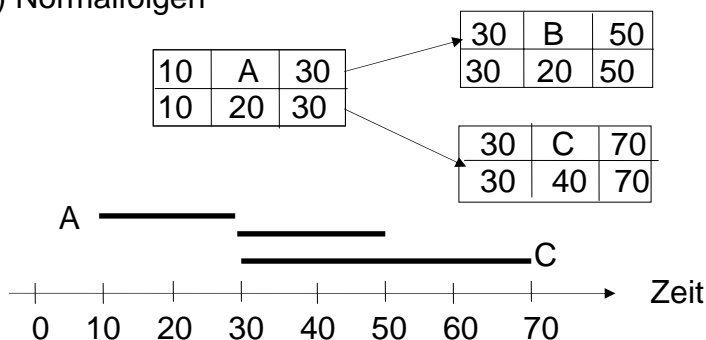


Abb.:

c) In dem folgenden Bsp. darf der Vorgang B erst 10 ZE nach Beendigung von A beginnen:

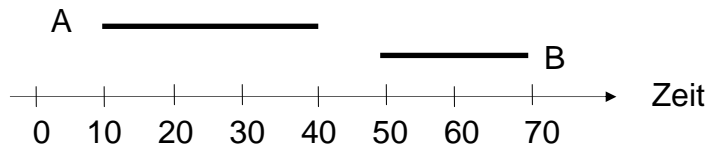


Abb.:

Zur Darstellung solcher Verbindungen gibt es **Anordnungsbeziehungen** zwischen den Vorgängen (logisch zeitliche Verbindungen)

2) Quantifizierbare Abhängigkeiten

Möglichkeiten

Man unterscheidet die folgenden quantifizierbaren Beziehungen bzgl. der Abhängigkeiten:

1. Anfang - Anfang - Beziehung (**Anfangsfolge**, AA)
2. Anfang - Ende - Beziehung (**Sprungfolge**, AE)
3. Ende - Anfang - Beziehung (**Normalfolge**, NN)
4. Ende - Ende - Beziehung (**Endfolge**, EF)

Diese 4 Folgearten lassen sich in der Vorgangsknotentechnik ohne Schwierigkeiten darstellen (im Gegensatz zur Vorgangspfeilknotentechnik).

MPM benutzt ausschließlich die **Anfangsfolge**.

AA-Beziehung

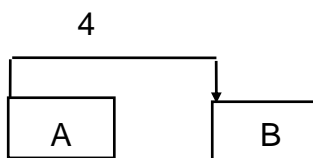


Abb.:

Vorgang B kann frühestens 4 ZE nach dem Anfang von A beginnen

EA-Beziehung

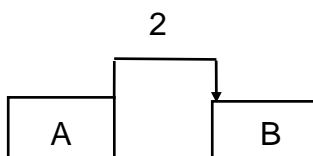


Abb.:

Vorgang B kann frühestens 2 ZE nach dem Ende von A beginnen

EE-Beziehung

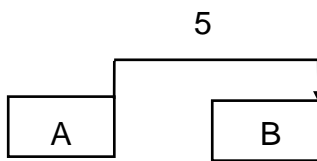


Abb.:

Vorgang B kann frühestens 5 ZE nach dem Ende von A beendet sein.

AE-Beziehung

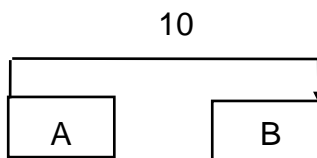


Abb.:

Vorgang B kann frühestens 10 ZE nach dem Anfang von A beendet sein.

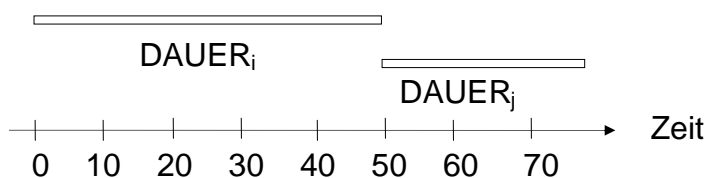
Systematik der Anordnungsbeziehungen und Zeitabstände im Vorgangspfeil- und Vorgangsknotennetzplan

Der Zeitwert a_{ij} der Anordnungsbeziehung kann größer, kleiner oder gleich der Vorgangsdauer sein. 2 Vorgänge können

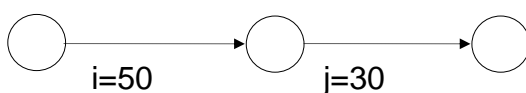
- direkt aufeinander folgen
- durch eine Wartezeit getrennt sein
- sich überschneiden
- parallel ablaufen
- im speziellen Fall gleichzeitig beginnen oder enden

1. Anordnungsbeziehungen zwischen 2 Vorgängen

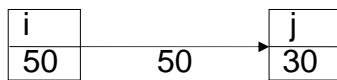
Direkte Aufeinanderfolge



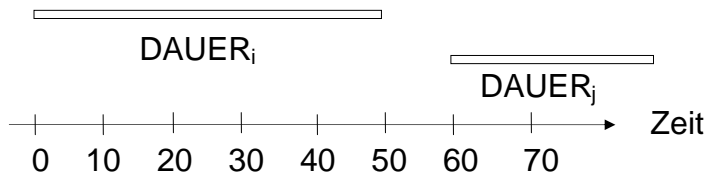
a) Darstellung im CPM-Netzplan



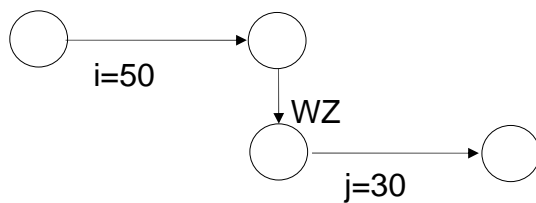
b) Darstellung in MPM-Netzplan



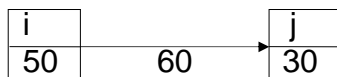
Wartezeit



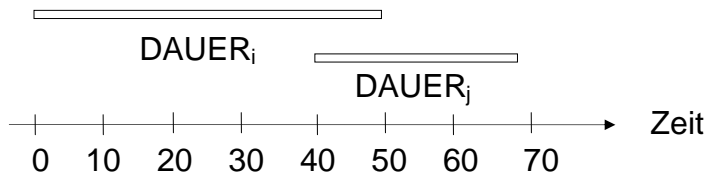
a) Darstellung im CPM-Netzplan



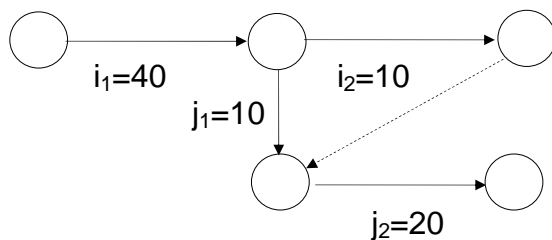
b) Darstellung in MPM-Netzplan



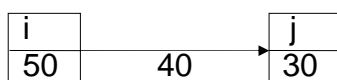
Überschneidung



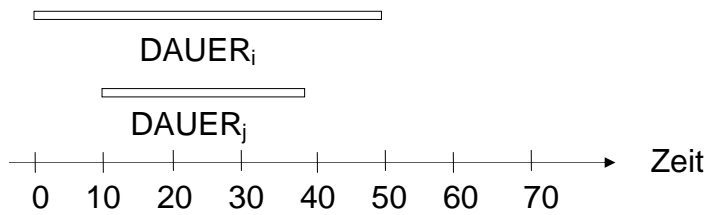
a) Darstellung im CPM-Netzplan



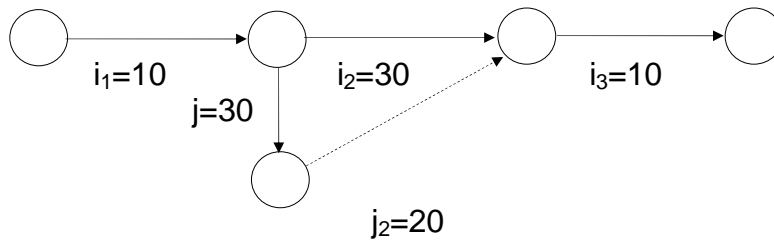
b) Darstellung in MPM-Netzplan



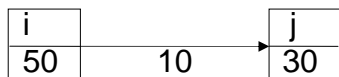
Parallelität



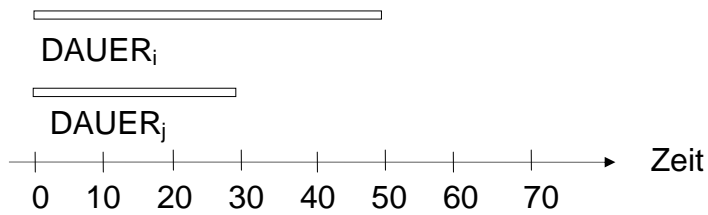
a) Darstellung im CPM-Netzplan



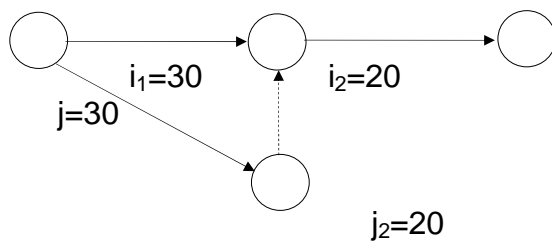
b) Darstellung in MPM-Netzplan



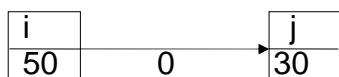
Gleichzeitiger Beginn



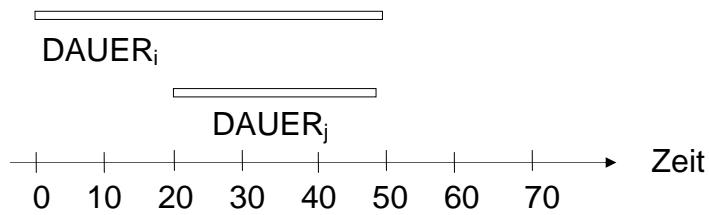
a) Darstellung im CPM-Netzplan



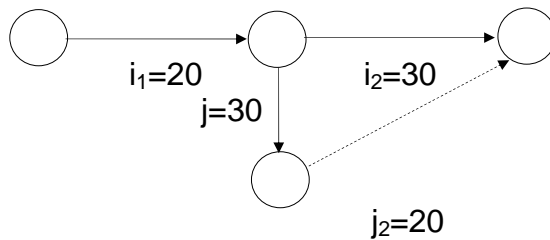
b) Darstellung in MPM-Netzplan



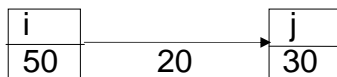
Gleichzeitiger Abschluß



a) Darstellung im CPM-Netzplan



b) Darstellung in MPM-Netzplan



2. Die zeitlichen Restriktionen in den Anordnungsbeziehungen

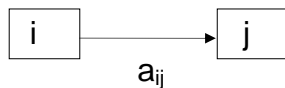
Für Vorgangsknotennetzpläne lassen sich 4 Typen zeitlicher Restriktionen in den Anordnungsbeziehungen unterscheiden:

a) Der Vorgang j kann frühestens $\min\{a_{ij}\}$ [ZE] nach Beginn des Vorgangs i anfangen.

$$t_j \geq t_i + a_{ij}$$

$$a_{ij} \geq 0$$

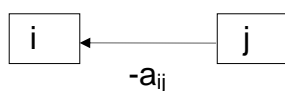
$$t_j - t_i \geq a_{ij}$$



b) Der Vorgang j soll spätestens $\max\{a_{ij}\}$ [ZE] nach Beginn des Vorgangs i anfangen

$$t_j \leq t_i + a_{ij}$$

$$t_i - t_j \geq -a_{ij}$$

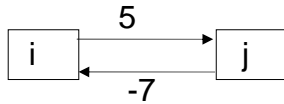


c) Der Vorgang j muß genau a_{ij} [ZE] nach Beginn des Vorgangs i anfangen.

$$-a_{ij} + a_{ij} = 0$$

d) Der Vorgang muß innerhalb $\min\{a_{ij}\}$ und $\max\{a_{ij}\}$ [ZE] nach Beginn des Vorgangs i anfangen

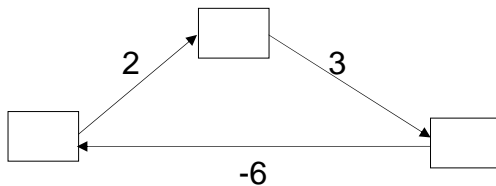
Bsp.: Die Tätigkeit j muß zwischen den Zeitpunkten 5 und 7 beginnen



Bei Vorgangsknotennetzwerken können demnach Zyklen auftreten, falls die Summe der Minimalabstände nicht positiv sind.

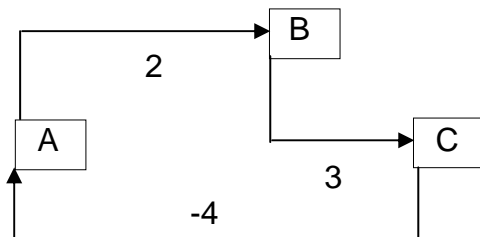
Bsp.: Interpretiere die folgenden Abbildungen

1)



B kann frühestens 2 ZE nach dem Beginn von A, C frühestens 5 Zeitpunkte nach dem Beginn von A. aber höchstens 6 ZE vor dem beginn von A anfangen.

2)



Aus dem Netzplan folgt, daß der Anfang von C mindestens 5 ZE nach dem Beginn von A, aber höchstens 4 ZE nach dem Beginn von A liegt. Dies ist ein Widerspruch.

3. Zeitrechnung

1) Vorwärtsrechnung

(Berechnung der frühest möglichen Anfangs- und Endzeitpunkte)

Vom Ausgangspunkt $FAZ = 0$ werden die frühest möglichen Beginnzeitpunkte (FAZ) ohne Berücksichtigung der negativen Anordnungsbeziehungen ermittelt. Es gelten folgende 3 Regeln:

1. Ein Vorgang i kann nur einen Termin FAZ_i zugeteilt erhalten, wenn alle Vorgänger $V(i)$ bereits einen Termin haben

2. Hat ein Vorgang i nur einen Vorgänger h, so ergibt sich der Termin FAZ_i aus:

$$FAZ_i = FAZ_h + a_{hi}$$

3. Hat ein Vorgang i mehrere Vorgänger, so wird das Maximum aus den jeweiligen Summen der Endtermine FAZ_h und den Zeitwerten $a_{(i)}$ ausgewählt:

$$FAZ_i = \max_{h \in V(i)} \{FAZ_h + a_{hi}\}$$

Sind alle FAZ ermittelt, ist zu prüfen, ob diese Zeitpunkte wegen der negativen Anordnungsbeziehungen geändert werden müssen.

Der Termin FAZ_n des Vorgangs n muß korrigiert werden, wenn

$$FAZ_m + \{-a_{nm}\} > FAZ_n$$

ist.

Die frühest möglichen Endtermine FEZ ergeben sich durch Addition der Vorgangsdauer, z.B. für den Vorgang i aus

$$FEZ_i = FAZ_i + DAUER_i$$

2) Rückwärtsrechnung

Man geht vom Projektende aus. Es gelten 3 Regeln:

1. Es müssen alle Nachfolger $N(i)$ des Vorgangs i einen Termin haben.

2. Bei nur einem Nachfolger j ergibt sich der Termin SAZ_j aus:

$$SAZ_i = SAZ_j - a_{ij}$$

3. Bei mehreren Nachfolgern gilt:

$$SAZ_i = \min_{j \in N(i)} \{SAZ_j - a_{ij}\}$$

Eine Korrektur des spätest zulässigen Beginnzeitpunktes (SAZ_m) wegen der negativen Anordnungsbeziehungen ist erforderlich, wenn

$$SAZ_m + \{-a_{nm}\} > SAZ_n$$

ist.

Der spätest zulässige Endtermin SEZ_i ergibt sich aus

$$SEZ_i = SAZ_i + DAUER_i$$

3) Bestimmung der Gesamtpufferzeit und des kritischen Weges

Gesamtpufferzeit: $GP_i = SAZ_i - FAZ_i$

Die gesamte Pufferzeit des Vorgangs i ist die Zeitspanne, um die der Termin FAZ_i maximal verschoben werden kann, ohne daß der Abschlußtermin des Projekts sich verzögert.

kritischer Vorgang: GP = 0

4) Bsp.: Gegeben ist der folgende MPM-Netzplan. Berechne die frühest möglichen und spätest zulässigen Anfangs- und Endzeitpunkte:

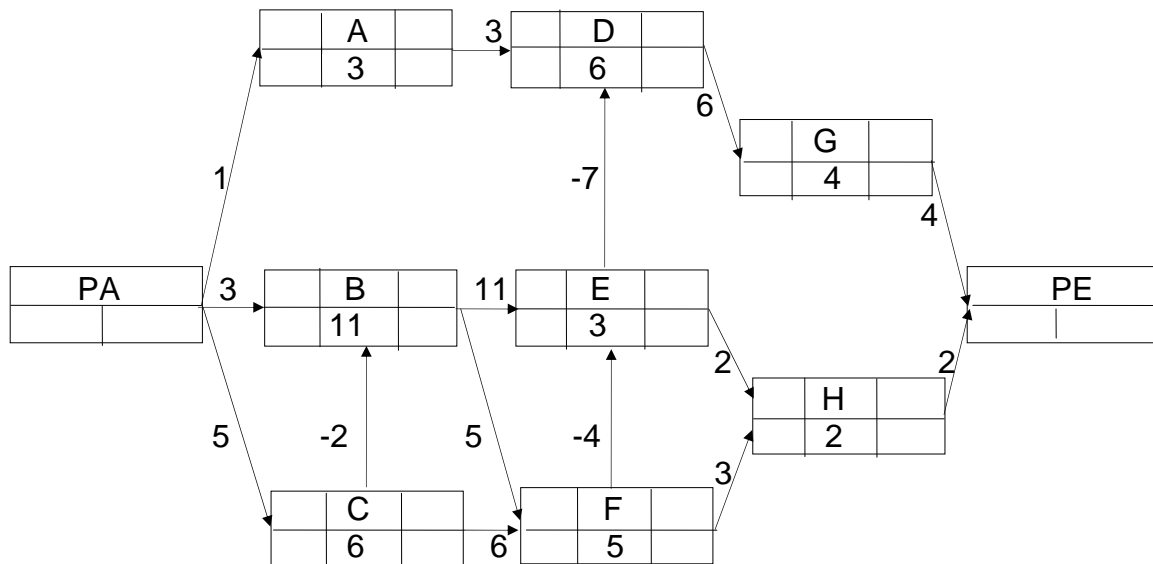


Abb.:

Berechnung der frühest möglichen Anfangs- und Endzeitpunkte:

1. Schritt: Vorwärtsrechnung, bei der nur die Anordnungsbeziehungen mit positiven Zeitwerten betrachtet werden.

(Berechnung der frühest möglichen Anfangszeitpunkte)

2. Schritt: Überprüfen, ob Anordnungsbeziehungen mit negativen Zeitwerten Einfluß auf die frühest möglichen Anfangszeitpunkte haben.

$$(C - B) = -2$$

C muß spätestens 2 ZE nach Beginn von B beginnen. Diese Bedingung ist erfüllt.

$$(F - E) = -4$$

F muß spätestens 4 ZE nach dem Anfang von E beginnen. Diese Bedingung ist erfüllt.

$$(E - D) = -7$$

F muß spätestens 7 ZE nach Beginn von Vorgang D beginnen. Das ist nach den bisherigen Berechnungen erfüllt. Die früheste Zeit des Vorgangs E ist 14. Um den Maximalabstand von 7 ZE zwischen den Beginnzeitpunkten von D und E einzuhalten, muß die früheste Zeit von D auf 7 ZE erhöht werden. Das hat Konsequenzen für die früheste Zeit von G. Sie muß auf 13 erhöht werden.

Berechnung der spätestzulässigen Anfangs- und Endzeitpunkte

1. Schritt: Rückwärtsrechnung

2. Schritt: Überprüfen der negativen Anordnungsbeziehungen

$$(C-B) = -2$$

C muß spätestens 2 ZE nach B beginnen. Der zuerst erreichte Wert von / ZE muß auf 5 ZE herabgesetzt werden

$$(E-D) = -7$$

E muß spätestens 7 ZE nach D beginnen. Die Bedingung ist erfüllt.

Ergebnis der Berechnungen

Tätigkeit	Dauer	FAZ	FEZ	SAZ	SEZ	Gesamt-Puffer	krit. Pfad
A	3	1	4	5	8	4	
B	11	2	14	3	14	0	x
C	6	5	11	5	11	0	x
D	6	7	13	8	14	1	x
E	3	14	17	14	17	0	x
F	5	11	16	13	18	2	
G	4	13	17	14	18	1	
H	2	16	18	16	18	0	x

4.2.3 Flüsse in Transportnetzen

4.2.3.1 Das Problem des maximalen Flusses (Algorithmus von Ford / Fulkerson)

Ein Transportnetz ist ein schleifenloser Graph, in dem jeder Strecke eine als Streckenkapazität bezeichnete Zahl $k(e) > 0$ zugeordnet wird. Im Transportnetz existiert ein Netzeintritt x_A und ein Netzaustritt x_E . Jeder Strecke e ist eine als Fluß bezeichnete Größe $PHI(e)$ zugeordnet:

$$(1) 0 \leq PHI(e) \leq k(e)$$

$$(2) \sum_{e \in E_x^-} PHI(e) - \sum_{e \in E_x^+} PHI(e) = 0 \quad (, \text{ falls } x \neq x_A, x \neq x_E)$$

E_x^-, E_x^+ : Menge der in x einmündenden bzw. von x ausgehenden Strecken. Damit beschreibt (2) die sog. konservative Eigenschaft des Flusses (der Fluß zum Knoten ist gleich dem Fluß vom Knoten).

$$PHI(z) = \sum_{e \in E_z^-} PHI(e) \text{ ist der Flußwert im Knoten } x_z.$$

Die Bestimmung des maximalen Flusses durch ein Netz ist gleichwertig der Bestimmung des maximalen Flusses im Netzaustritt. Er kann folgendermaßen bestimmt werden:

1. Schritt: "Bestimmen eines zulässigen Ausgangsflusses"

Zuordnung eines Flusses $PHI(e)$ zu jeder Kante.

$PHI(e)$ genügt den Bedingungen (1), (2). Danach ist das Ausgangsnetz bestimmt.

2. Schritt: "Bestimmen eines vollständigen Flusses"

Der Fluß ist vollständig, wenn jeder Weg zwischen x_A und x_E mindestens eine gesättigte Strecke enthält, d.h. eine Strecke für die $PHI(e) = k(e)$ ist. Zur Bestimmung eines vollständigen Flusses genügt es, den auf nicht gesättigten Strecken beschränkten Teil des Graphen zu betrachten.

Man vergrößert dann auf diesen "ungesättigten" Wegen den Fluß um eine Einheit und setzt das solange fort bis der Weg gesättigt ist, d.h. mindestens eine gesättigte Strecke enthält.

3. Schritt

a) Markieren des Eintritts durch +

b) Falls eine Knoten x_i markiert ist, dann ist zu markieren:

- Durch ein $+x_j$ alle nicht markierten Ecken x_j , für die eine ungesättigte Strecke (x_i, x_j) vorhanden ist.
- Durch ein $-x_j$ alle jene nicht markierten Knoten x_j , für die eine Strecke (x_i, x_j) existiert, deren Fluß nicht Null ist.

Gelingt es damit den Austritt zu markieren, so ist der Fluß nicht maximal. Es besteht dann zwischen x_A und x_E eine Kette, deren sämtliche Ecken voneinander verschieden sind. Jeder Knoten ist (bis auf das Vorzeichen) mit dem Index des vorhergehenden Knoten markiert:

Falls $e \notin \mathcal{E}$, wird $\text{PHI}'(e) = \text{PHI}(e)$ gesetzt.

Falls $e \in \mathcal{E}$ und von x_A nach x_E gerichtet ist, wird $\text{PHI}'(e) = \text{PHI}(e) + 1$ gesetzt.

Falls $e \in \mathcal{E}$ und von x_E nach x_A gerichtet ist, wird $\text{PHI}'(e) = \text{PHI}(e) - 1$ gesetzt.

$\text{PHI}'(e)$ ist ein Fluß. Da $\text{PHI}'(e) = \text{PHI}(e) + 1$ ist, folgt ein verbesserter Flußwert.

4. Schritt

Falls sich keine Verbesserung mehr erzielen läßt, d.h. der Knoten x_E nicht mehr markiert werden kann, dann hat der Flußwert seinen maximalen Wert erreicht.

Bsp.:

Bestimme mit Hilfe des Algorithmus von Ford-Fulkerson den maximalen Fluß durch das folgende Netz:

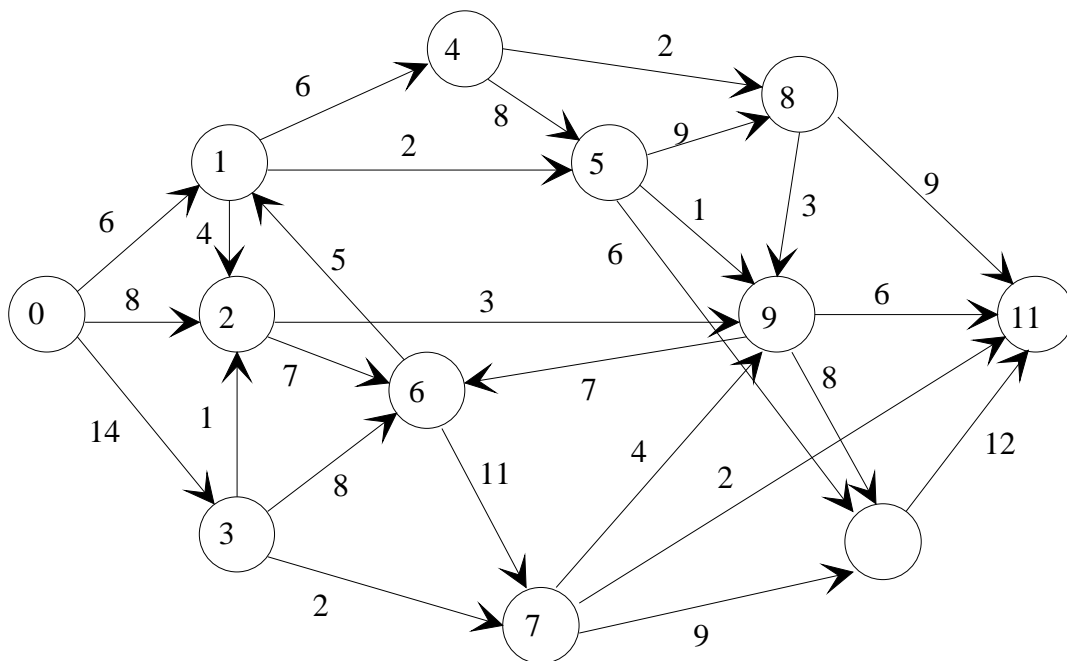


Abb.:

Lösungsschritte:

1. Bestimmen eines zulässigen Ausgangsflusses

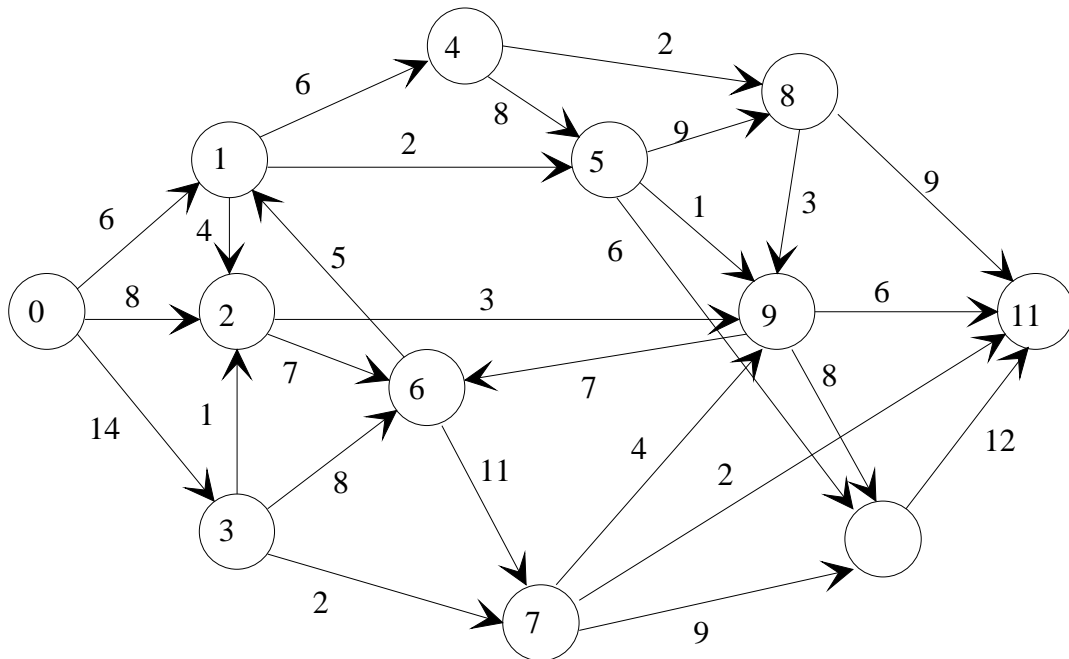


Abb.:

2. Bestimmen eines vollständigen Flusses

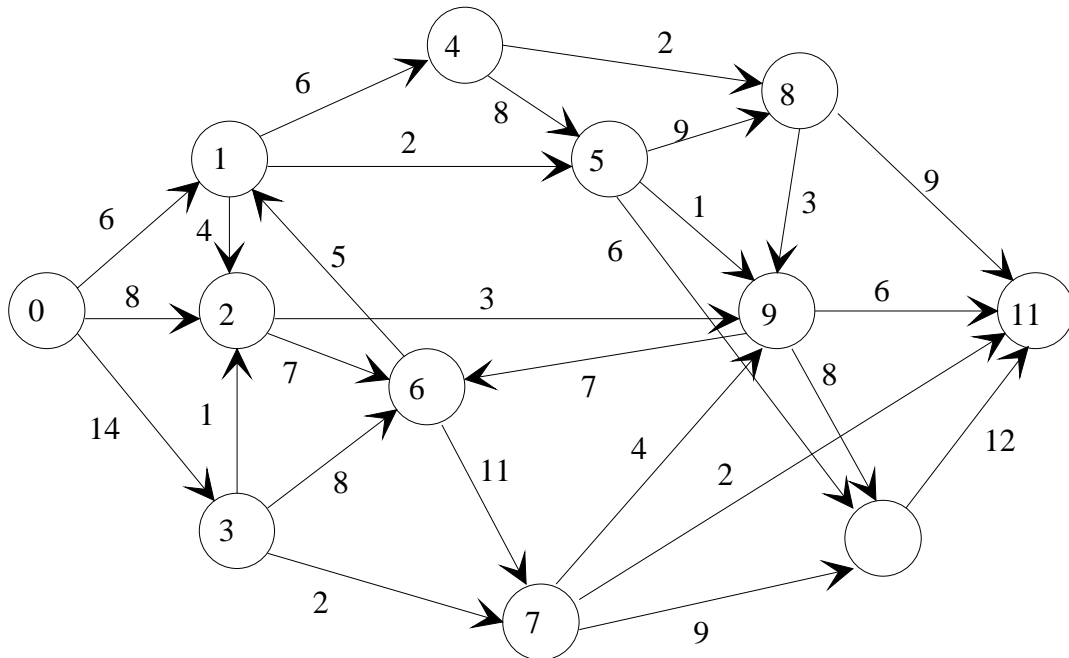


Abb.:

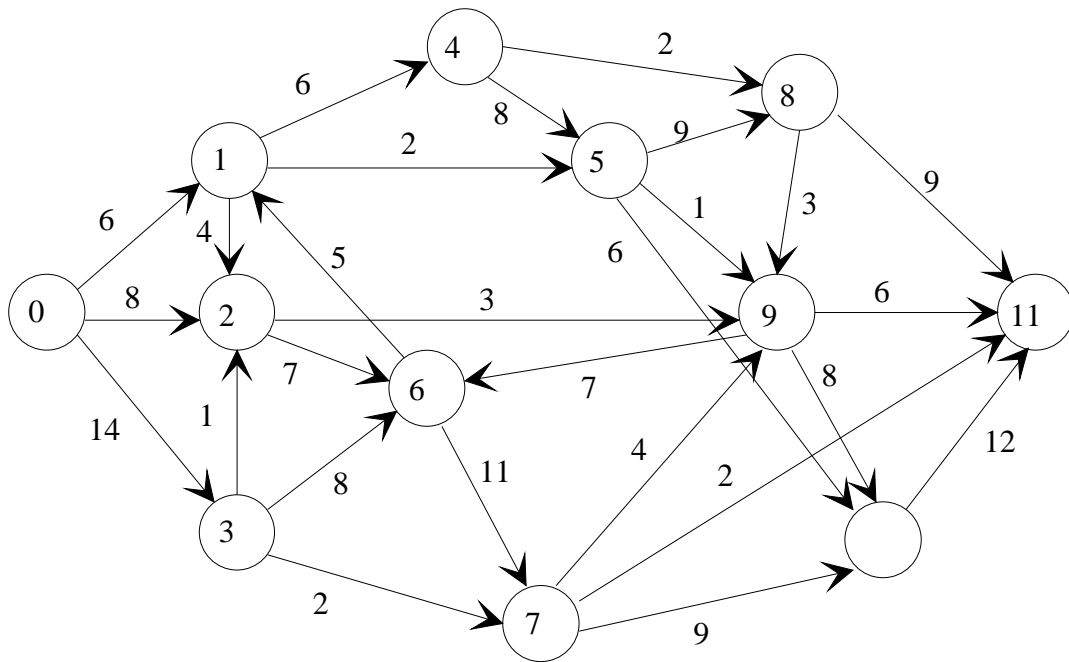


Abb.:

3. Markieren der Knoten

a) Markiere den Eintritt durch + (Knoten 0)

b) Falls ein Knoten x_i markiert ist, markiere durch $+x_i$ alle nicht markierten Ecken x_j für die eine ungesättigte Strecke (x_i, x_j) vorhanden ist, markiere durch $-x_j$ alle nicht markierten Knoten x_j für die eine Strecke (x_i, x_j) existiert, deren Fluß nicht Null ist. Gelingt es, mit diesem Verfahren den Austritt zu markieren, ist der Fluß nicht maximal.

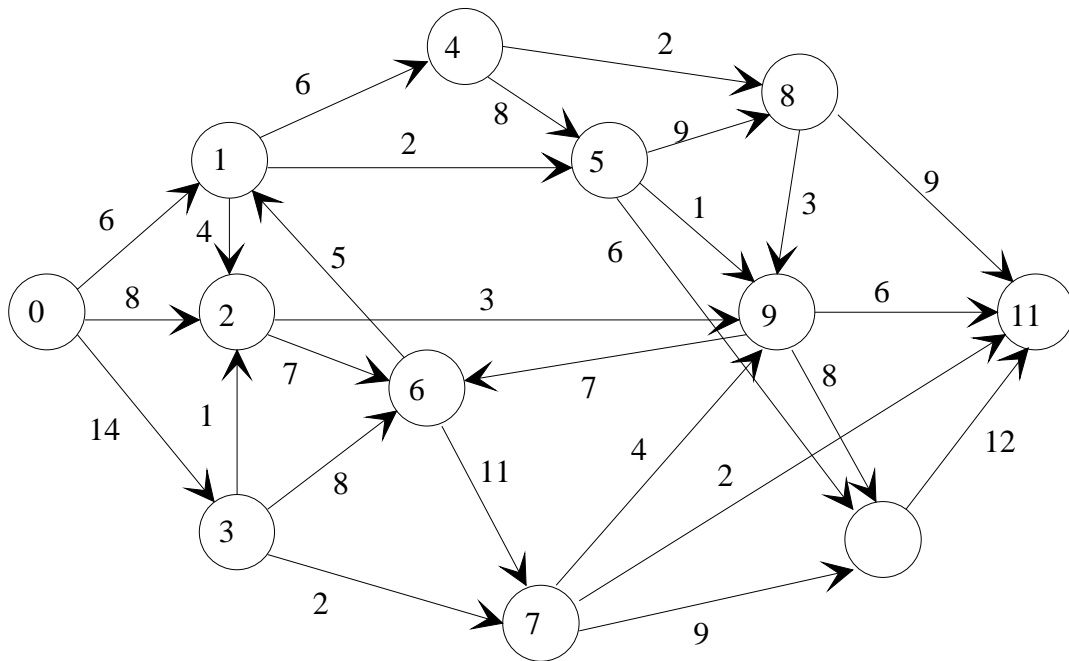
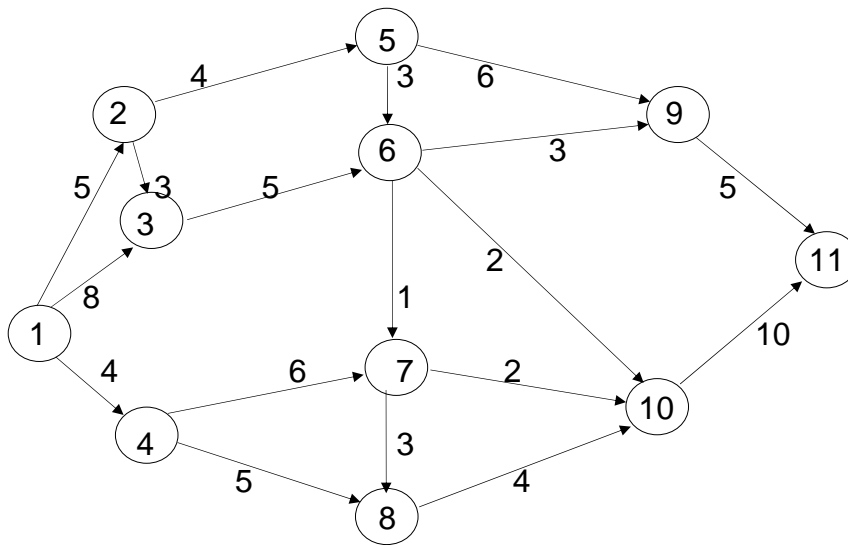


Abb.:

Ein maximaler Gesamtfluß wurde gefunden. Um seinen Wert zu bestimmen, ist eine Kurve derart zu zeichnen, daß die markierten Knoten von den nichtmarkierten getrennt werden. Dieser Schnitt durchschneidet entweder gesättigte Strecken oder eine ungesättigte Strecke, die von einem nicht markierten zu einem markierten Knoten gerichtet ist. Hieraus folgt unmittelbar auch die Unmöglichkeit einer weiteren Flußvergrößerung.

Der Schnitt ergibt einen maximalen Flußwert: $6 + 2 + 3 + 0 + 11 + 2 = 24$. Jeder andere Schnitt liefert den gleichen Wert!.

Aufgabe: Bestimme mit Hilfe des Algorithmus von Ford-Fulkerson den maximalen Fluß durch folgenden Grafen:



Anwendung des Ford/Fulkerson-Algorithmus auf Transportprobleme

Gegeben ist das folgende Transportproblem

von \ nach	5	6	7	8	Vorräte (t)
1	70	30	20	0	120
2	50	40	10	0	100
3	0	20	40	80	100
4	0	20	40	80	100
Bedarf (t)	100	80	90	150	420

Die gegebenen Vorräte sind nach dem vorliegenden Bedarf über die Transportkapazitäten (in t) zu verteilen. Vorrangig ist zu erfüllen:

Bedarf 80 t von (6)

Bedarf 150 t von (8)

(Es wird sich zeigen, daß diese Prioritäten keinen Einfluß auf den Optimierungsprozeß besitzen).

Ohne Berücksichtigung der Transportkosten soll der Bedarf weitgehend gedeckt werden. Stelle dafür einen optimalen Plan auf.

Lösung

Ausgangsdarstellung

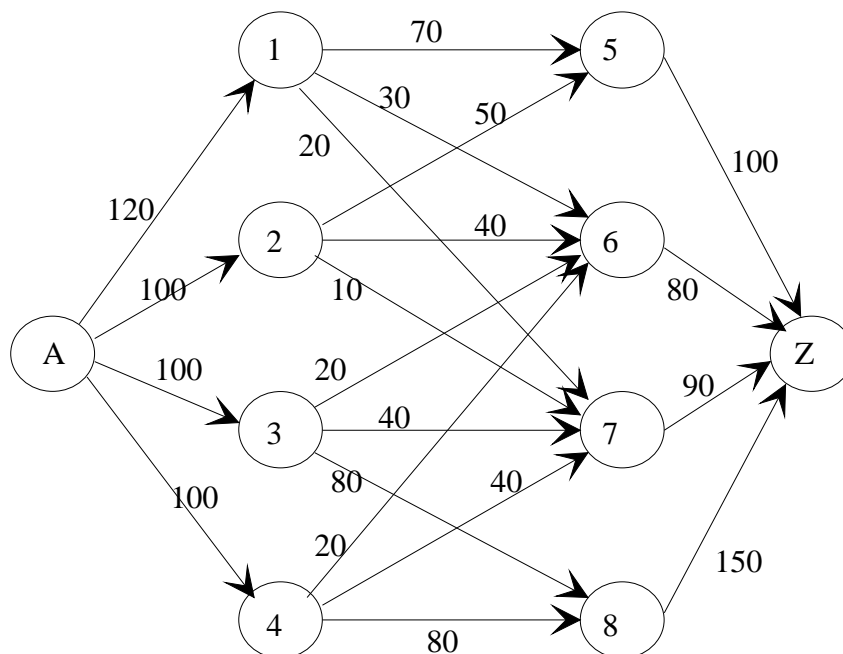


Abb.:

Zur Erfüllung vorrangiger Bestellungen erfolgen Sättigungen zunächst für

- die Kante (8,Z)
Dazu wird dem Weg (A,4,8,Z) der Fluß 80 zugewiesen und (A,3,(Z) erhält den Fluß 70, der die Kante (8,Z) sättigt.
- die Kante (6,Z)

Damit ergibt sich schließlich folgende Lösung:

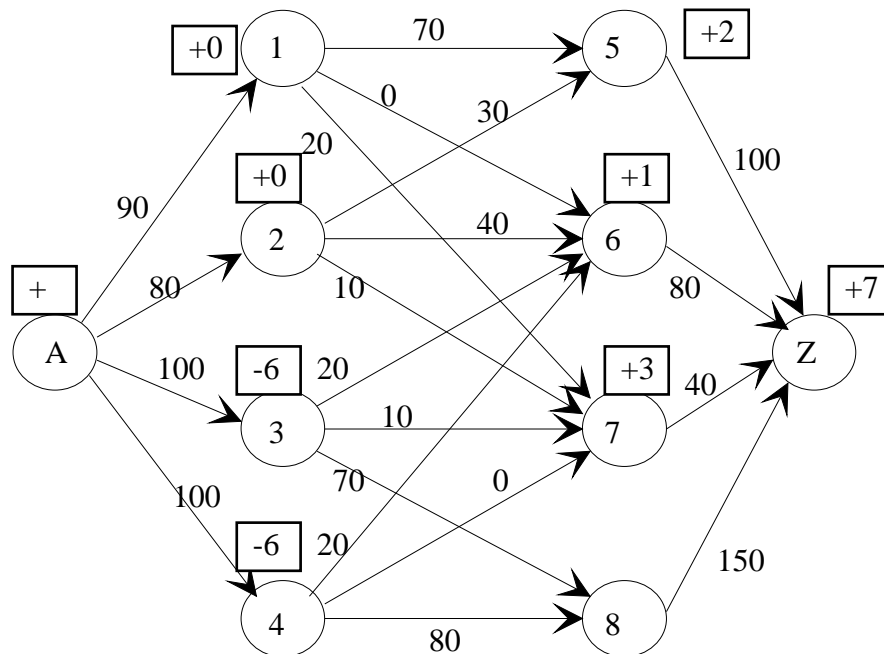
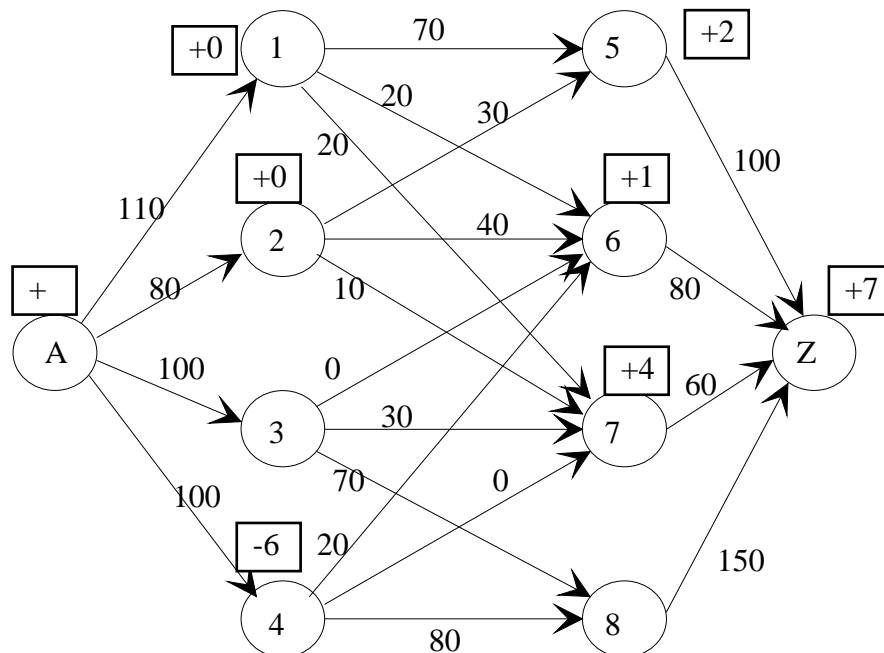


Abb.:

Man kann leicht nachprüfen, daß es keinen weiteren Weg gibt, der nicht zumindestens eine gesättigte Kante enthält. Ein "Erhöhen des Gesamtflusses" ist nur über den 3. Lösungsschritt möglich. Betrachtet wird dabei die Folge der markierten Punkte (A, 1, 6, 3, 7, Z). Der Fluß von A nach 6 kann um 20 Einheiten erhöht werden, wenn (3,6) um 20 Einheiten vermindert wird.



Längs der markierten Knoten ist die kleinste Kapazität 10. Damit ergibt sich die folgende Aufteilung des Gesamtflusses:

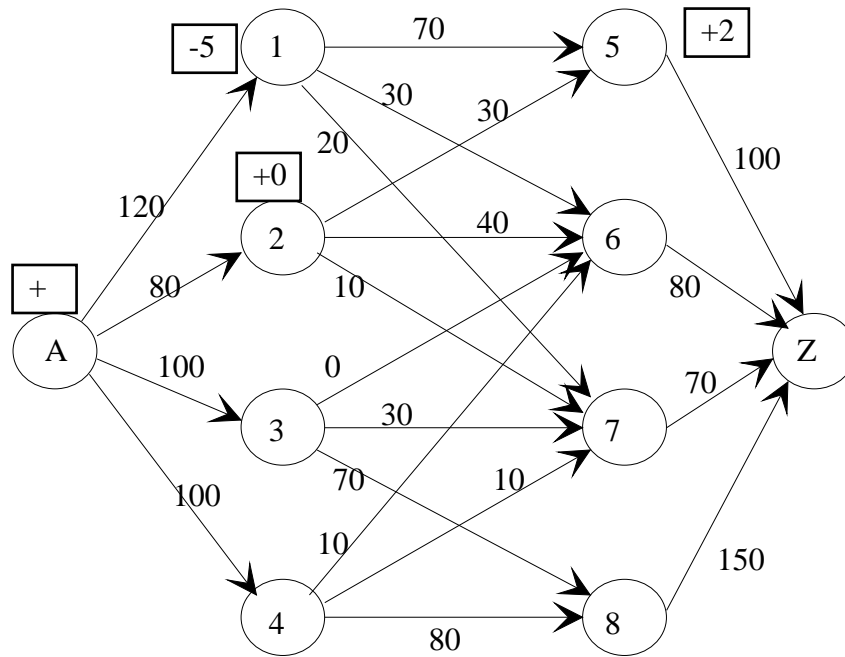


Abb.:

Betrachtet man die Menge aller nicht markierten Punkte (3, 4, 6, 7, 8, Z), dann sind alle Kanten, deren Ausgangspunkte markiert sind und die in der angegebenen Menge enden, gesättigt. Sonst hätte man wenigstens einen der Punkte aus der Menge markieren können. Daraus folgt die Lösung:

- Von 1 werden 70t nach 5 transportiert, 30 t nach 6, 20 t nach 7
- von 2 werden 40 t nach 6 transportiert, 10 t nach 7
- usw.

Optimum

Die Kapazität einer Schnittmenge ist gleich der Summe aus den Kapazitäten jeder ihrer Kanten. Da Z in der Schnittmenge enthalten ist, geht jeder Fluß notwendigerweise durch eine in der Schnittmenge hineingehenden Kanten. Deshalb ist der Fluß kleiner oder höchstens gleich der Kapazität einer Schnittmenge.

Findet man demzufolge einen Fluß, dessen Größe mit der Schnittmenge übereinstimmt, so ist der Fluß maximal und die Kapazität der Schnittmenge minimal.

Hauptsatz der Netztheorie

Ford und Fulkerson haben gezeigt (Hauptsatz der Netzwerktheorie, Max-Flow-Min-Cut - Theorem):

Für jedes Transportnetz gilt: $\Phi_{\max} = \text{Schnittmenge minimaler Kapazität}$

(Der maximale Wert eines Flusses ist gleich dem minimalen Wert seines Schnitts)

4.2.3.2 Kostenminimaler Gesamtfluß

4.2.3.2.1 Einführung

Häufig wird nicht nach einem Maximalfluß, sondern nach einem kostenminimalen Maximalfluß gefragt. Hier bestimmt man zunächst den maximalen Fluß ohne Rücksicht auf die Kosten und steuert anschließend die einfachen Flüsse so um, bis das Kostenminimum erreicht ist.

Bsp.: Gegeben ist das Verkehrsnetz

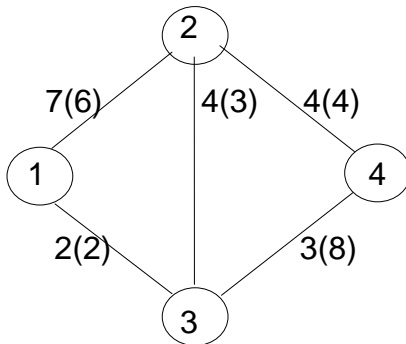


Abb.:

Jede Strecke des Netzes (Kante des Graphen) hat eine begrenzte Kapazität (bezeichnet durch die 1. Zahl an den Kanten). Die Zahl in den Klammern an den Kanten gibt die Kosten des Transports (je Einheit) an. Gesucht ist der maximale Fluß durch das Netz vom Knoten 1 zum Knoten 4, wobei die Kosten möglichst niedrig sein sollen.

1. Berechnung des maximalen Flusses ohne Berücksichtigung der Kosten

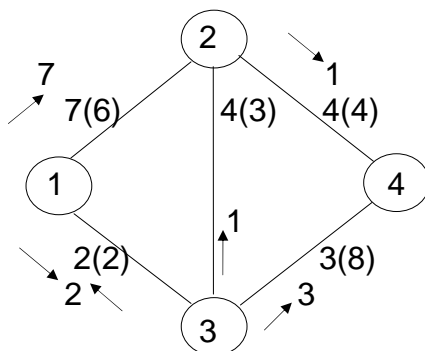


Abb.:

Der berechnete maximale Fluß besteht aus den Einzelflüssen:

- 7 (Einheiten) von 1 nach 2
- 3 (Einheiten) von 2 nach 3
- 4 (Einheiten) von 2 nach 4
- 3 (Einheiten) von 3 nach 4

Die Kosten betragen 91 [Kosteneinheiten]. Die Lösung ist nicht kostenminimal.

2. Kostenoptimale Lösung

Zwischen Knoten 1 und 3 bestehen 2 Alternativwege (1 - 2- 3 - 2) und (1 - 3). (1 - 3) wird nicht benutzt. Dort betragen die Kosten nur 2 [Kosteneinheiten]. Eine Umverteilung von 2 [Mengeneinheiten] führt hier zur Verbesserung. Man erhält die Optimallösung mit 77 [Kosteneinheiten].

4.2.3.2.2 Transportprobleme

1. Beschreibung

Gesucht werden hier die optimalen Transportzuordnungen eines homogenen Guts von einer Reihe von Angebotsorten (oder Lieferorten) zu einer Reihe von Bedarfsorten (oder Nachfrageorten). Gegeben sind die Angebotsmengen a_i der m Angebotsorte i sowie die Bedarfsmengen b_j der n Bedarfsorte j . Ferner sind die Kosten c_{ij} für den Transport je Mengeneinheit (ME) von den Angebotsorten i zu den Bedarfsorten j bekannt. x_{ij} sind die Transportmengen, die von den Orten i zu den Orten j befördert werden. Es gilt das folgende mathematische Modell:

(1) Nichtnegativitätsbedingungen (Distributform)

$$x_{ij} \geq 0 \quad (\text{für alle } i \text{ und } j)$$

(2) Nebenbedingungen

a) Angebotsgleichungen

$$\sum_{j=1}^n x_{ij} = a_i \quad (\text{für alle } i = 1, 2, \dots, m)$$

b) Nachfragegleichungen

$$\sum_{i=1}^m x_{ij} = b_j \quad (\text{für alle } j = 1, 2, \dots, n)$$

c) Gesamtangebot und Gesamtnachfrage gleichen sich aus:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

(3) Zielfunktion

Minimiere:

[illegible]

oder zusammengefaßt dargestellt

$$Z_{\min} = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

Rückführung von Verteilungsproblemen auf die lösbare Distributionsform

Die Bedingung für die Existenz einer Lösung ist:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

Besteht für das Verteilungsproblem die Ungleichung

$$\sum_{i=1}^m a_i \neq \sum_{j=1}^n b_j$$

dann haben die mathematischen Modelle folgende Form:

(1) $x_{ij} \geq 0$ (für alle i, j)

(2)

$$\sum_{i=1}^m a_i > \sum_{j=1}^n b_j \qquad \sum_{i=1}^m a_i < \sum_{j=1}^n b_j$$

a) $\sum_{j=1}^n x_{ij} \leq a_i \qquad \sum_{j=1}^n x_{ij} = a_i$

b) $\sum_{i=1}^m x_{ij} = b_j \qquad \sum_{i=1}^m x_{ij} \leq b_j$

(3) $Z_{\min} = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$

Zurückführung auf die Distributionsform

Die Ungleichungen sind durch Einführen von Schlupfvariablen zu Gleichungen zu ergänzen

$$(1) x_{ij} \geq 0 \text{ (für alle } i, j)$$

$$(2)$$

$$\begin{array}{ll} \sum_{j=1}^n x_{ij} + x_i = a_i & \sum_{j=1}^n x_{ij} = a_i \\ \sum_{i=1}^m x_{ij} = b_j & \sum_{i=1}^m x_{ij} + x_j = b_j \\ \sum_{i=1}^m x_i = \sum_{i=1}^m a_i - \sum_{j=1}^n b_j = b_{n+1} & \sum_{j=1}^n x_j = \sum_{j=1}^n b_j - \sum_{i=1}^m a_i = a_{m+1} \end{array}$$

fiktiver Besteller

fiktiver Anbieter

$$(3) Z_{\min} = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

Verbotstarife

Falls der Absender einen Besteller nicht beliefern kann, dann betrachtet man die Kosten zu der entsprechenden Relation als unendlich groß .

Solche Relationskosten heißen Verbotstarife und werden in der Transkostenmatrix durch den Buchstaben M bezeichnet.

Bsp.: Aus der folgenden Transportkostenmatrix

	B1	B2	B3
A1	M	1	10
A2	4	M	8

geht hervor:

A1 kann B1 nicht liefern

A2 kann B2 nicht liefern.

Verbotsfarife im Falle von $\sum_{i=1}^m a_i > \sum_{j=1}^n b_j$

Der Absender darf den fiktiven Besteller nicht beliefern, die Sperre erfolgt durch unendlich hohe Relationskosten.

Bsp.: Bei A1 darf kein Vorrat (für den fiktiven Besteller) übrig bleiben

	B1	B2	PHI	a_i
A1	$(x_{11})2$	$(x_{12})4$	$(x_{13})M$	10
A2	$(x_{21})3$	$(x_{22})5$	$(x_{23})0$	50
A3	$(x_{31})10$	$(x_{32})6$	$(x_{33})0$	200
b_j	100	100	60	

Verbotstarife im Falle von $\sum_{i=1}^m a_i < \sum_{j=1}^n b_j$

Der fiktive Absender darf dem Besteller nicht liefern, die Sperre erfolgt durch (unendlich) hohe Relationskosten.

Bsp.: Die Bestellmengen B1 und B4 müssen vollkommen erfüllt werden. Es ergibt sich dann die folgende Matrix der Transportmethode:

	B1	B2	B3	B4	b_j
A1	$(x_{11})2$	$(x_{12})1$	$(x_{13})4$	$(x_{14})5$	20
A2	$(x_{21})3$	$(x_{22})2$	$(x_{23})6$	$(x_{24})10$	50
A3	$(x_{31})7$	$(x_{32})1$	$(x_{33})9$	$(x_{34})4$	70
PHI	$(x_{41})M$	$(x_{42})0$	$(x_{43})0$	$(x_{44})M$	20
a_i	40	40	40	40	

Modifizierte Maximumsaufgabe

Sucht man in einem Verteilungsproblem eine Lösung bei der die Zielfunktion ein Maximum wird, d.h.

$$Z_{\max} = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

so gelangt man durch Umbenennen des Ziels (Multiplikation mit -1) zur üblichen Distributionsform.

2. Lösung von Transportproblemen

Bsp.: "Ein Transportproblem"

Die folgende Tabelle gibt die Transportkosten (c_{ij}), die Bedarfs- (b_j) und Angebotsmengen (a_i), die Bedarfs- (B_1, B_2, B_3, B_4) und Angebotsorte (A_1, A_2, A_3) an.

	B1	B2	3	B4	a_i
A1	100	220	90	140	90
A2	110	200	100	230	160
A3	30	100	160	140	80
b_j	70	60	70	130	330

Gesucht ist eine optimale Lösung für dieses Problem!

Dieses Problem läßt sich auch durch den folgenden Graphen beschreiben:

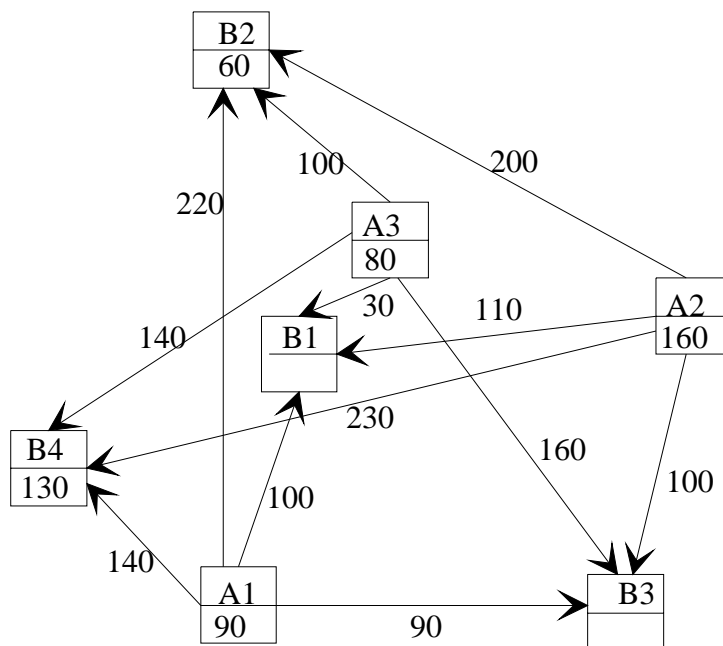


Abb.:

Bestimmen der ersten zulässigen Lösung (Zeilen-Spalten-Sukzession)

Ausgangspunkt ist die vorliegende Tabelle der Transportkosten. Es sind $m+n-1$ Basisvariable zu bestimmen (vgl.: die optimale Lösung des Transportproblems).

Die Zuordnung beginnt im kostengünstigsten Feld der gesamten Tabelle. Die maximal mögliche Menge wird zugeordnet (70 ME). Je nach Mengenüberschuß in

Zeile oder Spalte wird in der gleichen Zeile bzw. Spalte das nächstgünstigere Feld gesucht (A3 - B2). In der Überschuß-Spalte (B2) wird das nächstgünstigere Feld gesucht (A2 - B2). In der Überschußzeile (A2) erhält das nächstgünstigere Feld (A2 - B3) die Menge von 70 (ME). Es bleiben schließlich die Felder A2 - B4 bzw. A1 - B4, denen die Mengen von 40 bzw. 90 (ME) zugewiesen werden.

optimale Lösung?

Ein Teil der Lösung besteht in den Sendungen von A3 nach B1 und B2 sowie A2 nach B2. Der ihnen entsprechende Kantenzug besteht aus den hervorgehobenen Kanten von B1 über A3 und B2 nach A2. Fügt man in diesen offenen Kantenzug die Kante von A2 nach B1 ein, so erhält man einen geschlossenen Kantenzug.

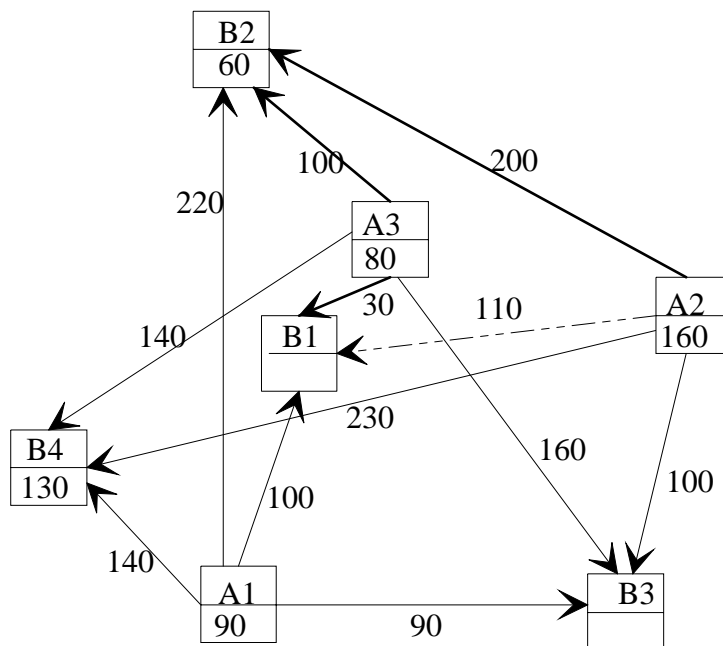


Abb.:

Innerhalb dieses geschlossenen Kantenzugs läßt sich der Fluß umverteilen, indem man B1 statt von A3 aus ganz oder teilweise von A2 aus beliefert. Entsprechend mehr müßte A3 an B2 liefern. Der Ort B2 wiederum benötigt dann eine entsprechend geringere Menge von A2. Diese ist schließlich gleich der Menge, die A2 dann an B1 schickt. In einer Tabelle läßt sich dieser Umverteilungsprozeß dann so darstellen:

		B1	B2	B3	B4	a_j
		0	70	-30	100	
A1	40	60	110	80	(90)	90
A2	130	$+\vartheta$ -20	$-\vartheta$ (50)	(70)	(40)	160
A3	30	$-\vartheta$ (70)	$+\vartheta$ (10)	160	10	80
b_i		70	60	70	130	330

Wann lohnt sich ein derartiger Umverteilung?

Eine derartige Umverteilung lohnt sich nur, wenn die Kosten sinken.

Jede ME, die von A2 nach B1 geht, kostet 110.- (GE). Gleichzeitig spart man 30.- GE für die Verringerung der Menge zwischen A3 und B1 ein. Entsprechend kommen 100.- GE dazu (zwischen A3 und B2) und schließlich gehen 200.- GE (zwischen A2 und B2) davon ab. Man erhält eine Summe von $(110 - 30 + 100 - 200) = -20$. Für jede umverteilte ME lassen sich die Kosten also um 20.- GE senken. Darum wird man möglichst viel umverteilen. Die maximal mögliche Menge ist gleich der kleinsten Transportmenge (ϑ) unter den Mengen, die verringert werden. Das wäre hier die Menge zwischen A2 und B2 von 50 (ME).

Man erhält folgende Lösung:

	B1	B2	3	B4	a_j
A1				(90)	90
A2	(50)		(70)	(40)	160
A3	(20)	(60)			80
b_i	70	60	70	130	330

Berechnung der Alternativkosten

Zu jedem nicht berücksichtigten Weg gibt es nur einen offenen Kantenzug mit berücksichtigten Wegen, deren Kosten zur Berechnung der Alternativkosten heranzuziehen sind. Für die Strecke von A2 nach B1 sind sie soeben mit -20 errechnet worden. Für andere Strecken kann man sie entsprechend ermitteln. Bspw. für die Strecke A1 nach B1: $100 - 30 + 100 - 200 + 230 - 140 = 60$.

Reduktion der Kosten

Der ausgewählte Transport zwischen A3 und B1 kostet 30 (GE)/(ME). Man kann diese Kosten dem Ort A3 anstatt der Strecke A3 - B1 zuordnen und die Kosten der von Ort A3 ausgehenden Strecken jeweils um 30 reduzieren. Das bedeutet:

- die Kosten von A3 nach B2 verringern sich auf 70
- die Kosten von A3 nach B3 verringern sich auf 130
- die Kosten von A3 nach B4 verringern sich auf 110

Zu diesen Kosten können jeweils die dem Ort A3 angelasteten Kosten von 30 hinzu. Der anschließend ausgewählte Transport von A3 nach B2 kostet 70 (GE)/(ME). Rechnet man diese Kosten dem Ort B2 zu, so sind die Kosten der zum Ort B2 führende Strecke von A1, A2, A3 auf 150, 130 bzw. 0 zu verringern. Fährt man weiter so fort, so sind den Orten A1, A2, A3 die Kosten 40, 130, 30 und den Orten B1, B2, B3, B4 die Kosten von 0, 70, -30 bzw. 100 zuzurechnen.

Generell gilt für alle $(m + n - 1)$ Basisvariablen (Reduktion der Kosten):

$$c_{ij} = u_i + v_j$$

(u_i, v_j sind die den Angebots- bzw. Bedarfsorten zugeordneten Kosten).

Das durch die vorstehende Gleichung gegebene lineare Gleichungssystem hat mit $m + n - 1$ linear unabhängigen Gleichungen und $m + n$ Variablen den Freiheitsgrad 1, d.h.: Bei Vorgabe einer Variablen sind die restlichen eindeutig zu berechnen. Der Wert einer Variablen (generell der Variablen v_1 wird immer deshalb mit 0 angegeben. Die restlichen Werte werden direkt in der Transporttabelle bestimmt. Für jedes Element, dem eine Nichtbasisvariable zugeordnet wurde, ergibt sich aus den u_i und v_j der Kostenveränderungswert:

$$\bar{c}_{ij} = c_{ij} - u_i - v_j$$

Falls alle $\bar{c}_{ij} > 0$, dann ist die Minimallösung gefunden. Negative Alternativkosten bedeuten: Eine Lösungsverbesserung ist möglich.

	B1 0	B2 70	B3 -10	B4 120	a _j
A1 20	80	130	80	(90)	90
A2 110	(50)	10	(70)	(40)	160
A3 30	(20)	(60)	140	-10	80
b _i	70	60	70	130	330

Eine weitere Lösungsverbesserung ergibt sich nach Berechnung der Alternativkosten durch Benutzen der Strecke A3 nach B4. Maximal können 20 ME umverteilt werden.

	B1 0	B2 80	B3 -10	B4 120	a _j
A1 20	80	120	80	(90)	90
A2 110	(70)	10	(70)	(20)	160
A3 20	10	(60)	150	(20)	80
b _i	70	60	70	130	330

Die jetzt erreichte Lösung ist die Optimallösung, denn sie erhält keine negativen Alternativkosten mehr.

4.2.3.2.3 Das Umladeproblem

1. Ein einführendes Beispiel

Beim Transportproblem sind lediglich direkte Transporte (vom Produzenten zum Verbraucher) erlaubt. Häufig führen Transporte über Zwischenstationen, die als Umschlagplätze dienen. Es gibt dann

Produzenten, z.B. P1, P3

Verbraucher, z.B. V4, V5

Umschlagplätze, z.B. U2, U6, U7, U8 und U9

P1 und P3 erzeugen $a_1 = 10$ (ME) bzw. $a_3 = 10$ (ME). V4 und V5 benötigen $b_4 = 5$ (ME) bzw. $b_5 = 15$ (ME).

Die verfügbaren Verbindungswege mit den Transportkosten für eine Mengeneinheit (ME) zeigt folgendes Netzwerk.

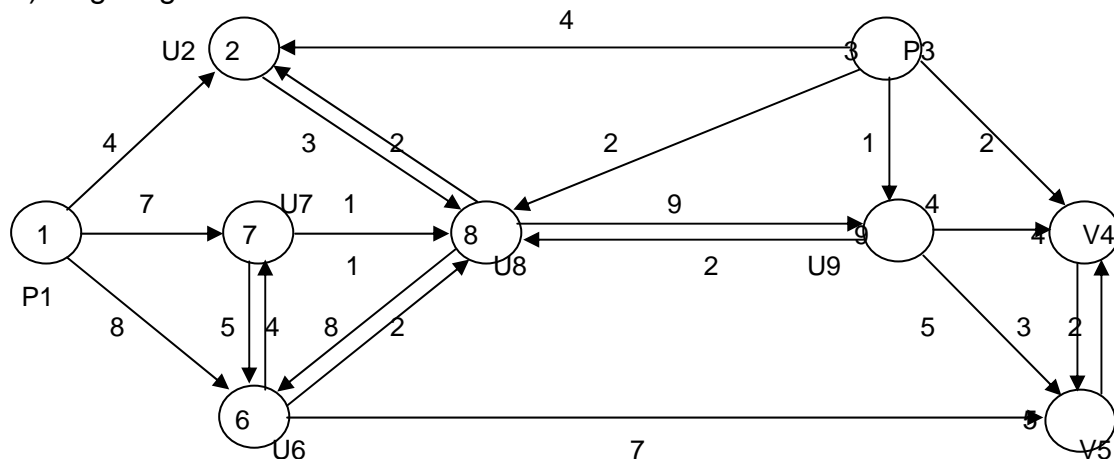


Abb.:

2. Verallgemeinerung

Generell sind in den Orten $O(m)$ ($m = 1, 2, \dots, l$) die Produzenten P_m und in den Orten $O(n)$ ($n = 1, 2, \dots, l$) die Verbraucher V_n . Die übrigen Orte dienen als Umschlagorte U_k . Auf den verfügbaren Verbindungswegen zwischen den Orten $O(i)$ und $O(j)$ gibt es Transportkosten zu c_{ij} (Geldeinheiten) je ME.

Es ist dann ein Transportplan zu erstellen, der den folgenden Bedingungen genügt:

- (1) Die Transportmengen sind nicht negativ
- (2) - Der Produktionsumfang eines jeden Produzenten wird ausgeschöpft
 - Die Nachfrage eines jeden Verbrauchers wird befriedigt
 - Der gesamte Produktionsumfang stimmt mit der gesamten Produktionsnachfrage überein

$$\sum_m a_m = \sum_n b_n$$

- (3) Minimale Transportkosten sind zu erreichen.

3. Das mathematische Modell

Vereinbarungen

- 1) Zuordnung eines Produktionsumfangs a_i und einer Produktionsnachfrage b_j zu jedem Ort $O(i)$ (unter dem Begriff $O(i)$ sind Produzenten und Umschlagplätze zusammengefaßt). " a_i, b_i " können 0 sein (reine Umschlagplätze)
- 2) Ergänzung der verfügbaren Transportwege um die nicht vorhandenen Transportwege zwischen 2 Orten (Zweck: übersichtliche Formulierung)
- 3) Belegung der nicht verfügbaren Verbindungswege mit ∞ Transportkosten.
" $c_{ij} = \infty$ ", falls keine direkte Verbindung vom Ort $O(i)$ zum Ort $O(j)$ ($i \neq j$) existiert.
- 4) x_{ij} (Anzahl der vom Ort $O(i)$ zum Ort $O(j)$ zu transportierenden ME)
- 5) " $\infty \cdot 0$ " (Variable zu den nachträglich hinzugefügten Transportwegen haben den wert 0)
- 6a) Summe der ankommenden bzw. selbstproduzierenden ME bezgl. eines jeden Ortes

$$\sum_{i=1, i \neq k}^l x_{ik} + a_k$$

- 6b) Summe der weitergeleiteten sowie selbstverwendeten ME bzgl. eines jeden Orts

$$\sum_{j=1, j \neq k}^l x_{kj} + b_k$$

Mathematisches Modell

$$(1) x_{ij} \geq 0$$

$$(2a) \sum_{i=1, i \neq k}^l x_{ik} - \sum_{j=1, j \neq k}^l x_{kj} = b_k - a_k \quad (k = 1, 2, \dots, l)$$

$$(2b) \sum_{i=1}^l a_i = \sum_{j=1}^l b_j$$

$$(3) Z_{\min} = \sum_{i=1}^l \sum_{j=1}^l c_{ij} \cdot x_{ij}$$

Mathematisches Modell zum einführenden Beispiel:

$$(1) x_{ij} \geq 0, i \neq j \quad (i = 1, 2, \dots, 9), (j = 1, 2, \dots, 9)$$

$$(2) \begin{aligned} -x_{12} - x_{16} - x_{17} &= -10 \\ -x_{12} + x_{32} + x_{82} - x_{28} &= 0 \\ -x_{32} - x_{34} - x_{38} - x_{39} &= -10 \\ x_{34} + x_{54} + x_{94} - x_{45} &= 5 \\ x_{45} + x_{65} + x_{95} - x_{54} &= 15 \\ x_{16} + x_{76} + x_{86} - x_{65} - x_{67} - x_{68} &= 0 \\ x_{17} + x_{67} + x_{76} - x_{78} &= 0 \\ x_{28} + x_{38} + x_{68} + x_{98} - x_{82} - x_{86} - x_{89} &= 0 \\ x_{39} + x_{89} - x_{94} - x_{95} - x_{98} &= 0 \end{aligned}$$

$$(3) Z_{\min} = 4x_{12} + 8x_{16} + 7x_{17} + 3x_{28} + 4x_{32} + 2x_{34} + 2x_{38} +$$

$$x_{39} + 3x_{45} + 2x_{54} + 7x_{65} + 4x_{67} + 2x_{68} + 5x_{76} + \\ x_{78} + 2x_{82} + 8x_{86} + 9x_{89} + 4x_{94} + 5x_{95} + 2x_{98}$$

4. Algorithmus zum Umladeproblem

Definition

\bar{x}_{kk} : Gesamtvorrat im Orte $O(k)$, $k = 1, 2, \dots, l$

Damit nimmt das mathematische Modell folgende Gestalt an:

$$(1) \ x_{ij} \geq 0$$

$$\bar{x}_{kk} \geq 0$$

(2)

$$\sum_{i=1, i \neq k}^l x_{ik} - \bar{x}_{kk} = -a_k \quad k = 1, 2, \dots, l$$

$$\sum_{j=1, j \neq k}^l x_{kj} - \bar{x}_{kk} = -b_k \quad k = 1, 2, \dots, l$$

$$\sum_{i=1}^l a_i = \sum_{j=1}^l b_j$$

$$(3) \ Z_{\min} = \sum_{i=1}^l \sum_{j=1}^l c_{ij} \cdot x_{ij}$$

Vereinbarungen:

1. Den Variablen \bar{x}_{kk} werden die Koeffizienten der Kostenmatrix $c_{kk} = 0$ zugeordnet
2. Nicht vorhandene Verbindungen zwischen 2 Orten i und j erhalten die Variable $x_{ij} = 0$ zugeordnet.

Beispiel:

j	1	2	3	4	5	6	7	8	9	-b _k
i										
1	\bar{x}_{11}	x ₁₂	α	α	α	x ₁₆	x ₁₇	α	α	0
2	α	\bar{x}_{22}	α	α	α	α	α	x ₂₈	α	0
3	α	x ₃₂	\bar{x}_{33}	x ₃₄	α	α	α	x ₃₈	x ₃₉	0
4	α	α	α	\bar{x}_{44}	x ₄₅	α	α	α	α	-5
5	α	α	α	x ₅₄	\bar{x}_{55}	α	α	α	α	-15
6	α	α	α	α	x ₆₅	\bar{x}_{66}	x ₆₇	x ₆₈	α	0
7	α	α	α	α	α	x ₇₆	\bar{x}_{77}	x ₇₈	α	0
8	α	x ₈₂	α	α	α	x ₈₆	α	\bar{x}_{88}	x ₈₉	0
9	α	α	α	x ₉₄	x ₉₅	α	α	x ₉₈	\bar{x}_{99}	0
-a _k	-10	0	-10	0	0	0	0	0	0	

Unterschiede zum allgemeinen Transportproblem

1. Die Einschränkung $0 \leq x_{ij} \leq \min(a_i, b_j)$ gilt nicht
2. Variable können nicht mit negativem Vorzeichen auftreten.

Überführung des Umladeproblems in ein "klassisches Transportproblem"

Es gilt folgender Satz: Sind in einem Umladeproblem sämtliche $c_{ij} \geq 0$ und besitzt das Problem eine optimale Lösung, dann gilt für die Variablen \bar{x}_{kk} in einer optimalen Lösung die Beziehung:

$$\bar{x}_{kk} \leq S + a_k + b_k$$

Der Schlupf S berechnet sich aus: $S = \sum_{i=1}^l a_i = \sum_{j=1}^l b_j$

\bar{x}_{kk} kann damit ersetzt werden: $\bar{x}_{kk} = S + a_k + b_k - x_{kk}$

Das in ein "klassisches Transportproblem" überführte Umladeproblem ist dann:

(1) $x_{ij} \geq 0$

(2) $\sum_{i=1}^l x_{ik} = S + b_k$

$$\sum_{j=1}^l x_{kj} = S + a_k$$

$$(3) Z_{\min} = \sum_{i=1}^I \sum_{j=1}^I c_{ij} \cdot x_{ij}$$

Das zum Umladeproblem des einführenden Beispiels äquivalente Transportproblem ist dann:

j	1	2	3	4	5	6	7	8	9	S+a _k
i										
1	(x ₁₁)	(x ₁₂)	-----	-----	-----	(x ₁₆)	(x ₁₇)	-----	-----	30
2	-----	(x ₂₂)	-----	-----	-----	-----	-----	(x ₂₈)	-----	20
3	-----	(x ₃₂)	(x ₃₃)	(x ₃₄)	-----	-----	-----	(x ₃₈)	(x ₃₉)	30
4	-----	-----	-----	(x ₄₄)	(x ₄₅)	-----	-----	-----	-----	20
5	-----	-----	-----	(x ₅₄)	(x ₅₅)	-----	-----	-----	-----	20
6	-----	-----	-----	-----	(x ₆₅)	(x ₆₆)	(x ₆₇)	(x ₆₈)	-----	20
7	-----	-----	-----	-----	-----	(x ₇₆)	(x ₇₇)	(x ₇₈)	-----	20
8	-----	(x ₈₂)	-----	-----	-----	(x ₈₆)	-----	(x ₈₈)	(x ₈₉)	20
9	-----	-----	-----	(x ₉₄)	(x ₉₅)	-----	-----	(x ₉₈)	(x ₉₉)	20
S+b _k	20	20	20	25	35	20	20	20	20	

5. Zulässige Basislösung für das Umladeproblem

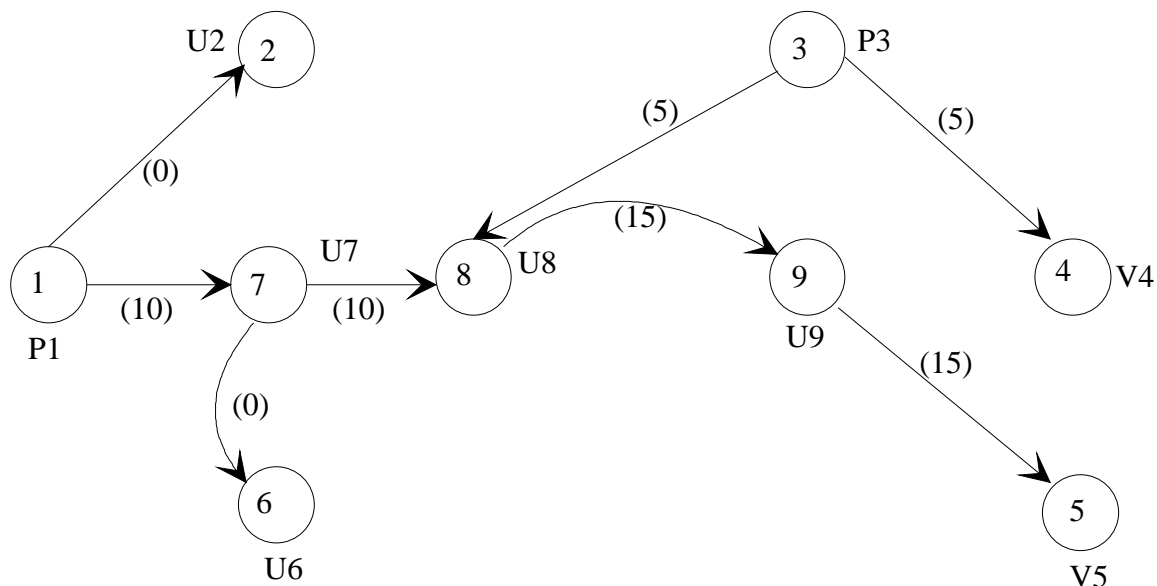
Aus der Beziehung $\bar{x}_{kk} \leq S + a_k + b_k$ folgt für jede Zahl $d > 0$:

$$\bar{x}_{kk} \leq S + a_k + b_k + d, \quad x_{kk} = S + a_k + b_k + d - \bar{x}_{kk}$$

Wegen $\bar{x}_{kk} \leq S + a_k + b_k + d$ sind die x_{kk} positiv und gehören damit stets zur Basis:

Es gilt folgender Satz: In der Darstellung des Umladeproblems als "klassisches Transportproblem" können die Variablen x_{kk} stets als Basisvariablen betrachtet werden. Damit sind l Variable x_{kk} , die zu einer Basis gehören, bekannt.

Zur Bestimmung der restlichen $(l - 1)$ Variablen der Basislösung betrachtet man das folgende Teilnetzwerk, in dem jeder Knoten mit einem anderen verbunden ist:



Die in dem Teilnetz errechneten Angaben trägt man in die Transporttabelle ein. Anschließend bestimmt man die Werte der Variablen x_{kk} .

Die Potentiale u_i und v_j können beim Umladeproblem so gewählt werden, daß sie die folgende Beziehung erfüllen:

$$-u_k = v_k \quad \text{für } k = 1, 2, \dots, l$$

Beweis:

$$c_{kk} = u_i + v_j - c_{kk} \text{ gilt für alle Basisvariablen}$$

x_{kk} ist in jeder basismenge vorhanden

c_{kk}, \bar{c}_{kk} sind immer 0, es gilt die angegebene Beziehung

Die gemeinsamen Werte von $v_k = -u_k$ heißen Π_k

Transporttabelle zur 1. zulässigen Basislösung

j Π_k	1 0	2 4	3 6	4 8	5 22	6 12	7 7	8 8	9 17	S+a _k
i - Π_k										
1 0	(20)	(0)	-----	-----	-----	-4	(10)	-----	-----	30
2 -4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3 -6	-----	6	(20)	(5)	-----	-----	-----	(5)	-10	30
4 -8	-----	-----	-----	(20)	-11	-----	-----	-----	-----	20
5 -22	-----	-----	-----	16	(20)	-----	-----	-----	-----	20
6 -12	-----	-----	-----	-----	-3	(20)	9	6	-----	20
7 -7	-----	-----	-----	-----	-----	(0)	(10)	(10)	-----	20
8 -8	-----	6	-----	-----	-----	4	-----	(5)	(15)	20
9 -17	-----	-----	-----	13	(15)	-----	-----	11	(5)	20
S+b _k	20	20	20	25	35	20	20	20	20	

j Π_k	1 0	2 4	3 6	4 8	5 22	6 12	7 7	8 8	9 17	S+a _k
i - Π_k										
1 0	(20)	(0)	-----	-----	-----	-4	(10)	-----	-----	30
2 -4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3 -6	-----	6	(20)	(5)	-----	-----	-----	(5)	-10	30
4 -8	-----	-----	-----	(20)	-11	-----	-----	-----	-----	20
5 -22	-----	-----	-----	16	(20)	-----	-----	-----	-----	20
6 -12	-----	-----	-----	-----	-3	(20)	9	6	-----	20
7 -7	-----	-----	-----	-----	-----	(0)	(10)	(10)	-----	20
8 -8	-----	6	-----	-----	-----	4	-----	(5)	(15)	20
9 -17	-----	-----	-----	13	(15)	-----	-----	11	(5)	20
S+b _k	20	20	20	25	35	20	20	20	20	

j Π_k $i - \Pi_k$	1 0	2 4	3 17	4 19	5 22	6 12	7 7	8 8	9 17	$S+a_k$
1 0	(20)	(0)	-----	-----	-----	-4	(10)	-----	-----	30
2 -4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3 -17	-----	6	(20)	(5)	-----	-----	-----	11	1	30
4 -19	-----	-----	-----	(15)	(5)	-----	-----	-----	-----	20
5 -22	-----	-----	-----	5	(20)	-----	-----	-----	-----	20
6 -12	-----	-----	-----	-----	-3	(20)	9	6	-----	20
7 -7	-----	-----	-----	-----	-----	(0)	(10)	(10)	-----	20
8 -8	-----	6	-----	-----	-----	4	-----	(10)	(10)	20
9 -17	-----	-----	-----	2	(10)	-----	-----	11	(10)	20
$S+b_k$	20	20	20	25	35	20	20	20	20	

j Π_k $i - \Pi_k$	1 0	2 4	3 17	4 19	5 22	6 8	7 7	8 8	9 17	$S+a_k$
1 0	(20)	(0)	-----	-----	-----	(0)	(10)	-----	-----	30
2 -4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3 -17	-----	6	(20)	(5)	-----	-----	-----	11	1	30
4 -19	-----	-----	-----	(20)	(5)	-----	-----	-----	-----	20
5 -22	-----	-----	-----	5	(20)	-----	-----	-----	-----	20
6 -8	-----	-----	-----	-----	-7	(20)	5	2	-----	20
7 -7	-----	-----	-----	-----	-----	8	(10)	(10)	-----	20
8 -8	-----	6	-----	-----	-----	8	-----	(10)	(10)	20
9 -17	-----	-----	-----	2	(10)	-----	-----	11	(10)	20
$S+b_k$	20	20	20	25	35	20	20	20	20	

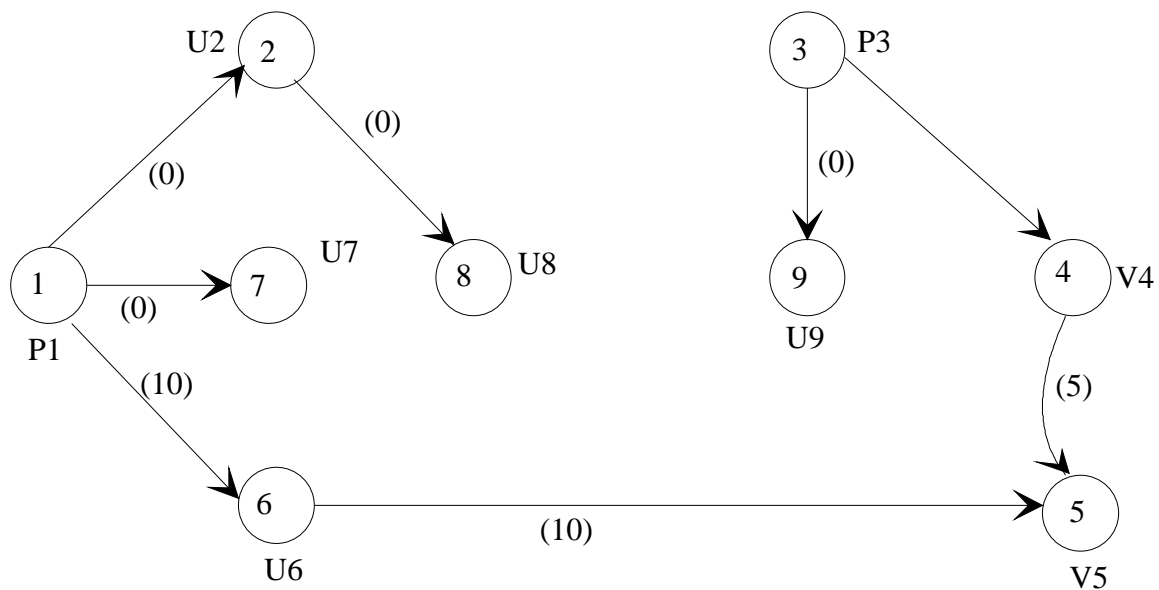
j Π_k $i - \Pi_k$	1 0	2 4	3 10	4 12	5 15	6 8	7 7	8 8	9 17	$S+a_k$
1 0	(20)	(0)	-----	-----	-----	(10)	(0)	-----	-----	30
2 -4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3 -10	-----	-----	(20)	(10)	-----	-----	-----	4	-6	30
4 -12	-----	-----	-----	(15)	(5)	-----	-----	-----	-----	20
5 -15	-----	-----	-----	-----	(20)	-----	-----	-----	-----	20
6 -8	-----	-----	-----	-----	(10)	(10)	5	2	-----	20
7 -7	-----	-----	-----	-----	-----	4	(20)	(0)	-----	20
8 -8	-----	-----	-----	-----	-----	8	-----	(20)	(0)	20
9 -17	-----	-----	-----	9	7	-----	-----	11	(20)	20
$S+b_k$	20	20	20	25	35	20	20	20	20	

j Π_k $i - \Pi_k$	1 0	2 4	3 10	4 12	5 15	6 8	7 7	8 8	9 11	$S+a_k$
1 0	(20)	(0)	-----	-----	-----	(10)	(0)	-----	-----	30

2	-4	-----	(20)	-----	-----	-----	-----	-----	-1	-----	20
3	-6	-----	10	(20)	(10)	-----	-----	-----	4	(0)	30
4	-8	-----	-----	-----	(15)	(5)	-----	-----	-----	-----	20
5	-22	-----	-----	-----	5	(20)	-----	-----	-----	-----	20
6	-12	-----	-----	-----	-----	(10)	(10)	5	2	-----	20
7	-7	-----	-----	-----	-----	-----	4	(20)	(0)	-----	20
8	-8	-----	6	-----	-----	-----	8	-----	(20)	6	20
9	-11	-----	-----	-----	3	1	-----	-----	5	(20)	20
S+b _k	20	20	20	25	35	20	20	20	20	20	

$\begin{matrix} j \\ \Pi_k \\ i - \Pi_k \end{matrix}$	1	2	3	4	5	6	7	8	9		S+a _k
Π_k	0	4	10	12	15	8	7	7	11		
$i - \Pi_k$	0	(20)	(0)	-----	-----	-----	(10)	(0)	-----	-----	30
1	-4	-----	(20)	-----	-----	-----	-----	-----	(0)	-----	20
2	-10	-----	10	(20)	(10)	-----	-----	-----	5	(0)	30
3	-12	-----	-----	-----	(15)	(5)	-----	-----	-----	-----	20
4	-15	-----	-----	-----	5	(20)	-----	-----	-----	-----	20
5	-8	-----	-----	-----	-----	(10)	(10)	5	3	-----	20
6	-7	-----	-----	-----	-----	-----	4	(20)	1	-----	20
7	-7	-----	6	-----	-----	-----	7	-----	(20)	7	20
8	-11	-----	-----	-----	3	1	-----	-----	6	(20)	20
9	-11	-----	-----	-----	3	1	-----	-----	6	(20)	20
S+b _k	20	20	20	25	35	20	20	20	20	20	

Teilnetzwerk zur optimalen Lösung



$$Z_{\min} = 10 \cdot 8 + 10 \cdot 7 + 10 \cdot 2 + 5 \cdot 3 = 185 \text{ [GE]}$$

4.2.3.2.4 Zuordnungsprobleme

1. Beschreibung

Zuordnungsprobleme sind bspw. von der folgenden Art:

Ein Werkstück hat bis zu seiner Fertigstellung 4 Arbeitsplätze in der Folge A1, A2, A3 und A4 zu durchlaufen. Zur Besetzung der Arbeitsplätze stehen 4 Personen P1, P2, P3 und P4 zur Verfügung, die jedoch bzgl. der verschiedenen Arbeitsplätze unterschiedliche Qualifikationen aufweisen. In der folgenden Tabelle sind die Arbeitszeiten der verschiedenen Personen an den 4 Arbeitsplätzen zur Fertigung des Werkstücks zusammengestellt:

	A1	A2	A3	A4
P1	9	7	8	6
P2	6	2	6	7
P3	7	1	3	8
P4	1	3	3	5

Die Aufgabe besteht darin, die Verteilung der Personen auf die Arbeitsplätze in der folgenden Weise vorzunehmen:

- Jeder Person ist genau ein Arbeitsplatz zuzuordnen
- Die Gesamtarbeitszeit zur Fertigstellung des Werkstücks ist zu minimieren.

2. Formulierung

Das Zuordnungsproblem besteht in der optimalen Zuordnung von n Elementen zu n anderen Elementen einer Menge, z.B.:

Der Person P_i wird entweder der Arbeitsplatz A_j zugeordnet oder nicht

$$(1) \ x_{ij} = \begin{cases} 1, & \text{falls der Person } P_i \text{ Arbeitsplatz } A_j \text{ zugeordnet wird} \\ 0 & \end{cases}$$

(2)

a) Jedem P_i muß ein Arbeitsplatz zugewiesen werden

$$\sum_{i=1}^n x_{ij} = 1 \quad (j = 1, 2, \dots, n)$$

b) Jedes A_j muß belegt werden

$$\sum_{i=1}^n x_{ij} = 1 \quad (i = 1, 2, \dots, n)$$

(3) Zielforderung (Minimierung der Gesamtbearbeitungszeit)

$$Z_{\min} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}$$

Die folgende Zuordnungstabelle legt dann die problemspezifischen Aufgaben für das Zuordnungsproblem fest:

	A_1	A_2	A_3	A_n
P_1	c_{11}	c_{12}	c_{13}	c_{1n}

P_2	c_{21}	c_{22}	c_{23}	...	c_{2n}
....
....
P_n	c_{n1}	c_{n2}	c_{n3}	...	c_{nn}

3. Standardmodell zum einführenden Beispiel

(1)

$$x_{ij} = \begin{cases} 0 & (i = 1, 2, 3, 4) \\ 1 & (j = 1, 2, 3, 4) \end{cases}$$

(2)

$$\begin{aligned} \text{a) } x_{11} + x_{12} + x_{13} + x_{14} &= 1 \\ x_{21} + x_{22} + x_{23} + x_{24} &= 1 \\ x_{31} + x_{32} + x_{33} + x_{34} &= 1 \\ x_{41} + x_{42} + x_{43} + x_{44} &= 1 \end{aligned}$$

$$\begin{aligned} \text{b) } x_{11} + x_{21} + x_{31} + x_{41} &= 1 \\ x_{12} + x_{22} + x_{32} + x_{42} &= 1 \\ x_{13} + x_{23} + x_{33} + x_{43} &= 1 \\ x_{14} + x_{24} + x_{34} + x_{44} &= 1 \end{aligned}$$

(3)

$$Z_{\min} = 9x_{11} + 7x_{12} + 8x_{13} + 8x_{14} + \dots$$

4. Maximumproblem

Das Zuordnungsproblem läßt sich auch folgendermaßen formulieren:

Man ordne n Personen genau n Aufgaben zu, daß jeder Person nur eine Aufgabe zugeteilt wird. Der Wert c_{ij} der Zuordnung der i -ten Person zur j -ten Aufgabe ist durch Leistungstests feststellbar. Man bestimme x_{ij} so, daß

$$Z_{\max} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}$$

maximal wird.

Mathematisches Modell

$$(1) \quad x_{ij} = \begin{cases} 1, & \text{falls der Person } P_i \text{ die } j\text{-te Aufgabe } A_j \text{ zugeordnet wird} \\ 0 & \end{cases}$$

(2)

a) Jedem P_i muß eine Aufgabe zugewiesen werden

$$\sum_{i=1}^n x_{ij} = 1 \quad (j = 1, 2, \dots, n)$$

b) Jedes A_j muß belegt werden

$$\sum_{i=1}^n x_{ij} = 1 \quad (i = 1, 2, \dots, n)$$

(3) Zielforderung

$$Z_{\max} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}$$

5. Lösung von Zuordnungsproblemen

Zur Lösung von Zuordnungsproblemen wird die "**Ungarische Methode**" (benannt nach dem Theorem des ungarischen Mathematikers König) benutzt. Die Methode stützt sich auf das folgende grundlegende Prinzip:

Die optimale Lösung eines Zuordnungsproblems ändert sich nicht, wenn man in der Matrix c_{ij} alle Elemente einer Zeile (oder Spalte) um den gleichen Wert erhöht (oder vermindert). Das ist selbstverständlich, da eine mögliche Lösung nur ein Element mit dem Wert 1 je Zeile oder Spalte enthalten kann. Erhöhen (bzw. vermindern) um einen konstanten Wert kann nur die Gesamtsumme einer Lösung ändern, nicht aber die optimale Lösung. Der 1. Lösungsschritt kann demnach in der Erzeugung von Nullelementen bestehen.

Lösungsschritt 1: "Erzeugen von Nullelementen"

$$c_{ij}^{(1)} = c_{ij} - \min(c_{ij}) \quad (i: \text{Zeilenindex}, j: \text{Spaltenindex})$$

Von allen Elementen jeder Spalte wird das kleinste Element dieser Spalte subtrahiert. Die neue Tabelle ist $c_{ij}^{(1)}$. Man erhält mindestens ein Nullelement je Spalte. Danach behandelt man die Zeilen auf die gleiche Weise:

$$c_{ij}^{(2)} = c_{ij}^{(1)} - \min(c_{ij}^{(1)})$$

Lösungsschritt 2: "Suche nach der optimalen Lösung"

Eine optimale Lösung erfüllt die Bedingung $\sum_{i=1}^n \sum_{j=1}^n c_{ij}^{(2)} x_{ij} = 0$

Eine solche Lösung wird erreicht, falls eine Zuordnung so festgelegt werden kann, daß alle $c_{ij}^{(2)}$, die Bestandteil einer Lösung sind, den Wert 0 haben. Dazu wird zunächst eine Zeile (oder eine der Zeilen) betrachtet, die die geringste Anzahl an 0-Elementen enthält. Um eines der 0 Elemente wird ein Kästchen gezeichnet und alle Nullen gestrichen, die sich in der gleichen Zeile oder Spalte wie die gerade eingerahmten Nullen befinden. Die weiteren Zeilen werden der gleichen Prozedur unterzogen, bis keine Nullen mehr eingerahmt werden können. Die Anwendung des Lösungsschritts auf das folgende Beispiel ergibt:

Bsp.:

Gegeben ist die Kostenmatrix c_{ij} :

17.5	15	9	5.5	12
16	16.5	10.5	5	10.5
12	15.5	14.5	11	5.5
5.5	8	14	17.5	13
13	9.5	8.5	12	17.5

Diese Kostenmatrix kann reduziert werden zu $c_{ij}^{(1)}$

13	7	0.5	0.5	6.5
11.5	8.5	3	0	5
7.5	7.5	6	6	0
0	0	5.5	12.5	7.5
8.5	1.5	0	7	12

Eine weitere Reduktion führt zu $c_{ij}^{(2)}$

12.5	6.5	0	0	6.0
11.5	8.5	3	0	5
7.5	7.5	6	6	0
0	0	5.5	12.5	7.5
8.5	1.5	0	7	12

Wurde hier tatsächlich die größtmögliche Anzahl von Nullelementen (maximaler Verbund) zugeordnet?

Die ungarische Methode beruht auf dem Satz von D. König:

Für jede quadratische Matrix, deren Elemente teils Null, teils von Null verschiedene positive Zahlen sind, ist die minimale Anzahl der **Decklinien** gleich der maximalen Anzahl der **unabhängigen Punkte**.

Decklinien sind Nullen führende Zeilen und Spalten. **Unabhängige Punkte** sind solche Nullelemente, die keinen Zeilen und Spalten gemeinsam haben.

Ein Zuordnungsproblem ist dann gelöst, wenn n unabhängige Punkte gefunden sind. Sie garantieren als Nullelemente die kleinste Kostensumme.

Nach dem Algorithmus der ungarischen Methode müssen im vorstehenden Beispiel 5 unabhängige Nullelemente entstehen. Die Konstruktion der Decklinien erfolgt über die Lösungsschritte 3 und 4:

Lösungsschritt 3: Bestimmen der minimalen Anzahl Zeilen und Spalten, die sämtliche Nullelemente enthalten

1. Kennzeichnen (mit einem Kreuz) alle Zeilen, die kein gerahmtes Nullelement enthalten
2. Kennzeichnen in gleicher Weise jede Spalte, die in einer oder mehreren der unter 1. markierten Zeilen ein gestrichenes Nullelement enthält.
3. Kennzeichnen jeder Zeile, die ein eingerahmtes Nullelement in einer markierten Spalte enthält.
4. Wiederhole 2. und 3. bis keine Zeile oder Spalte mehr markiert werden kann.

Lösungsschritt 4: Konstruktion der Deckungslinien

Jede nicht markierte Zeile und jede markierte Spalte wird gestrichen. Man bestimmt dadurch die minimale Anzahl von Zeilen und Spalten, die alle eingerahmten oder gestrichenen Nullen enthalten (Deckungslinien).

12.5	6.5	0	0	6	X
11.5	8.5	2	0	5	X
7.5	7.5	6	6	0	
0	0	5.5	12.5	7.5	
8.5	1.5	0	7	12	X
		X	X		

Lösungsschritt 5:

Betrachtet werden nun alle nicht gestrichenen Elemente. Daraus ist das Element mit dem kleinsten Wert zu bestimmen. Dieser Zahlenwert ist von allen Elementen nicht gestrichener Spalten zu subtrahieren und zu allen Elementen von gestrichenen Zeilen zu addieren.

Daraus ergibt sich (auf das vorliegende Beispiel bezogen) die Matrix

12.5 -1.5	6.5 -1.5	0	0	6 -1.5	X
11.5 -1.5	8.5 -1.5	2	0	5 -1.5	X
+1.5 7.5 -1.5	+1.5 7.5 -1.5	+1.5 6	+1.5 6	+1.5 0 -1.5	
+1.5 0 -1.5	+1.5 0 -1.5	+1.5 5.5	+1.5 12.5	+1.5 7.5 -1.5	
8.5 -1.5	1.5 -1.5	0	7 -1.5	12 -1.5	X
		X	X		

Das ergibt, auf das vorliegende Beispiel bezogen, die Matrix $c_{ij}^{(3)}$:

11	5	<div>0</div>	<div>0</div>	6
10	8.5	2	<div>0</div>	5
7.5	7.5	6	6	<div>0</div>
<div>0</div>	<div>0</div>	5.5	12.5	7.5
7	<div>0</div>	<div>0</div>	7	12

Hier ergibt sich bereits die optimale Lösung.

0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
1	0	0	0	0
0	1	0	0	0

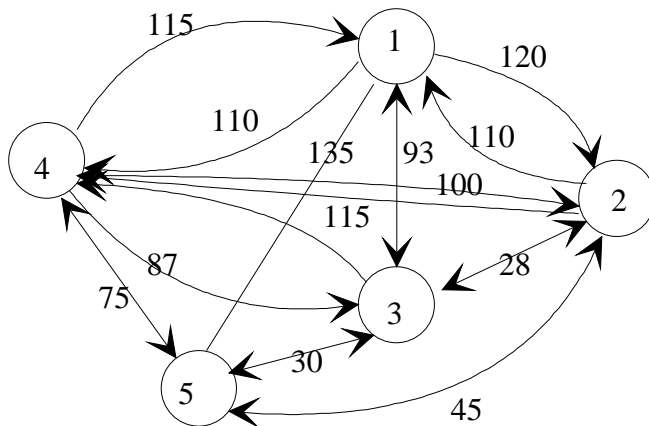
8. Die Anwendung der ungarischen Methode auf zyklische Probleme

- Das Travelling Salesman Problem

Man hat versucht, die ungarische Methode auf gewisse zyklische Probleme und insbesondere auf das berühmte Problem des Handlungsreisenden anzuwenden:

Gegeben sind n Städte 1, 2, 3, ..., n , deren jeweilige Entfernung zueinander bekannt ist (oder für die auch die Transportkosten von einer Stadt zu anderen bekannt sind).

Gesucht ist die wirtschaftlichste Rundfahrt mit dem gleichen Anfangs- und Endpunkt, bei der jede Stadt einmal berührt wird.



Der Reisende befindet sich im Ausgangsort 1. Er soll die übrigen Orte (Knoten 2 bis 5) so anfahren, daß die zurückgelegten Entfernungen minimal sind.

Entfernungsmatrix

	1	2	3	4	5
1	∞	120	93	110	135
2	110	∞	28	115	45
3	93	28	∞	87	30
4	115	100	87	∞	75
5	135	45	30	75	∞

- Formulierung des Problems als Zuordnungsproblem

Ähnlichkeiten sind:

1. Es gibt eine Kostenmatrix, deren Anzahl Zeilen und Spalten der Anzahl Städte entspricht ($c_{ij} = \infty$, da man voraussetzungsgemäß nicht in einer Stadt bleibt)
2. Eine Zuordnung der zeilen zu den Spalten sucht. Diese Zuordnung muß jedoch so beschaffen sein, daß die Folge der Zeilen- und Spaltenindizes eine vollständige Rundreise ergibt.

Das TSP-Problem kann in folgender Form als Zuordnungsmatrix angegeben werden: c_{ij} sind die Kosten (Entfernungen, Zeiten) für den Weg i nach j . Mit $x_{ij} = 1$ bzw. 0 ist gekennzeichnet, ob die Folge " $i - j$ " in einer Lösung erhalten ist ($= 1$) oder nicht ($= 0$).

Es gilt:

(1) Nichtnegativitätsbedingungen oder Ganzzahligkeitsbedingung

$$x_{ij} = \begin{cases} 0 \\ 1 \end{cases}$$

(2) Restriktionen

Jeder Ort kann nur einmal angereist werden. Damit gilt für jeden Ort j die Bedingung:

$$\sum_i x_{ij} = 1 \quad \text{für } j = 1, 2, \dots, n$$

Außerdem muß man von jedem Ort einmal zu einem anderen Ort fahren. Daher gilt für jeden Ort i : $\sum_j x_{ij} = 1$

Diese Restriktionen bewirken, daß jeder Ort sowohl einen Vorgänger als auch einen Nachfolger erhält. Sie verhindern aber nicht das Auftreten von Kurzzyklen.

Kurzzyklen sind Rundreisen mit weniger als n Orten. Sie können durch folgende Restriktionen ausgeschlossen werden:

$x_{ij} = 0$ (Solche Zyklen wurden bereits verhindert, da die Diagonalelemente c_{ij} der Entfernungsmatrix auf ∞ gesetzt wurden)

Kurzzyklen zwischen 2 Orten können durch die Restriktionen $x_{ij} + x_{jk} + x_{ki} \leq 2$ verhindert werden.

Das kann fortgesetzt werden, führt aber zu einer sehr großen Anzahl von Ungleichungen. So hat man bereits für $n=13$ Orte 241738 Bedingungen und 169 Variable errechnet.

Man kann aber trotzdem die Verwandtschaft des TSP zum Zuordnungsproblem nutzen. Die beiden Probleme unterscheiden sich nur durch die zusätzlichen Restriktionen. Das bedeutet: Der Wert der Zielfunktion kann nie kleiner sein als der der Optimallösung des Zuordnungsproblems. Die Lösung des Zuordnungsproblems ist somit eine Art Untergrenze, die vom TSP lediglich erreicht, aber nie unterschritten werden kann.

Behandlung des TSP als Zuordnungsproblem

Gegeben ist die Entfernungsmatrix

	1	2	3	4	5
1	∞	120	93	110	135
2	110	∞	28	115	45
3	93	28	∞	87	30
4	115	100	87	∞	75
5	135	45	30	75	∞

Gesucht ist die wirtschaftlichste Rundfahrt!

Reduktion der Entfernungsmatrix ergibt:

	1	2	3	4	5
1	∞	27	0	0	42
2	42	∞	0	70	17
3	25	0	∞	42	2
4	0	25	17	∞	0
5	65	15	0	28	∞

Lösung

$+15$ ∞ -15	$+15$ 27 -15	$+15$ 0 -15	$+15$ 0 -15	$+15$ 42 -15	
42 -15	∞ -15	0 -15	70 -15	$+15$ 17 -15	x
$+15$ 25 -15	0 $+15$ -15	$+15$ ∞ -15	$+15$ 42 -15	2 -15	
0 $+15$ -15	$+15$ 25 -15	$+15$ 12 -15	$+15$ ∞ -15	0 $+15$ -15	
65 -15	15 -15	0 -15	28 -15	∞ -15	x
					x

$+2$ ∞ -2	$+2$ 27 -2	$+2$ 15 -2	$+2$ 0 -2	$+2$ 42 -2	
27 -2	∞ -2	0 -2	55 -2	2 -2	x
25 -2	0 -2	∞ -2	42 -2	2 -2	x
0 $+2$ -2	$+2$ 25 -2	$+2$ 27 -2	$+2$ ∞ -2	0 $+2$ -2	
50 -2	0 -2	0 -2	13 -2	∞ -2	x
	x	x			

∞	29	17	0	42
25	∞	0	53	2
23	0	∞	40	0
0	27	29	∞	0
48	0	0	11	∞

Der Wert der Zielfunktion ist $Z = 328$. Diesen wert kann die exakte Lösung des Travelling Salesman Problem nicht unterschreiten. Die optimale Lösung des TSP ist $Z = 351$.

4.2.3.2.5 Das Travelling-Salesman-Problem

1. Fallbeispiel

Ein Molkereiunternehmen in A beliefert jeden Tag 4 Großmärkte in B, C, D und E mit Milchprodukten. Die Belieferung erfolgt mit einem LKW, der in A beladen wird und dann auf seiner Auslieferungsfahrt die 4 Großmärkte ansteuert, bevor er schließlich wieder nach A zurückfährt. Die Entfernungen in km zwischen dem Molkereiunternehmen und den Großmärkten sind in der folgenden Entfernungsmatrix wiedergegeben.

	A	B	C	D	E
A	∅	14	8	23	16
B	12	∅	14	19	9
C	10	14	∅		
D	22	18	19	∅	14
E	16	12	26	12	∅

Es soll für den LKW eine Auslieferungsrouten so bestimmt werden, daß vom Auslieferungspunkt A aus jeder Ort B, C, D und E einmal angefahren wird und der LKW dann wieder nach A zurückkehrt. Die zurückgelegte Strecke soll so kurz wie möglich sein.

2. Die Lösung des TSP mit Hilfe der begrenzten Enumeration

(1) Bestimmen der Ausgangslösung

Von jedem Ort aus wird zum nächstgelegenen Nachbarort gefahren. Ergibt sich dabei ein Kurzzyklus wird der zweitnächste Nachbarort gewählt.

Im Bsp. ergibt sich die Route: A - C - B - E - D - A mit der Streckenlänge 65 km.

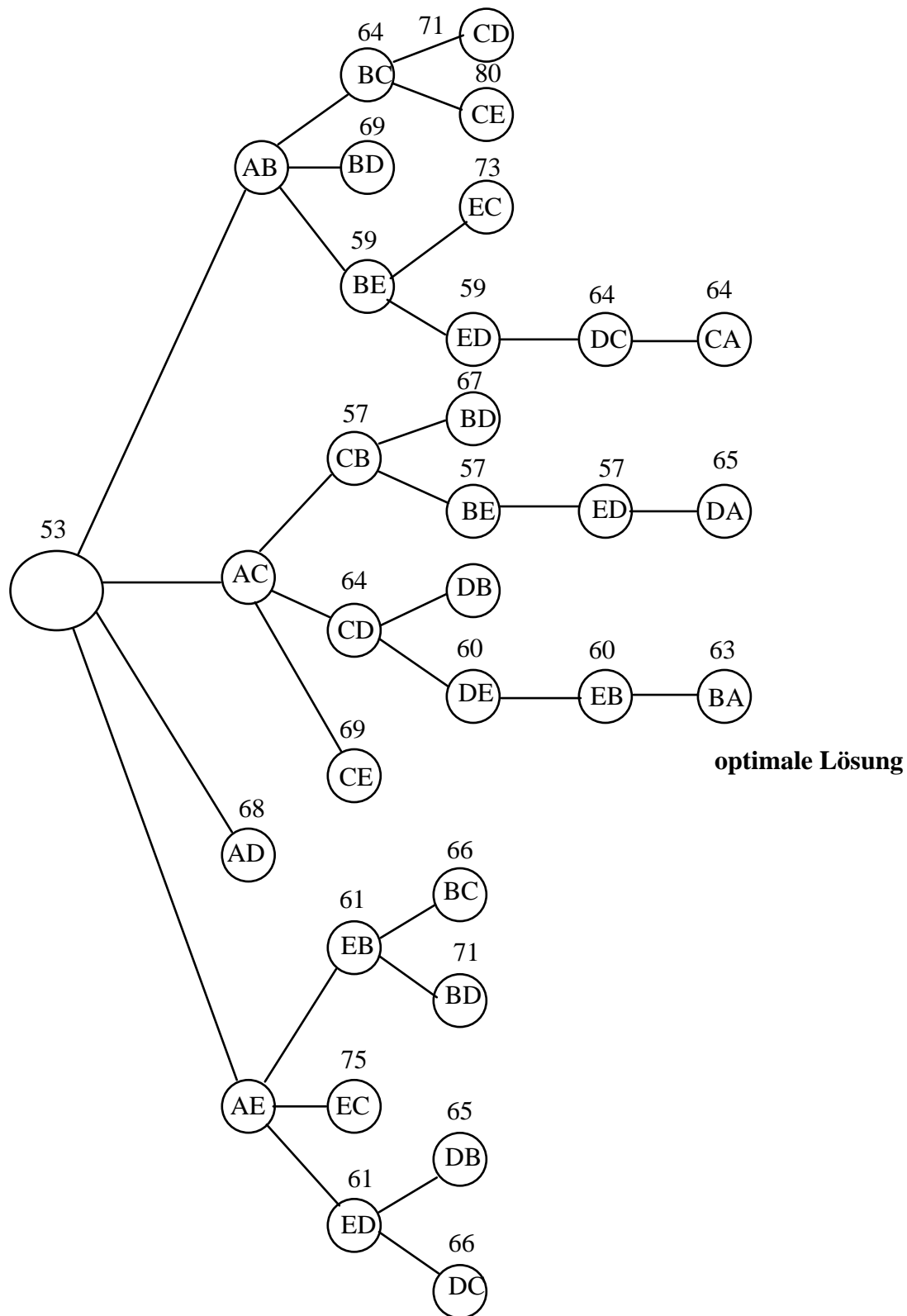
(2) Ermitteln der optimalen Lösung

- Bestimmen der reduzierten Entfernungsmatrix.

	A	B	C	D	E	Reduktion
A	∅	6	0	15	8	8
B	3	∅	5	10	0	9
C	0	4	∅	7	16	10
D	8	4	5	∅	0	14
E	4	0	14	0	∅	12
Reduktion						53

Im Beispiel ergibt sich die kürzestmögliche Ausgangslänge von 53 Km.

(3) Aufbau des Entscheidungsbaums der begrenzten Enumeration



4.2.3.2.6 Das Chinese Postman's Problem

Aufgabenstellung

Beim TSP wird ein Rundweg durch einen Graphen gesucht, der jeden Knoten einmal berührt. Beim Chinese Postman's Problem¹² ist ein „Rundweg“ gefragt, der jede Kante eines Graphen mindestens einmal enthält. Derartige Rundwege werden täglich von Briefträgern gegangen, die die Anlieger der durchschrittenen Straßen mit ihrer Post versorgen. Gesucht ist dabei jeweils der optimale, d.h. der kürzeste oder kostenminimale Rundweg.

Da alle Wege des Netzes einmal durchlaufen werden müssen, besteht die gesuchte Rundreise aus einer einfach durchzuführenden Aneinanderreihung aller Strecken, soweit zur Verbindung dieser Strecken keine zusätzlichen unproduktiven Wege erforderlich sind. Die Möglichkeit der einfachen Aneinanderreihung in zusammenhängenden ungerichteten Graphen besteht immer dann, wenn zu jedem Knoten eine gerade Anzahl von Kanten führt. In zusammenhängenden gerichteten Graphen ist die Möglichkeit dann gegeben, wenn für jeden Knoten die Zahl der zu ihm führenden Pfeile gleich der Zahl der zu ihm fortführenden Pfeile ist.

Das Königsberger Brückenproblem

Das Chinese Postman's Problem hat einen berühmten Vorgänger in dem Königsberger Brückenproblem¹³. In Königsberg bilden einige Brücken über den Fluß Pregel“ folgendes Schema:

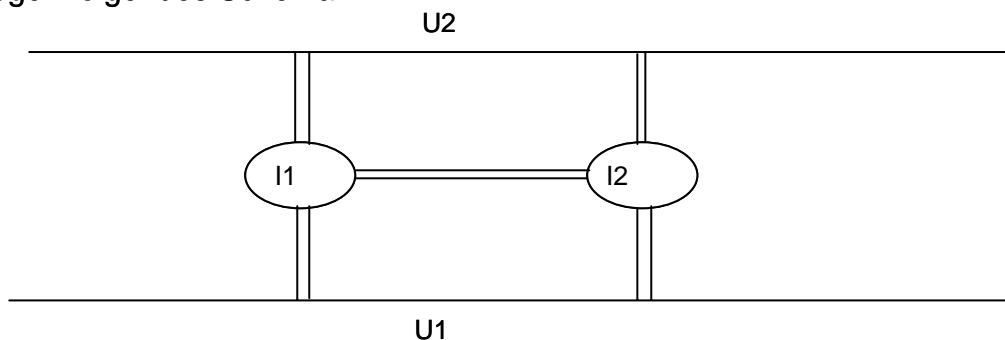


Abb.: Skizze zum Königsberger Brückenproblem (U1, U2: Ufer, I1, I2: Inseln)

Die Situation kann durch folgenden Graphen beschrieben werden:

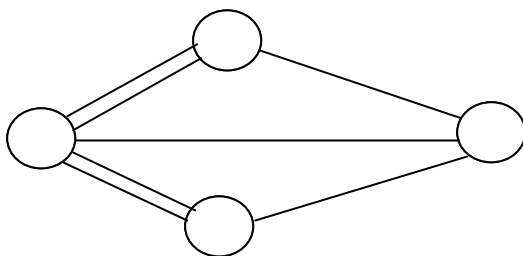


Abb.: Graph zum Königsberger Brückenproblem

¹² erstmals 1960 von dem Chinesen Mei-Ko Kwan untersucht

¹³ 1736 von Euler beschrieben und gelöst

Euler wies nach, daß dann und nur dann ein alle Kanten maximal benutzender Rundweg in einem zusammenhängenden ungerichteten Graphen besteht, wenn in alle Knoten eine gerade Anzahl von Kanten einmündet. Falls Start und Zielknoten verschieden sind, müssen in beide eine ungerade Anzahl Kanten einmünden, weil eine zusätzliche Kante vom Ziel zum Start den Rundweg vervollständigen muß. Diese Bedingungen sind notwendig: Falls in einen Knoten eine ungerade Zahl n von Kanten einmündet, dann wird man diesen Ort $\frac{n-1}{2}$ mal besuchen und ebenso oft

verlassen. Schließlich wird man den Ort noch einmal besuchen können und dann auf einer nicht benutzten Kanten verlassen. Es wäre ein zusätzlicher Weg erforderlich, der beim Chinese Postman's Problem als „unproduktiv“ bezeichnet werden soll.

Ein weiteres Problem vom Typ des Königsberger Brückenpüblems bildet das mit einem Strich zu zeichnende Häuschen (Haus des Nikolaus). Anfangsknoten und Endknoten haben eine ungerade Kantenzahl.

Soweit es alle Strecken umfassende Rundwege ohne unproduktiven Wege gibt, ist das Chinese Postman's Problem kein Optimierungsproblem. Erst wenn unproduktive Wege unumgänglich sind, liegt ein Optimierungsproblem vor.

Das Chinese Postman's Problem in ungerichteten Graphen

Während es Euler nur um die Existenz von Rundwegen ging, die alle Kanten benutzen, interessiert beim Chinese Postman's Problem insbesondere die Wahl der unproduktiven Wege, die eingefügt werden müssen, wenn ohne sie kein Rundweg durch alle Kanten existiert.

Wenn alle Kanten eine gerade Zahl einmündender Kanten aufweisen, ist eine Aneinanderreihung dieser Kanten erforderlich. Im allgemeinen Fall wird die Kantenzahl jedoch nicht für alle Knoten gerade sein. Unproduktive Wege verbinden die Knoten mit ungerader Kantenzahl. Ist die Anzahl der Knoten m , dann sind zur Verbindung der m Knoten $\frac{m}{2}$ unproduktive Wege erforderlich. Sie bestehen aus neu

einzufügenden Kanten bzw. Kantenzüge. Soweit man an die bestehenden Kanten eines Graphen gebunden ist, sind die unproduktiven Wege als Mehrfachbenutzung der bestehenden Kanten zu verstehen. In den entsprechenden Abbildungen werden sie meistens gestrichelt dargestellt.

Gesucht sind diejenigen unproduktiven Wege mit der geringsten Kantenlängensumme.

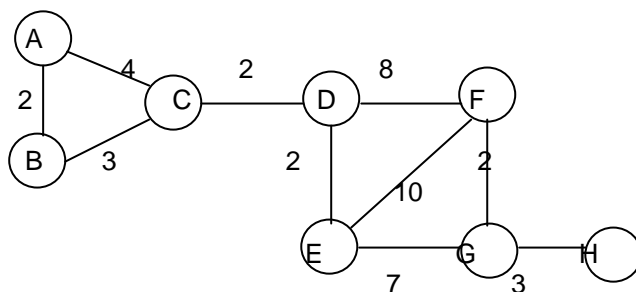


Abb.

Kürzester Weg von A durch alle Kanten zurück zum Ort A. Nur A und B haben eine gerade Zahl einmündender Kanten. Unter vielen gleich optimale Wegen lautet einer:

A – B – C – D – E – F – G – E – G – H – F – D – C – A

Die unproduktiven wege enthalten die Kanten C – D, E – G, F – G und G – H mit einer Länge von 14 Einheiten. Der gesamte Rundweg ist 57 Einheiten lang.

Das Chinese Postman's Problem in gerichteten Graphen

Soweit der Graph zusammenhängend ist, ist es eine notwendige und hinreichende Bedingung für die Existenz eines Rundwegs, der jede Kante mindestens einmal enthält, daß für jeden Knoten die Zahl der hinführenden Kanten gleich der fortführenden Kanten ist.

Wenn ein Knoten i genau a_i mehr zu ihm hinführende als fortführende Pfeile aufweist, so sind a_i von ihm fortführende Pfeile als unproduktive Wege zusätzlich einzufügen. Hat dagegen ein Knoten j genau b_j mehr von ihm fortführende als zu ihm hinführende Pfeile, so müssen b_j zu ihm hinführende Pfeile als unproduktive Wege zusätzlich eingeplant werden. Jeder unproduktive Weg geht von einem Knoten mit mehr hinführenden Pfeilen zu einem Knoten mit mehr fortführenden Pfeilen. Die Länge des kürzesten wegs von einem Knoten i zu einem Knoten j sei mit d_{ij} bezeichnet. Gesucht sind diejenigen unproduktiven Wege, deren Kantenlängensumme minimal ist.

Mit $1 = 1, 2, \dots, m$ als Index für die Wege mit mehr hinführenden Pfeilen und $j = 1, 2, \dots, n$ als Index für die Wege mit mehr fortführenden Pfeilen und x_{ij} als Zahl der unproduktiven Wege vom Knoten i zum Knoten j läßt sich das Problem der optimalen unproduktiven Wege folgendermaßen definieren:

$$Z_{\min} = \sum_{i=1}^m \sum_{j=1}^n d_{ij} \cdot x_{ij}$$

unter den Restriktionen

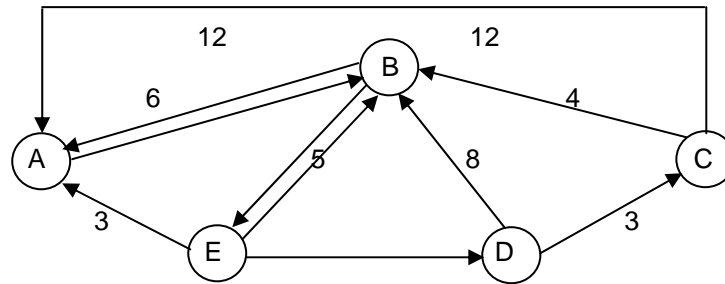
$$\sum_{i=1}^m x_{ij} = b_j \quad \text{für } j = 1, 2, \dots, n$$

$$\sum_{j=1}^n x_{ij} = a_i \quad \text{für } i = 1, 2, \dots, m$$

$$x_{ij} \geq 0$$

Dieses Problem ist identisch mit dem Transportproblem der linearen Planungsrechnung.

Bsp.: Zur Bestimmung der unproduktiven Wege in dem folgenden Graphen



kann eine Lösung des folgenden Transportproblem herangezogen werden:

Von \ Bis	C	D	E	a_i
A	18	15	11	3
B	12	9	5	2
b_j	1	2	2	5

Es gibt mehrere gleichwertige Lösungen, u.a. $x_{AC} = 1$ und $x_{AD} = x_{BE} = 2$

In den dadurch gekennzeichneten unproduktiven Wegen sind die Kanten B – E fünfmal, A – B und E – D dreimal sowie D – C enthalten. Die Länge der unproduktiven Wege beträgt 58 Einheiten. Eine der vielen möglichen Rundreisen mit der Länge von 123 Einheiten lautet:

A – B – A – D – E – A – D – E – B – E – D – A – B – E – D – B – E – D – C – B – E – D – C – A

Zusammenstellung der Rechenschritte:

1. Schritt: Bestimmung der Knoten mit Überschuß an hinführenden bzw. mit Überschuß an fortführenden Kanten. Hat kein Knoten einen derartigen Überschuß gehe zum 4. Schritt
2. Schritt: Berechnung (und Speicherung) der kürzesten Wege zwischen den Knoten mit Überschuß an hinführenden Kanten und den Knoten mit Überschuß an fortführenden Kanten. Zusammenstellung der Entfernungen als Transportproblematrix.
3. Schritt: Lösen des Transportproblems und Einfügen der unproduktiven Kanten in den Graphen.
4. Schritt: Ermitteln des Rundwegs durch alle Kanten

5. Warteschlangenmodelle

5.1 Stochastische Prozesse

5.1.1 Einführende Beispiele

Vielfach ist ein Zusammenhang zwischen Ursache und Wirkung nicht eindeutig nachweisbar. Vorgänge sind in der Regel determiniert. Es gibt aber auch sich wiederholende Vorgänge, die unter gleichen Bedingungen jeweils unterschiedliche Resultate liefern. Solche Vorgänge haben stochastischen Charakter. Es kann nur ausgedrückt werden, mit welchen Wahrscheinlichkeiten ein bestimmtes Ereignis auftritt.

Brownsche Molekularbewegung

(Bewegung kleiner Teilchen in einer ruhenden, homogenen Flüssigkeit)

Die Lage eines Teilchens in einer Flüssigkeit kann durch die Koordinaten $X(t)$, $Y(t)$ und $Z(t)$ beschrieben werden. Zum Zeitpunkt $t_0 = 0$ befindet sich das Teilchen im Koordinatenursprung. In Folge der dauernden Zusammenstöße ist die Lage des Teilchens zu einem beliebigen Zeitpunkt nicht vorher bestimmt, sondern zufälliger Art.

$X(t)$, $Y(t)$ und $Z(t)$ sind Zufallsgrößen. Die Menge der von $t \geq 0$ abhängigen Größen $\{X(t), Y(t), Z(t)\}$ bezeichnet man als **stochastischen Prozeß**.

Warteschlangen und Bedienung

Zu zufälligen Zeitpunkten treffen "Kunden" vor einem "Schalter" ein und fordern eine "Bedienung". Diese Bedienung erfordert eine bestimmte zufällige Abfertigungszeit. Stochastischer Prozeß ist hier die Anzahl der zur Zeit t wartenden Kunden.

Die Warteschlangen-Theorie stellt Methoden bereit, um wichtige Kenngrößen (z.B.: mittlere Wartezeit eines Kunden, Auslastungsgrad eines Wartesystems) zu ermitteln.

Lagerhaltung

Ein bestimmter Artikel wird in einem Lager gehalten. Am Ende der Periode t (Tag, Woche, ...) wird der Lagerbestand $X(t)$ festgestellt. Dieser Lagerbestand hängt vom Bedarf $Y(t)$ während der Periode und von den Bestellregeln zur Steuerung der Lagerzufuhr ab.

Zuverlässigkeitsprobleme

Ein Gerätetyp besitzt eine zufällige Lebensdauer. Nach dem Ausfall eines Gerätes wird dieses sofort durch ein neues Gerät ersetzt oder nach einer (zufälligen) Reperaturzeit wieder instand gesetzt. Die Anzahl der Erneuerungen bis zu einem Zeitpunkt t bildet einen stochastischen Prozeß.

ökonomische Zeitreihen

Volkswirtschaftliche Größen $X(t)$, (z.B. Bruttosozialprodukt, Arbeitslosenziffer, Lebenshaltungsindex) werden periodisch erhoben. Die in aufeinanderfolgenden Perioden festgestellten Daten $X(1), X(2), \dots, X(t)$ bilden eine Zeitreihe. Auch solche Zeitreihen können als stochastischer Prozeß aufgefaßt werden.

Zusammenfassung

In den vorliegenden Beispielen gibt es eine Menge von Zufallsgrößen $X(t)$, die von einem nichtzufälligen Parameter t einer bestimmten Parametermenge T (Indexmenge) abhängen.

5.1.2 Wahrscheinlichkeitstheoretische Grundlagen**5.1.2.1 Zufallsvariable**

Eine Zufallsvariable ist eine Funktion, die das Ergebnis eines zufallsbedingten Vorgangs (Zufallsexperiments) ausdrückt. So kann bspw. das Ergebnis des Zufallsexperiments "Werfen eines einzelnen Würfels" durch das Zufallsexperiment "Geworfene Augenzahl", die die möglichen Werte 1, 2, ..., 6 annehmen kann, beschrieben werden. Eine Zufallsvariable ist auch die Zeit zwischen der Ankunft aufeinanderfolgender Aufträge bei einem Rechensystem oder die Antwortzeit oder der Durchsatz eines Rechensystems. Zufallsvariable können kontinuierlich oder diskrete Werte annehmen (kontinuierliche oder diskrete Zufallsvariable).

1. Diskrete Zufallsvariable

Sie wird beschrieben durch die möglichen Werte, die sie annehmen kann (im allg. ganze Zahlen) und die Wahrscheinlichkeiten für diese Werte, z.B.:

$$p_k = P(X = k) \text{ für } k = 0, 1, 2, \dots$$

$$\text{Es muß sein } P(X = k) \geq 0 \text{ und } \sum_k P(X = k) = 1$$

Für das Zufallsexperiment "Werfen eines Würfels" ergibt sich bspw. folgende Verteilung: $P(X = k) = \frac{1}{6}$ für $k = 1, 2, \dots, 6$

Andere diskrete Verteilungen sind

- Bernoulli - Verteilung

Hier steht ein Zufallsexperiment im Mittelpunkt, das nur 2 mögliche Ausgänge haben kann, z.B. das Werfen einer Münze ($k = 1, 2$). Für die Wahrscheinlichkeitsverteilung der dazugehörigen Zufallsvariablen X gilt dann:

$$P(X = 1) = p$$

$$P(X = 2) = (1 - p) \text{ mit } 0 < p < 1$$

- Geometrische Verteilung

Das Zufallsexperiment mit 2 Ausgängen wird mehrfach durchgeführt, wobei die Zufallsvariable X jetzt die Anzahl der Versuche angeben soll, bis zum ersten Mal die Eins erscheint (inklusive dieses Versuchs). Für die Wahrscheinlichkeitsverteilung von X gilt:

$$P(X = k) = p \cdot (1-p)^{k-1} \quad k = 1, 2, \dots$$

- Binomialverteilung (B(n,p))

Falls der Zufallsvorgang n -mal wiederholt wird, wobei es zwischen den einzelnen Wiederholungen keine wechselseitigen Beeinflussungen (Unabhängigkeit) gibt, dann kann die i -te Wiederholung (der Ereignisse) durch die Zufallsvariable

$$X_i = \begin{cases} 1 \\ 0 \end{cases}$$

beschrieben werden. Die Zufallsvariable $X = \sum_{i=1}^n X_i$ mißt die Anzahl der Experimente, bei denen das Ereignis eintritt. Die Wahrscheinlichkeitsverteilung der Binomialverteilung läßt sich folgendermaßen ableiten:

a) Alle X_i ($i=1, \dots, n$) sind voneinander stochastisch unabhängig. Für jedes X_i gilt:

$$P(X_i = 1) = p \text{ bzw. } P(X_i = 0) = (1-p)$$

b) Für $x_i \in \{0,1\}$ und $k = \sum_{i=1}^n x_i$ gilt (wegen der Unabhängigkeit der X_i)

$$P(X_1 = x_1, \dots, X_n = x_n) = p^k \cdot (1-p)^{n-k}$$

c) Jeder Vektor (x_1, \dots, x_n) entspricht einer speziellen Anordnung von " k " Einsen und $(n-k)$ Nullen. Es gibt $\frac{n!}{k! \cdot (n-k)!}$ derartige Anordnungsmöglichkeiten.

d) Aus b) und c) ergibt sich die Wahrscheinlichkeit $P(X=k)$ dafür, daß bei n -maliger Wiederholung des Bernoulli-Experiments (n Stichproben) genau k -mal auftritt:

$$P(X = k) = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

Es gibt $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k}$ Möglichkeiten, k Ereignisse aus n möglichen Ereignissen auszuwählen¹⁴.

- Poisson-Verteilung

Sind beliebige Ereignisse voneinander unabhängig und gibt die Zufallsvariable X die Anzahl der Ereignisse (Häufigkeit) je Zeiteinheit an, dann ist X poissonverteilt:

$$P(X = k) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t} \quad k = 0, 1, 2, \dots$$

λ ist die Rate, mit der Ereignisse eintreten, d.h. im Mittel treten je Zeitspanne t λ Ereignisse ein.

¹⁴ Es handelt sich um Binomialkoeffizienten, die auch im Pascal'schen Dreieck zu finden sind.

Es gibt eine Vielzahl von zufallsbedingten Vorgängen, bei denen die Einzelereignisse unabhängig sind und deren Anzahl damit poissonverteilt ist, z.B.:

- Zahl der Anrufe in einem Fernsprechnet
- Zahl der ankommenden Aufträge in einem Rechensystem

Die Poisson-Verteilung bietet eine Approximationsmöglichkeit für die $B(n,p)$ -Verteilung bei kleinem "p" und großem "n".

Aus der Wahrscheinlichkeitsverteilung einer diskreten Verteilung können weitere wichtige Parameter abgeleitet werden:

- Mittelwert oder Erwartungswert: $E(X) = \sum_k k \cdot P(X = k)$

- n-tes Moment: $E(X^n) = \sum_k k^n \cdot P(X = k)$

(d.h.: Das n-te Moment ist der Erwartungswert der n-ten Potenz von X. Das erste Moment von X ist der Mittelwert von X).

2. Kontinuierliche Zufallsvariable

Sie kann alle Werte in einem Intervall annehmen und wird beschrieben durch die Verteilungsfunktion:

$$F_X(x) = P(X \leq x)$$

Statt der Verteilungsfunktion kann auch die Dichtefunktion $f_X(x)$ verwendet werden:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

- Mittelwert oder Erwartungswert: $E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$

- n-tes Moment: $E(X^n) = \int_{-\infty}^{+\infty} x^n \cdot f_X(x) dx$

Wichtige kontinuierliche Verteilungen sind:

- die Exponentialverteilung

$$F_X(x) = 1 - e^{(-x/\lambda)}, \quad x \geq 0$$

wobei $\bar{X} = \begin{cases} 1/\lambda & \text{für einen Ankunftsprozeß} \\ 1/\mu & \text{für einen Bedienprozeß} \end{cases}$

$\frac{1}{\lambda}$ für einen Ankunftsprozeß

Somit gilt z.B. für einen Ankunftsprozeß mit exponentiell verteilten Zwischenankunftszeiten:

Dichtefunktion: $f_X(x) = e^{-\lambda x}$, $x \geq 0$

Mittelwert: $X = \frac{1}{\lambda}$

Die Exponentialverteilung ist vollständig durch ihren Mittelwert bestimmt.

Falls sich der stochastische Prozeß der Ankunft von Einheiten in einem Wartesystem durch eine Poisson-Verteilung beschreiben läßt, dann sind die Zwischenankunftszeiten der Kunden exponentialverteilt.

5.1.2.2 Mehrdimensionale Zufallsvariable, Abhängigkeiten zwischen mehreren Zufallsvariablen

In manchen Fällen werden durch den Ausgang ein und desselben Zufallsexperiments die Werte mehrerer Zufallsvariablen bestimmt, wobei sich die Werte gegenseitig beeinflussen können.

Sind X_1, X_2, \dots, X_n diskrete Zufallsvariable, dann ist $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ die Verbundwahrscheinlichkeit dafür, daß $X_1 = x_1$ und $X_2 = x_2$ und $X_n = x_n$ ist.

Im kontinuierlichen Fall bezeichnet die Verbundverteilung $P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$ die Wahrscheinlichkeit dafür, daß $X_1 \leq x_1$ und $X_2 \leq x_2$ und \dots $X_n \leq x_n$.

1. Unabhängigkeit

Gilt im kontinuierlichen Fall

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdot \dots \cdot P(X_n \leq x_n)$$

bzw. im diskreten Fall

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \dots \cdot P(X_n = x_n)$$

dann werden die Zufallsvariablen X_1, X_2, \dots, X_n (statistisch) **unabhängig** genannt.

Im anderen Fall heißen sie (statistisch) **abhängig**.

2. Bedingte Wahrscheinlichkeit

Im diskreten Fall gibt die **bedingte** Wahrscheinlichkeit $P(X_1 = x_1 \mid X_2 = x_2, \dots, X_n = x_n)$ die Wahrscheinlichkeit für $X_1 = x_1$ an, wenn die Bedingungen $X_2 = x_2$ und \dots und $X_n = x_n$ erfüllt sind. Es gilt:

$$P(X_1 = x_1 \mid X_2 = x_2, \dots, X_n = x_n) = \frac{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}{P(X_2 = x_2, \dots, X_n = x_n)}$$

Entsprechend gilt für kontinuierliche Zufallsvariable:

$$P(X_1 \leq x_1 \mid X_2 \leq x_2, \dots, X_n \leq x_n) = \frac{P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)}{P(X_2 \leq x_2, \dots, X_n \leq x_n)}$$

5.1.2.3 Stochastische Prozesse

Prozesse, wie sie durch die einführenden Beispiele (vgl. 5.1.1) charakterisiert wurden, werden mit Hilfe von Zufallsvariablen beschrieben, z.B. im Zeitpunkt t durch die Zufallsvariable $X(t)$.

Zu einem Zeitpunkt t' ist der Prozeß durch eine andere Zufallsvariable $X(t')$ charakterisiert. Ein geeignetes Beschreibungsverfahren ist demnach die Familie von Zufallsvariablen $\{X(t), t \in T\}$, bei der sich die einzelnen Zufallsgrößen durch einen Zeitparameter t voneinander unterscheiden. So eine Familie heißt **stochastischer Prozeß**.

Alle möglichen Werte, die die Zufallsvariablen $X(t), t \in T$ annehmen können, bilden den Zustandsraum Z des stochastischen Prozesses. Jeder einzelne dieser Werte $X(t)$ wird Zustand des Prozesses zur Zeit t genannt. Setzt sich die Menge Z aus einem oder mehreren Intervallen der reellen Zahlen zusammen, so heißt der Prozeß **zustandskontinuierlich**. Ist der Zustandsraum Z dagegen endlich oder abzählbar, dann spricht man von einem **zustandsdiskreten** Prozeß. Zustandsdiskrete Prozesse werden auch als Ketten bezeichnet, wobei der Zustandsraum in diesem Fall meistens durch die Menge $\{0, 1, 2, \dots\}$ angegeben wird.

In ähnlicher Weise können stochastische Prozesse auch bzgl. der Indexmenge T klassifiziert werden. Beinhaltet T nämlich ein beliebiges Intervall der Zeit, so wird der Prozeß **zeitkontinuierlich** genannt. Ist T dagegen endlich oder abzählbar, dann spricht man von einem **zeitdiskreten** Prozeß. Zeitdiskrete Prozesse werden auch als Folgen bezeichnet.

5.1.2.4 Markov-Prozesse

Ein stochastischer Prozeß gehört dann zur Klasse der Markov-Prozesse, wenn der zukünftige Verlauf des Prozesses nur vom augenblicklichen Zustand $X(t)$ zur Zeit t abhängt und nicht von der Vorgeschichte (**Markov-Eigenschaft**).

Von zentraler Bedeutung sind vor allem zeitkontinuierliche und zustandsdiskrete Markov-Prozesse.

Markov-Ketten

Die formale Definition für zeitkontinuierliche Markov-Ketten lautet:

Ein stochastischer Prozeß $\{X(t), t \geq 0\}$ bildet eine zeitkontinuierliche Markov-Kette, wenn für alle natürlichen Zahlen $x_k \in Z$ und alle $\{t_1, t_2, \dots, t_{n+1}\}$ mit $t_1 < t_2 < \dots < t_{n+1}$ gilt:

$$P(X(t_{n+1}) = x_{n+1} \mid X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_n) = x_n) = P(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n)$$

Diese Gleichung beschreibt die **Markov-Eigenschaft**. Sie besagt, daß die gesamte relevante Information über die Vergangenheit eines Markov-Prozesses vollständig im

augenblicklichen Zustand $x(t_n)$ enthalten ist. Auch hängt es nur vom augenblicklichen Zustand $x(t_n)$ ab, in welchen Zustand $x(t_{n+1})$ der Prozeß als nächstes übergehen wird. Dies hat zur Folge, daß die Zeit, wie lange sich der Prozeß noch im Zustand $x(t_n)$ aufhalten wird, vollkommen unabhängig von der Zeit ist, wie lange sich bereits der Prozeß vor dem Beobachtungszeitpunkt t_n in diesem Zustand befunden hat. Die Verteilung der Zustandszeiten hat demnach die Eigenschaft der Gedächtnislosigkeit. Die **Exponentialverteilung** besitzt als einzige kontinuierliche Verteilung diese Eigenschaft. Zustandszeiten einer zeitkontinuierlichen Markov-Kette müssen daher exponentiell verteilt sein.

$$P(T \leq s + t | T > s) = 1 - e^{-(t/T)} = P(T \leq t)$$

Bsp.: Hat man eine Bedienstation mit exponentiell verteilten Zwischenankunftszeiten T bereits s Zeiteinheiten beobachtet und während dieser Zeit keinen neuen Auftrag festgestellt, dann ist die Wahrscheinlichkeit noch weitere t Zeiteinheiten bis zur Ankunft des nächsten Auftrags warten zu müssen, vollkommen unabhängig von der bislang gewarteten Zeit s . Die Station erinnert sich nicht daran, bereits s Einheiten gewartet zu haben.

Exponentialverteilung liegt demnach immer genau dann vor, wenn die einzelnen Ereignisse eines (Ankunft- oder Bedien-) Prozesses voneinander unabhängig sind.

$$P(X(t_{n+1}) = x_{n+1} | X(t_n) = x_n)$$

bezeichnet die **Übergangswahrscheinlichkeit** der Markov-Kette. Dafür kann man auch schreiben:

$$p_{ij}(s, t) = P(X(t) = j | X(s) = i), \quad t > s.$$

$p_{ij}(s, t)$ ist die bedingte Übergangswahrscheinlichkeit dafür, daß der Prozeß zur Zeit t im Zustand j befindet, wenn er zur Zeit s im Zustand i ist.

Hängen die Übergangswahrscheinlichkeiten nicht vom Zeitpunkt t ab, sondern nur von der Zeitdifferenz $\tau = t - s$, dann wird die Markov-Kette **homogen** genannt.

$$p_{ij}(\tau) = P(X(s + \tau) = j | X(s) = i).$$

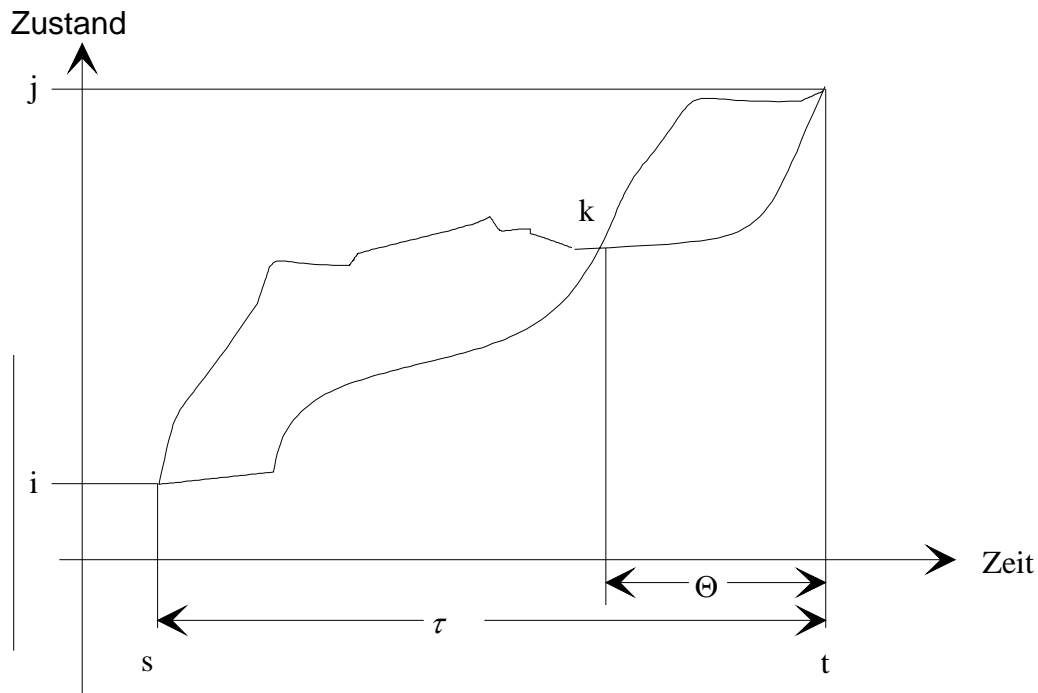
Es gilt

$$p_{ij}(\tau) \geq 0, \quad i, j \in Z, \quad \sum_{k \in Z} p_{ij}(\tau) = 1$$

und die **Chapman-Kolmogorov-Gleichung**

$$p_{ij}(\tau) = \sum_{k \in Z} p_{ik}(\tau - \Theta) \cdot p_{kj}(\Theta) \quad i, j \in Z$$

Geht ein Prozeß in der Zeit τ vom Zustand i in einen anderen Zustand j über, so muß er sich nach Ablauf des Zeitintervalls $\tau - \Theta$ in irgendeinem Zustand k befinden. Ausgehend von diesem Zwischenzustand gelangt er dann in der Zeit Θ zum Endzustand j .



Da sich der Prozeß nach Ablauf des Zeitintervalls $(\tau - \Theta)$ in irgendeinem von vielen möglichen Zwischenzuständen befinden kann, erhält man durch Summenbildung über alle möglichen Zwischenzustände k alle Prozeßverläufe, die in der Zeit τ von i nach j führen.

Das Ziel der Untersuchung von Markov-Prozessen ist die Bestimmung der Zustandswahrscheinlichkeiten $p_j(t) = P(X(t) = j)$. Diese geben an, mit welcher Wahrscheinlichkeit sich der Prozeß zur Zeit t im Zustand j befindet. Mit den allgemeinen Anfangsbedingungen $p_i(0) = P(X(0) = i)$ gilt:

$$p_j(t) = \sum_{i \in Z} p_{ij}(t) \cdot p_i(0)$$

Die Markov-Kette ist somit durch die Wahrscheinlichkeit $p_i(0)$ für die Anfangszustände und durch die Übergangswahrscheinlichkeiten vollständig definiert.

Bsp.: "Irrfahrt eines Teilchen"

Ein Teilchen verändert seine Lage ($1 \leq x \leq l$) längs einer Geraden zu den Zeitpunkten $t = 1, 2, 3, \dots$ nach folgender Vorschrift:

1. Von den Punkten $x = 2, \dots, l - 1$ wird es in der nachfolgenden Zeiteinheit um eine Einheit mit der Wahrscheinlichkeit p in positiver und mit der Wahrscheinlichkeit $q = 1 - p$ in negativer Richtung bewegt.
2. An den Punkten $x = 1$ und $x = l$ wird das Teilchen absorbiert. Es verbleibt mit der Wahrscheinlichkeit 1 in der nachfolgenden Zeiteinheit dort.

Zur Zeit $t = 0$ befindet sich das Teilchen mit der Wahrscheinlichkeit $p_i(0)$ bei $x = i$ ($i = 1, 2, 3, \dots, l$).

- 1) Berechne die Wahrscheinlichkeit, das Teilchen zur Zeit $t = 1$ bei $x = j$ anzutreffen
- 2) Berechne die Wahrscheinlichkeit, das Teilchen zur Zeit $t = n$ ($n > 1$) bei $x = j$ anzutreffen.

Zur Berechnung der Wahrscheinlichkeiten sind die Übergangswahrscheinlichkeiten zu bestimmen:

Für $i = j \in \{2, 3, \dots, l-1\}$ ist $p_{ij} = 0$

Für alle i, j mit $|i - j| \geq 2$ gilt ebenfalls $p_{ij} = 0$.

(Das Teilchen wird mit der Wahrscheinlichkeit 0 um 2 oder mehr Einheiten fortbewegt).

Für $x = 1$ bzw. $x = l$ ist $p_{11} = 1$ bzw. $p_{ll} = 1$

in den Zustand j vom Zustand i	1	2	3	4	...	l-2	l-1	l
1	1	0	0	0	...	0	0	0
2	q	0	p	0	...	0	0	0
3	0	q	0	p	...	0	0	0
4	0	0	q	0	...	0	0	0
...					...			
l-1	0	0	0	0	...	q	0	p
l	0	0	0	0	...	0	0	1

1) Zu bestimmen ist die Wahrscheinlichkeit $P(X(1) = j)$!

$$p_j = \sum_{i \in \mathbb{Z}} p_{ij}(t) \cdot p_i(0)$$

Setzt man bspw. $l = 5$ und $p = q = 1/2$ (, das Teilchen befindet sich zur Zeit $t = 0$ mit der Wahrscheinlichkeit $1/2$ im Punkt 2 und mit der Wahrscheinlichkeit $1/2$ im Punkt 4), dann ergibt sich:

$$p_1(0) = 0, p_2(0) = 1/2, p_3(0) = 0, p_4(0) = 1/2, p_5(0) = 0$$

Die Wahrscheinlichkeit, das Teilchen an den Stellen $j = 1, 2, 3, 4, 5$ anzutreffen ist $1/4, 0, 1/2, 0, 1/4$.

2) Zu bestimmen ist die Wahrscheinlichkeit $P(X(n) = j)$

$$p_j = \sum_{i \in \mathbb{Z}} p_{ij}(t) \cdot p_i(0) \text{ ergibt sich zu } p_j(n) = \sum_{i \in \mathbb{Z}} p_{ij}(n) \cdot p_i(0)$$

$$p_{ij}(n) = ?$$

$p_{ij}(n)$ lässt sich mit Hilfe der Chapman-Kolmogorov-Gleichung bestimmen.

$$p_{ij}(n) = \sum_k p_{ik} \cdot p_{kj}(n-1)$$

Aus den p_{ij} lassen sich dann nacheinander alle $p_{ij}(2)$, $p_{ij}(3)$ und schließlich auch $p_{ij}(n)$ bestimmen.

Es gilt demnach für ein ansonsten beliebiges, ganzzahliges k ($1 \leq k \leq n-1$):

$$p_{ij}(n) = \sum_k p_{ik} \cdot p_{kj}(n-1)$$

$p_{ij}(2)$ ist dann bspw. auf die Übergangsmatrix bezogen und führt zu: $p_{ij} \cdot p_{ij} = (p_{ij})^2$

Allgemein gilt für $n \geq 1$: $p_{ij}(n) = (p_{ij})^n$

Im vorstehenden Beispiel ("Irrfahrt eines Teilchen") ergibt sich dann für $n = 2$, $p = q = 1/2$, und $s = 5$

$$p_{ij}(2) = p_{ij} \cdot p_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/4 & 0 & 1/4 & 0 \\ 1/4 & 0 & 1/2 & 0 & 1/4 \\ 0 & 1/4 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

So erhält man für die Wahrscheinlichkeit, das Teilchen zur Zeit $t = 2$ bei $j = 2$ anzutreffen

$$P(X(2) = 2) \geq \sum p_i(0) \cdot p_{i2}(2) = 0 + 1/2 \cdot 1/4 + 0 + 1/2 \cdot 1/4 + 0 = 1/4$$

Zustands-, Übergangsgraph, Übergangsintensität

Darunter versteht man einen gerichteten Graphen, dessen Knoten den Zuständen entsprechen. Die Kanten werden mit p_{ij} bewertet.

Bsp.: Ein Benutzerprogramm kann in einer Rechenanlage einen der folgenden Zustände annehmen

INITIAL	(1) : initialisiert
AKTIV	(2) : aktiv (rechnend)
BLOCK	(3) : blockiert
BEREIT	(4) : Warten auf Prozessor-Zuteilung
AUS.BLOCK	(5) : auf dem Hintergrundspeicher ausgelagert
AUS.BEREIT	(6) : ausgelagert, auf die Wiedereinlagerung wartend
FERTIG	(7)

Die Zustandsübergänge bilden einen Markov-Prozeß .

Zustandsgraph

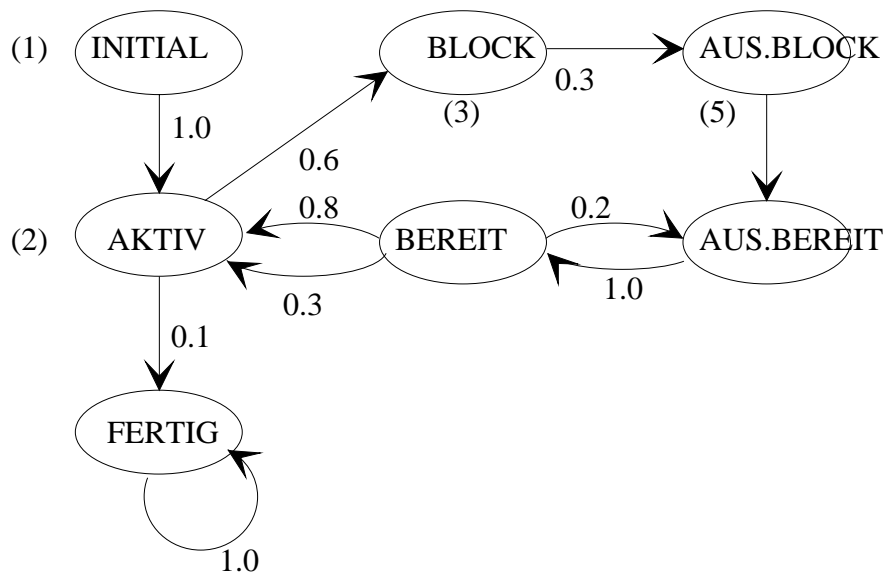


Abb.: Zustandsgraph

Übergangsmatrix:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	1.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.60	0.30	0.00	0.00	0.10
3	0.00	0.00	0.00	0.70	0.30	0.00	0.00
4	0.00	0.80	0.00	0.00	0.00	0.20	0.00
5	0.00	0.00	0.00	0.00	0.00	1.00	0.00
6	0.00	0.00	0.00	1.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Diese Matrix beschreibt das Verhalten eines (zufälligen) Übergangs in einem Schritt.
Bei 2 aufeinanderfolgenden Schritten gilt

$$p_{ij}^{(2)} = \sum_k p_{kj} \cdot p_{ki} = (p_{ij})^2$$

Übergangsmatrix nach 2 Schritten

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	0.00	0.60	0.30	0.00	0.00	0.10
2	0.00	0.24	0.00	0.42	0.18	0.06	0.10
3	0.00	0.56	0.00	0.00	0.00	0.44	0.00
4	0.00	0.00	0.48	0.44	0.00	0.00	0.08
5	0.00	0.00	0.00	1.00	0.00	0.00	0.00
6	0.00	0.80	0.00	0.00	0.00	0.20	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Die Übergangsmatrix nach 50 Schritten ergibt sich dann zu:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	0.07	0.04	0.08	0.01	0.03	0.77
2	0.00	0.06	0.04	0.08	0.01	0.03	0.78
3	0.00	0.07	0.04	0.08	0.01	0.03	0.76
4	0.00	0.07	0.04	0.08	0.01	0.03	0.77
5	0.00	0.07	0.04	0.09	0.01	0.03	0.76
6	0.00	0.07	0.04	0.08	0.01	0.03	0.76

7	0.00	0.00	0.00	0.00	0.00	0.00	1.00
---	------	------	------	------	------	------	------

Eine weitere wichtige Matrix kann aus der Matrix $p_{ij}^{(n)}$ abgeleitet werden: Die erwartete Anzahl der Besuche im Zustand j, falls im Zustand i gestartet wurde. Sie ergibt sich zu:

$$\sum_{n=0}^{\infty} p_{ij}^{(n)}$$

Für das vorliegende Beispiel ermittelt sich die Übergangswahrscheinlichkeit für 500 Schritte zu

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	0.00	0.00	0.00	0.00	0.00	1.00
2	0.00	0.00	0.00	0.00	0.00	0.00	1.00
3	0.00	0.00	0.00	0.00	0.00	0.00	1.00
4	0.00	0.00	0.00	0.00	0.00	0.00	1.00
5	0.00	0.00	0.00	0.00	0.00	0.00	1.00
6	0.00	0.00	0.00	0.00	0.00	0.00	1.00
7	0.00	0.00	0.00	0.00	0.00	0.00	1.00

und die "erwartete Anzahl Besuche nach 500 Schritten" ist:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	10.00	6.00	11.25	1.80	4.05	466.90
2	0.00	9.00	6.00	11.25	1.80	4.05	467.90
3	0.00	10.00	6.00	12.50	2.10	4.60	464.80
4	0.00	10.00	6.00	11.50	1.80	4.30	466.40
5	0.00	10.00	6.00	12.50	1.80	5.30	464.40
6	0.00	10.00	6.00	12.50	1.80	4.30	465.40
7	0.00	0.00	0.00	0.00	0.00	0.00	500.00

Im Grenzfall $n \rightarrow \infty$ ergibt sich:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.00	10.00	6.00	11.25	1.80	4.05	∞
2	0.00	9.00	6.00	11.25	1.80	4.05	∞
3	0.00	10.00	6.00	12.50	2.10	4.60	∞
4	0.00	10.00	6.00	11.50	1.80	4.30	∞
5	0.00	10.00	6.00	12.50	1.80	5.30	∞
6	0.00	10.00	6.00	12.50	1.80	4.30	∞
7	0.00	0.00	0.00	0.00	0.00	0.00	∞

Mit Hilfe von Zustandsgraphen können demnach Übergangsmatrizen beschrieben werden.

Aufgabe: Ein Sessellift kann alle 30s zwei Personen befördern. Die Bedienrate ist demnach $M = 2$. Die Wartekapazität K ist ebenfalls 2.

Wie groß ist die Warteschlangenlänge?

Wie hoch ist die Auslastung des Lifts?

Folgende Referenzverteilung $q_k = P(A(n)=k)$, $k=0,1,2,2^*$ ist gegeben.

Anzahl der eintreffenden Personen	0	1	2	2*
-----------------------------------	---	---	---	----

Wahrscheinlichkeiten q_k	1/8	1/2	1/4	1/8
----------------------------	-----	-----	-----	-----

$A(n)$ ist die zufällige Ankunft von Personen in der Warteschlange vor dem Lift zur Zeit $n\Delta t$

- 1) Modelliere die zufällige Warteschlangenlänge $X(n)$ als homogene Markovkette und führe die Berechnung der Übergangswahrscheinlichkeiten aus.
- 2) Stelle die Tabelle für die Übergangsmatrix auf
- 3) Berechne die ergodischen Wahrscheinlichkeiten.
- 4) Berechne die Systemkenngrößen in der stabilen Phase.

Übergangsintensität

Die direkte Benutzung der Chapman-Kolmogorov-Gleichung zur Bestimmung der Übergangswahrscheinlichkeiten ist schwierig. Gewöhnlich werden daher unter Zuhilfenahme der Grenzwerte

$$q_{ij} = \lim_{t \rightarrow 0} \frac{p_{ij}(t) - \delta_{ij}}{t}$$

$$\text{mit } \delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{sonst.} \end{cases}$$

bestimmt. q_{ij} ist die **Übergangsrate** (-intensität) des Prozesses vom Zustand i in den Zustand j . Da $-q_{ii}$ die Rate ist, mit der Zustand i verlassen wird, gilt:

$$q_{ii} + \sum_{j \in Z, j \neq i} q_{ij} = 0$$

Aus der Chapman-Kolmogorov-Gleichung folgt damit:

$$p_{ij}(t + \Delta t) - p_{ij}(t) = \sum_{k \in Z} (p_{ik}(t + \Delta t - \theta) - p_{ik}(t - \theta)) \cdot p_{kj}(\theta)$$

Dividiert man beide Seiten durch Δt und läßt Δt gegen 0 und θ gegen t gehen, dann erhält man die sog. **Kolmogorov-Rückwärtsgleichungen**:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in Z} p_{jk}(t) \cdot q_{ik}$$

Auf ähnliche Weise erhält man die **Kolmogorov-Vorwärtsgleichung**:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in Z} p_{ik}(t) \cdot q_{kj}$$

Differenziert man $p_j(t) = \sum_{i \in Z} p_{ij}(t) \cdot p_i(0)$ und setzt die soeben gewonnenen Ergebnisse ein, so stellt man fest, daß die zeitabhängigen Zustandswahrscheinlichkeiten des Prozesses durch Lösung des folgenden Differentialgleichungssystems bestimmt werden können:

$$\frac{d}{dt} p_j(t) = \sum_{i \in Z} p_i(t) \cdot q_{ij}$$

Zur Darstellung dieser Gleichung verwendet man häufig die Matrizenschreibweise:

$$\frac{d}{dt} \mathbf{p}^T(t) = \mathbf{p}^T(t) \cdot \mathbf{Q}$$

mit dem Zeilenvektor $\mathbf{p}^T(t) = (p_1(t), p_2(t), p_3(t), \dots)$

Q wird Übergangsratenmatrix oder **Generatormatrix** genannt:

q ₁₁	q ₂₁	q ₃₁
q ₁₂	q ₂₂	q ₃₂
q ₁₃	q ₂₃	q ₃₃
....
....

Markovmodelle

Generell gilt für Markovprozesse im Gleichgewicht: Für jeden Zustand muß der Fluß in diesem Zustand gleich dem Fluß aus dem Zustand sein.

$$\frac{d}{dt} \mathbf{p}^T(t) = \mathbf{p}^T(t) \cdot \mathbf{Q}$$

Die zeitabhängigen Zustandswahrscheinlichkeiten $\mathbf{p}^T(t) = (p_1(t), p_2(t), p_3(t), \dots)$ hängen ab von der Übergangsmatrix

$$\begin{pmatrix} q_{11} & q_{12} & q_{13} & \dots \\ q_{21} & q_{22} & q_{23} & \dots \\ q_{31} & q_{32} & q_{33} & \dots \\ q_{41} & q_{42} & q_{43} & \dots \end{pmatrix}$$

$$q_{ii} - \sum_{j \neq i} q_{ij}$$

und können durch die Lösung des folgenden Differentialgleichungssystems bestimmt werden:

$$\frac{d}{dt} \mathbf{p}^T(t) = \mathbf{p}^T(t) \cdot \mathbf{Q}$$

Bsp.: "Geburts-, Sterbeprozesse"

"Geburts-, Sterbeprozesse sind spezielle Markov-Prozesse, die die einschränkende Bedingung besitzen, daß Übergänge ausschließlich zwischen benachbarten Zuständen des Prozesses möglich sind. Ist k ein beliebiger Prozeßzustand, dann geht der Prozeß mit der Geburtsrate $q_{k,k+1} = \lambda_k$ in den Zustand $k + 1$ über und mit der Sterberate $q_{k,k-1} = \mu_k$ in den Zustand $k - 1$. Für $|k - 1| > 1$ gilt $q_k = 0$.

Die Generatormatrix hat dann folgendes Aussehen:

$-\lambda_0$	λ_0	0	0
μ_1	$-(\lambda_1 + \mu_1)$	λ_1	0	...
0	μ_2	$-(\lambda_2 + \mu_2)$	λ_2	...
0		μ_3	$-(\lambda_3 + \mu_3)$...
...
...

Mit $\frac{d}{dt} \mathbf{p}^T(t) = \mathbf{p}^T(t) \cdot \mathbf{Q}$ und dem Vektor $\mathbf{p}^T(t) = (p_0(t), p_1(t), p_2(t), \dots)$ ergibt sich das folgende Differentialgleichungssystem zur Bestimmung der Zustandswahrscheinlichkeiten des Geburts-, Sterbeprozesses:

$$\frac{d}{dt} p_0(t) = -p_0(t) \cdot \lambda_0 + p_1(t) \cdot \mu_1$$

$$\frac{d}{dt} p_k(t) = -p_k(t) \cdot (\lambda_k + \mu_k) + p_{k-1}(t) \cdot \lambda_{k-1} + p_{k+1}(t) \cdot \mu_{k+1} \quad \text{für } k \geq 1$$

Spezialfall "Reiner Geburtsprozess": Dieser Prozeß besitzt die zusätzliche Eigenschaft $\lambda_k = \lambda$ und $\mu_k = 0$ für alle k. Mit den Anfangsbedingungen

$$p_k(0) = \begin{cases} 1 & \text{für } k = 0 \\ 0 & \text{für } k \neq 0 \end{cases}$$

(, d.h.: Am Anfang ist das Wartesystem leer,)

$$\text{erhält man über } \frac{d}{dt} p_0(t) = -\lambda \cdot p_0(t)$$

$$\text{bzw. } \frac{d}{dt} p_k(t) = \lambda \cdot p_{k-1}(t) - \lambda \cdot p_k(t)$$

die Lösungen

$$p_0(t) = e^{-\lambda t} \quad \text{bzw.} \quad p_k(t) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda t}$$

Ein neuer Geburtsprozeß ist demnach poissonverteilt.

Interpretiert man λ als die durchschnittliche Ankunftsrate, so wird durch einen Poisson-Prozeß die Ankunft von zu bedienenden Einheiten beschrieben. " $p_k(t)$ " ist dann die Wahrscheinlichkeit, daß k Einheiten im Zeitintervall 0 .. t ankommen.

$$\text{Erwartungswert: } \sum_{k=1}^{\infty} k \cdot p_k(t) = e^{-\lambda t} \cdot \sum_{k=1}^{\infty} \frac{(\lambda \cdot t)^k}{(k-1)!} = e^{-\lambda t} \cdot \lambda \cdot t \cdot \sum_{k=0}^{\infty} \frac{(\lambda \cdot t)^k}{k!}$$

$$\text{Da } e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \text{ ist, ergibt sich } \lambda \cdot t$$

Stationärer Zustand

Das Hauptinteresse bei der Betrachtung von Markov-Ketten gilt der Untersuchung des **stationären Zustands**. Dieser stationäre Zustand stellt sich genau dann ein, wenn der Prozeß hinreichend lange Zeit gelaufen ist, bis sämtliche Einschwingvorgänge abgeklungen sind. Die Wahrscheinlichkeit für einen bestimmten Prozeßzustand ist dann unabhängig vom Anfangszustand des Prozesses und ändert sich auch nicht mehr mit der Zeit. Man sagt auch der Prozeß befindet sich im **statistischen Gleichgewicht**. Die Existenz eines stationären Zustands ist an verschiedene Bedingungen geknüpft, die von der Markov-Kette bzw. deren Zuständen erfüllt werden müssen. Zur Beschreibung dieser Bedingungen sind noch einige Vereinbarungen zu treffen:

Eine Markov-Kette ist **irreduzibel**, wenn jeder Zustand der Kette von jedem anderen Zustand der Kette erreicht werden kann, d.h. $p_{ij}(t) > 0$ für alle $i, j \in Z$. Ein Zustand einer Markov-Kette heißt **transient**, wenn der Prozeß nach Verlassen dieses Zustands mit einer Wahrscheinlichkeit größer als 0 nie wieder in diesen Zustand zurückkehrt. Kehrt der Prozeß mit der Wahrscheinlichkeit Eins wieder dorthin zurück, dann wird der Zustand **rekurrent** genannt. Die nach Verlassen dieses Zustands bis zur ersten Rückkehr verstrichene Zeit heißt **Rekurrenzzeit**. Je nachdem, ob die mittlere Rekurrenzzeit für einen Zustand endlich oder unendlich ist, nennt man den Zustand **rekurrent nichtnull** oder **rekurrent null**.

Es kann gezeigt werden, daß für **irreduzible homogene** Markov-Ketten stets die Grenzwerte $p_j = \lim_{t \rightarrow \infty} p_{ij}(t)$ existieren und unabhängig von den Anfangsbedingungen sind.

Wenn dann sämtliche Zustände einer Markov-Kette transient oder rekurrent sind, folgt, daß die obigen Grenzwerte $p_j = 0$ sind. In diesem Fall existiert für diesen Prozeß kein stationärer Zustand.

Demgegenüber existiert ein stationärer Zustand, wenn sämtliche Zustände der Markov-Kette **rekurrent nichtnull** sind. In diesem Fall sind sämtliche Grenzwerte p_j größer als Null, und die Zustände der Markov-Kette, wie auch die Markov-Kette selbst, werden **ergodisch** genannt. Die Grenzwahrscheinlichkeiten einer ergodischen Markov-Kette heißen **stationäre Zustandswahrscheinlichkeiten** oder **Gleichgewichtszustandswahrscheinlichkeiten** und erfüllen die Beziehung:

$$p_j = \lim_{t \rightarrow \infty} p_{ij}(t)$$

Existieren diese Grenzwahrscheinlichkeiten, dann gilt: $\lim_{t \rightarrow \infty} \frac{d}{dt} p_j(t) = 0$

Mit $\frac{d}{dt} p_j(t) = \sum_{i \in Z} p_i(t) \cdot q_{ij}$ ergibt sich die Lösung für die stationären Zustandswahrscheinlichkeiten aus dem folgenden linearen Gleichungssystem $\sum_{i \in Z} p_i \cdot q_{ij}$ und der Normalisierungsbedingung: $\sum_{i \in Z} p_i = 1$

Es ist üblich, das vorstehende Gleichungssystem in Matrixform zu schreiben:

$$\mathbf{p}^T \mathbf{Q} = 0$$

mit dem Zeilenvektor $\mathbf{p}^T = (p_0, p_1, p_2, \dots)$ als dem Lösungsvektor für die Gleichgewichtszustandswahrscheinlichkeiten und \mathbf{Q} als Generatormatrix.

Bsp.:

1. Gegeben ist $p_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Existieren stationäre Zustandswahrscheinlichkeiten (Grenzwahrscheinlichkeiten einer ergodischen Markov-Kette).

$$\lim_{t \rightarrow \infty} p_{11}(t) = 1$$

$$\lim_{t \rightarrow \infty} p_{21}(t) = q$$

$$\lim_{t \rightarrow \infty} p_{31}(t) = 0$$

Nach dem Ergodentheorem müßte sein: $\lim_{t \rightarrow \infty} p_{ij}(t) = p_j$

Das ist hier nicht der Fall.

Begründung: $p_{ij}(t) = (p_{ij})^t$ ergibt bspw. für $p_{ij}(2) = (p_{ij})^2$

$$\begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix} \text{ bzw. für } t = n: \begin{pmatrix} 1 & 0 & 0 \\ q & 0 & p \\ 0 & 0 & 1 \end{pmatrix}$$

Auch $\sum_{j \in \mathbb{Z}} p_j = 1$ ist nicht erfüllt.

2. Für die "Irrfahrt eines Teilchens" existieren folgende Übergangswahrscheinlichkeiten:

$$p_{ij} = \begin{pmatrix} q & p & 0 & 0 & \dots & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & q & p \end{pmatrix}$$

Stationäre Verteilung

Das ist die Wahrscheinlichkeitsverteilung, falls $\mathbf{p}^T \mathbf{Q} = 0$ ist.

Für entsprechend groß bemessene Werte von t ist $p_{ij}(t-1) = p_{ij}(t) = p_{ij}(t+1) = p_j$. Daraus ergeben sich für die ergodischen Wahrscheinlichkeiten die Bestimmungsgleichungen:

$$p_j = \sum_{k \in \mathbb{Z}} p_k \cdot p_{kj} \text{ mit } \sum_{j \in \mathbb{Z}} p_j = 1$$

Für $l = 3$ nimmt die Übergangsmatrix zur Beschreibung der "Irrfahrt eines Teilchens" folgende Gestalt an:

$$p_{ij} = \begin{pmatrix} q & p & 0 \\ q & 0 & p \\ 0 & q & p \end{pmatrix}$$

Die Voraussetzungen für Ergodizität sind hier mit $p = q = 1/2$ erfüllt. Es gilt

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j$$

$$\text{und somit } \begin{pmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{pmatrix}$$

$$p_1 = 1/2 \cdot p_1 + 1/2 \cdot p_2$$

$$p_2 = 1/2 \cdot p_1 + 1/2 \cdot p_3$$

$$p_3 = 1/2 \cdot p_2 + 1/2 \cdot p_3$$

$$p_1 = p_1 + p_2 + p_3$$

Lösung: $p_1 = 1/3, p_2 = 1/3, p_3 = 1/3$

Aufgabe

Eine Reperaturwerkstatt bearbeitet 2 Motortypen M1 und M2. Die Reperatur von M1 dauert zwei Tage, die von M2 dauert einen Tag. An jedem Morgen ist die Wahrscheinlichkeit dafür, daß ein M1-Motor repariert werden muß $1/3$. Die Wahrscheinlichkeit, daß ein M2-Motor reperariert werden muß, ist $1/2$. Reperaturaufträge, die die Werkstatt nicht fertigstellen kann, werden anderweitig vergeben.

Ist ein Arbeitstag der erste Reperaturtag eines M1-Motors, dann wird jede neue Arbeit für den nächsten Tag zurückgewiesen. An jedem anderen Tag wird jeder Motortyp (jeweils nur eine Maschine) zur Reperatur angenommen.

2 Strategien sind möglich:

- a) Die Reperatur von M1
- b) Die Reperatur von M2

Welche Politik ist besser?

Lösung

Berechnung der Übergangsmatrizen:

Mögliche Zustände

Zustand 0 ... keine Arbeit
 Zustand 1 ... 1. Reperaturtag eines M1 - Motors
 Zustand 2 ... 2. Reperaturtag eines M1 - Motors
 Zustand 3 ... Reperaturtag eines M2 - Motors

Wahrscheinlichkeiten WS

$$WS(\text{kein M1 - Motor fällt zur Reperatur an}) = (1 - 1/3)$$

$$WS(\text{kein M2 - Motor " " " "}) = (1 - 1/2)$$

$$WS(\text{keine Arbeit}) = (1 - 1/3)(1 - 1/2)$$

Die WS für "keine Arbeit" ist demnach unabhängig vom Zustand des vorangegangenen Tages. Das gilt nicht, wenn mit der Reperatur eines M1-Motors begonnen wurde. In diesem Fall ist WS nämlich 0.

Übergangsmatrix zur 1. Politik

	0	1	2	3
0	$(1-1/2) \cdot (1-1/3)$	$1/3$	0	$1/2 \cdot (1-1/3)$
1	0	0	1	0
2	$(1-1/2) \cdot (1-1/3)$	$1/3$	0	$1/2 \cdot (1-1/3)$
3	$(1-1/2) \cdot (1-1/3)$	$1/3$	0	$1/2 \cdot (1-1/3)$

$$= \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \end{pmatrix}$$

Übergangsmatrix zur 2. Politik

	0	1	2	3
0	$(1-1/3) \cdot (1-1/2)$	$(1-1/2) \cdot 1/3$	0	$1/2$
1	0	0	1	0
2	$(1-1/3) \cdot (1-1/2)$	$1/3$	0	$1/2$
3	$(1-1/3) \cdot (1-1/2)$	$1/3$	0	$1/2$

$$= \begin{pmatrix} 1/3 & 1/6 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 1/3 & 1/6 & 0 & 1/2 \\ 1/3 & 1/6 & 0 & 1/2 \end{pmatrix}$$

Ergodizität ist gegeben, wenn $p_j = \lim_{t \rightarrow \infty} p_{ij}^{(n)}$

Das bedeutet: $p_{ij}^{(n)} = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix}$

Mit der Übergangsmatrix zur 2. Politik muß dann gelten:

$$\begin{pmatrix} p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix} \cdot \begin{pmatrix} 1/3 & 1/6 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \\ 1/3 & 1/6 & 0 & 1/2 \\ 1/3 & 1/6 & 0 & 1/2 \end{pmatrix} = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix}$$

Daraus folgen die Gleichungen:

$$1/3 \cdot p_0 + 0 \cdot p_1 + 1/3 \cdot p_2 + 1/3 \cdot p_3 = p_0$$

$$1/6 \cdot p_0 + 0 \cdot p_1 + 1/6 \cdot p_2 + 1/6 \cdot p_3 = p_1$$

$$p_1 = p_2$$

$$1/2 \cdot p_0 + 0 \cdot p_1 + 1/2 \cdot p_2 + 1/2 \cdot p_3 = p_3$$

$$\text{Außerdem gilt: } p_0 + p_1 + p_2 + p_3 = 1$$

Die Auflösung der Gleichungen ergibt: $p_0=2/7$, $p_1=p_2=1/7$, $p_3=3/7$.

Die 1. Politik führt zu: $p_0=1/4$, $p_1=1/4$, $p_2=1/4$, $p_3=1/4$.

Folgerung: Die 1. Politik, die die Werkstatt während $3/4$ der Gesamtzeit voll beschäftigt, ist günstiger als die 2. Politik. Die Werkstatt ist hier nur zu $5/7$ voll ausgelastet.

Gleichgewichtswahrscheinlichkeiten für Geburts-/Sterbeprozesse

Für die Existenz der Gleichgewichtswahrscheinlichkeiten eines Geburts-/Sterbeprozesses müssen sämtliche Zustände dieses Prozesses ergodisch sein. Die ist der Fall, wenn $\frac{\lambda_k}{\mu_k} < 1$ ist. dann existiert die Grenzwahrscheinlichkeit

$$p_j = \lim_{t \rightarrow \infty} p_j(t) \text{ und es gilt: } \lim_{t \rightarrow \infty} \frac{d}{dt} p_j(t) = 0$$

Das Differentialgleichungssystem zur Bestimmung des Geburts-/ Sterbeprozesses nimmt damit folgende Gestalt an:

$$0 = -p_0 \cdot \lambda_0 + p_1 \cdot \mu_1 \text{ bzw. } 0 = -p_k \cdot (\lambda_k + \mu_k) + p_{k-1} \cdot \lambda_{k-1} + p_{k+1} \cdot \mu_{k+1} \text{ (für } k \geq 1 \text{)}$$

$$\text{Daraus folgt: } p_1 = \frac{\lambda_0}{\mu_1} \cdot p_0$$

$$\text{bzw. für } k = 1: p_2 = \frac{\lambda_0 \cdot \lambda_1}{\mu_1 \cdot \mu_2} \cdot p_0$$

$$\text{bzw. für } k = 2: p_3 = \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2}{\mu_1 \cdot \mu_2 \cdot \mu_3} \cdot p_0$$

$$\text{bzw. für } k = n: p_3 = \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \dots \cdot \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdot \dots \cdot \mu_n} \cdot p_0$$

$$\text{bzw. allgemein } p_k = \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \cdot p_0 \text{ (für } k \geq 1 \text{)}$$

$$p_0 \text{ kann über } \sum_{i \in \mathbb{Z}} p_i = 1 \text{ bestimmt werden zu: } p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}}$$

Bsp.: Wartesystem M/M/1

Dieses Wartesystem ist ein spezieller Geburts-, Sterbeprozess mit der konstanten Geburtsrate λ und der konstanten Sterberate μ . Die Ergodizitätsbedingung ist $\frac{\lambda}{\mu} < 1$

bzw. $\lambda < \mu$ Setzt man die angegebenen Raten in die vorstehenden Gleichungen ein,

$$\text{dann erhält man: } p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda}{\mu}} = \frac{1}{1 + \sum_{k=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k}$$

$$\text{Da } \frac{\lambda}{\mu} < 1 \text{ ist, konvergiert die Summe zu: } p_0 = \frac{1}{1 + \frac{\lambda/\mu}{1 - \lambda/\mu}} = 1 - \frac{\lambda}{\mu}$$

$$\text{Mit } p_k = \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \cdot p_0 \text{ ist: } p_k = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^k = \left(1 - \frac{\lambda}{\mu}\right) \cdot \left(\frac{\lambda}{\mu}\right)^k$$

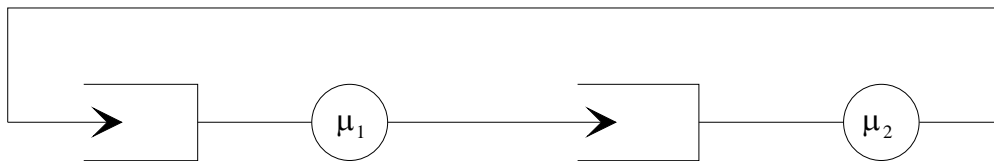
Globale Gleichgewichtsbedingungen

Viele Warteschlangenmodelle können mit Hilfe von Markov-Ketten beschrieben werden. Diese Markov-Ketten sind durch die Übergangsmatrizen zwischen Zuständen des jeweils zugrundeliegenden Modells vollständig definiert. Ist die Markov-Kette ergodisch, dann existiert ein Gleichgewichtszustand und sämtliche Zustandswahrscheinlichkeiten des Modells sind zeitunabhängig:

$$\vec{p}^T \cdot \vec{Q} = 0$$

Mit dieser Gleichung wird zum Ausdruck gebracht, daß im Gleichgewicht für jeden beliebigen Zustand eines Warteschlangenmodells der Fluß aus diesem Zustand gleich dem Fluß in diesen Zustand sein muß (Prinzip von der Erhaltung des statistischen Gleichgewichts).

Bsp.: Gegeben ist das folgende geschlossene Warteschlangennetz. In diesem Netz befinden sich 2 Knoten ($N = 2$) und 3 Aufträge ($K = 3$). Die Rechenzeiten sind exponentiell verteilt mit den Mittelwerten . Die Warteschlangendisziplin ist bei beiden Knoten FCFS (First-Come-First-Served). μ_1, μ_2



Der Zustandsraum der Markov-Kette, die das Verhalten dieses Beispielnetzes beschreibt, ergibt sich aus den möglichen Zuständen des Netzes:

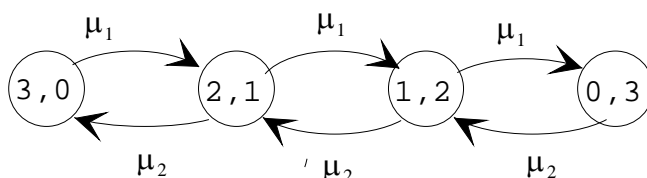
$$\{(3,0), (2,1), (1,2), (0,3)\}$$

Zustand (k_1, k_2) gibt an, daß sich k_1 Aufträge im Knoten 1 und k_2 Aufträge im Knoten 2 befinden. $p(k_1, k_2)$ bezeichnet die Wahrscheinlichkeit für diesen Zustand im Gleichgewicht.

Bestimmung der einzelnen Übergangsraten zwischen den einzelnen Zuständen:

Betrachtet man bspw. Zustand $(1,2)$, dann findet ein Übergang von $(1,2)$ in den Zustand $(0,3)$ genau dann statt, wenn ein Auftrag im Knoten 1 fertig bedient wird, die dazugehörige Rate ist μ_1 . Entsprechend ist dann μ_2 die Übergangsrate vom Zustand $(1,2)$ nach Zustand $(2,1)$.

Ein einfaches Hilfsmittel zur Beschreibung der Markov-Ketten sind **Zustandsübergangsdiagramme**. Darunter versteht man gerichtete Graphen, deren Knoten die Zustände und deren Kanten die möglichen Übergänge zwischen den Zuständen der Markov-Kette repräsentieren. Die Kanten sind zusätzlich mit den Übergangsraten markiert.



Mit Hilfe dieses Diagramms können die globalen Gleichgewichtsbedingungen formuliert werden: Der Fluß in einem bestimmten Zustand des Modells ist im Diagramm dargestellt durch alle Kanten, die genau in den betreffenden Zustand führen, und alle Kanten, die genau aus diesem Zustand herausführen. Die globalen Gleichgewichtsbedingungen für das Bsp. lauten somit:

$$p(3,0) \cdot \mu_1 = p(2,1) \cdot \mu_2$$

$$p(2,1) \cdot (\mu_1 + \mu_2) = p(3,0) \cdot \mu_1 + p(1,2) \cdot \mu_2$$

$$p(1,2) \cdot (\mu_1 + \mu_2) = p(2,1) \cdot \mu_1 + p(0,3) \cdot \mu_2$$

$$p(0,3) \cdot \mu_2 = p(1,2) \cdot \mu_1$$

Mit der Generatormatrix

$$\begin{pmatrix} -\mu_1 & \mu_1 & 0 & 0 \\ \mu_2 & -(\mu_1 + \mu_2) & \mu_1 & 0 \\ 0 & \mu_2 & -(\mu_1 + \mu_2) & \mu_1 \\ 0 & 0 & \mu_2 & -\mu_2 \end{pmatrix}$$

und dem Lösungsvektor $\mathbf{p}^T = (p(3,0), p(2,1), p(1,2), p(0,3))$ ergibt sich dann die Gültigkeit der Matrixgleichung $\vec{p} \cdot \vec{Q} = 0$. Mit dieser Gleichung können die Zustandswahrscheinlichkeiten und daraus alle anderen Leistungsgrößen berechnet werden. Zu diesen Gleichgewichtsbedingungen kommt hinzu, daß sich die Wahrscheinlichkeiten für die einzelnen Systemzustände zu Eins aufsummieren müssen:

$$p(3,0) + p(2,1) + p(1,2) + p(0,3) = 1$$

5.2 Elementare Wartesysteme

5.2.1 Beschreibung

Ein elementares Wartesystem besteht aus einer Warteschlange sowie aus einer oder mehreren identischen Bedieneinheiten, in denen Aufträge bedient werden. Solch ein Bediensystem wird in der Warteschlangentheorie auch Bedienstation oder Knoten genannt. Häufig ist auch die Beziehung "single server" für Wartesysteme mit einer Bedieneinheit bzw. "multiple server" für Wartesysteme mit mehreren Bedieneinheiten zu finden.

Eine einzelne Bedieneinheit eines Wartesystems kann immer nur einen Auftrag gleichzeitig bedienen. Sie befindet sich deshalb in einem der beiden Zustände "belegt" oder "frei". Sind bei der Ankunft eines Auftrags an einem Wartesystem alle Bedieneinheiten belegt, so muß sich der Auftrag in die gemeinsame Warteschlange einreihen. Wird ein Auftrag in einer Bedieneinheit fertig bedient und diese somit frei, dann wird aus der Warteschlange ein neuer Auftrag entsprechend einer **Warteschlangendisziplin** ausgewählt, mit dessen Bedienung daraufhin begonnen wird.

Zu den Zeitpunkten t_k (mit $0 < t_1 < t_2 < t_3 \dots$) kommen zu bedienende Einheiten an und reihen sich in der Warteschlange ein. $t_k - t_{k-1}$ ist dann die Zwischenankunftszeit der k -ten zu bedienenden Einheit. Ein Wartesystem ist charakterisiert durch die **Zwischenankunftszeit** t_a und durch die Bedienzeit t_b für die einzelnen Aufträge. Da der Zugang der Aufträge genau wie deren Abfertigung in der Regel zufällig erfolgt, sind die Zwischenankunftszeit und die Bedienzeit Zufallsgrößen und damit durch ihre Verteilung gegeben.

Aus der mittleren Zwischenankunftszeit erhält man durch Kehrwertbildung die **Ankunftsrate** λ .

λ kann als die mittlere Anzahl von Aufträgen, die je Zeiteinheit eintreffen, interpretiert werden.

Im Gegensatz zum Ankunftsprozeß ist der Bedienprozeß außer von den zu bearbeitenden Aufträgen zusätzlich noch vom gegebenen Wartesystem abhängig. Vorausgesetzt wird aber, daß alle Bedienanforderungen rein zufällig sind. So können auch die Bedienzeiten für alle Aufträge als Zufallsgrößen angesehen werden. Es gilt dabei die Annahme, daß sämtliche Aufträge im Wartesystem die gleiche Bedienzeitverteilung besitzen. Aus der mittleren Bedienzeit ergibt sich dann die **Bedienrate** μ .

μ bezeichnet die mittlere Anzahl von Aufträgen, die je Zeiteinheit bedient werden.

5.2.2 Leistungsgrößen

Leistungsgrößen für Warteschlangenmodelle sind zeitabhängig. In der Regel interessieren nur Ergebnisse im **stationären Zustand**. Hier sind sämtliche Einschwingvorgänge abgeklungen und die Leistungsgrenzen zeitunabhängig. Man sagt dann auch, das System befindet sich im statistischen Gleichgewicht. Das bedeutet: Die Rate mit der Aufträge ankommen ist gleich der Rate, mit der Aufträge abgehen.

Die wichtigsten Leistungsgrößen sind:

Zustandswahrscheinlichkeit p_k :

Es gilt: $p(k) = P(\text{es befinden sich } k \text{ Aufträge im System})$

Auslastung

Bei Wartesystemen mit einer Bedieneinheit (single servers) gibt die Auslastung den Bruchteil der Gesamtzeit an, den die Bedieneinheit aktiv (belegt) ist. Sie berechnet sich zu

$$\rho = \frac{\text{mittlere_Bedienzeit}}{\text{mittlere_Zwischenankunftszeit}} = \frac{\text{Ankunftsrate}}{\text{Bedienrate}} = \frac{\lambda}{\mu}$$

Bei mehreren Bedieneinheiten (multiple servers) gibt die Auslastung auch den mittleren Anteil der aktiven Bedieneinheiten an. Da $m \cdot \mu$ die Gesamtbedienrate ist, gilt

$$\rho = \frac{\lambda}{m \cdot \mu}$$

Für ein stabiles System (Gleichgewichtszustand) muß die Bedingung $\rho < 1$ erfüllt sein, d.h.: Es dürfen im Mittel nicht mehr Aufträge ankommen als bedient werden können.

Antwortzeit W

Als Antwortzeit (Verweilzeit) bezeichnet man die Zeitspanne, die ein Auftrag in einem Wartesystem verbringt

Wartezeit W^q

Die Zeitspanne gibt an, wie lange ein Auftrag in der Warteschlange warten muß bis seine Bearbeitung beginnt. Es gilt

Antwortzeit = Wartezeit + Bedienzeit

Da W bzw. W^q meistens als Zufallsgrößen gegeben sind, wird ihr Mittelwert berechnet: $\bar{W} = \bar{W}^q + \frac{1}{\mu}$

Warteschlangenlänge L^q

Die Anzahl der Aufträge in der Warteschlange wird Warteschlangenlänge genannt

Anzahl der **Aufträge im Wartesystem L** (bzw. k)

$$\bar{L} = \bar{k} = \sum_{k=1}^{\infty} k \cdot p(k)$$

Das Gesetz von Little

$$L = \lambda \cdot W \text{ bzw. } L^q = \lambda \cdot W^q$$

Diese Gesetzmäßigkeit gilt für alle Warteschlangendisziplinen und beliebige G/G/m-Systeme.

5.2.3 Charakterisierung (Kurzschreibweise) von Wartesystemen

Zur einheitlichen Beschreibung elementarer Wartesysteme hat sich die folgende (**Kendall'sche**) Schreibweise durchgesetzt:

A/B/m - Warteschlangendisziplin

A beschreibt die Verteilung der Zwischenankunftszeiten und **B** die Verteilung der Bedienzeiten des Wartesystems. "**m**" gibt die Anzahl der identischen (parallelen) Bedieneinheiten an ($m \geq 1$). Für **A** und **B** werden folgende Symbole verwendet:

M	Exponentialverteilung oder Markov
E_k	Erlang-Verteilung mit k Phasen
H_k	Hyperexponentialverteilung mit k Phasen
D	Deterministische Verteilung, d.h. die Zwischenankunft- bzw. Bedienzeiten sind konstant.
G	Allgemeine Verteilung
GI	Allgemeine unabhängige Verteilung

Durch **A** wird der Forderungsstrom gekennzeichnet. Dabei bedeutet u.a.:

- A = M: Ein Poissonscher Forderungsstrom liegt vor, d.h. in keinem Zeitpunkt trifft mehr als eine Forderung ein und die Zeiten zwischen zwei eintreffenden Forderungen sind unabhängige, positive, identisch verteilte Zufallsgrößen
- A = GI: (generell independent). Ein rekurrenter Forderungsstrom liegt vor, d.h. in keinem Zeitpunkt trifft mehr als eine Forderung ein und die Zeiten zwischen 2 eintreffenden Forderungen sind unabhängige, positive, identisch verteilte Zufallsgrößen
- A = D: (deterministic). Es liegt ein Forderungsstrom mit konstanten zeitlichen Abständen zwischen den einzelnen Forderungen vor.

Mit **B** wird die Folge der Bedienungen auf dem Bedienungsapparat erfasst. So bedeutet bspw.

- B = M: (Markov). Hinsichtlich der Bedienungszeit liegt bei jedem Bedienungsapparat eine Folge unabhängiger, identisch exponentiell verteilter Bedienungszeiten vor.
- B = G: (general). Hinsichtlich der Bedienungszeit liegt bei jedem Bedienungsapparat eine Folge unabhängiger, positiver, identisch verteilter Bedienungszeiten vor.

Die **Warteschlangendisziplin** (Bedienstrategie) legt fest, welcher Auftrag aus der Warteschlange als nächstes zur Bedienung ansteht.

Es gibt eine Vielzahl unterschiedlicher Warteschlangendisziplinen. Die wichtigsten sind:

FCFS (First-Come-First-Served)

Die Aufträge werden in der Reihenfolge ihrer Ankunft bedient. Ist in der Kendall'schen Notation keine Disziplin angegeben, so impliziert dies FCFS.

LFCS (Last-Come-First-Served)

Der zuletzt angekommene Auftrag wird als nächster bedient.

SIRO (Service-In-Random-Order)

Die Auswahl ist rein zufällig.

RR (Round Robin)

Ist die Bedienung eines Auftrags nach einer fest vorgegebenen Zeitscheibe noch nicht beendet, so wird der Auftrag verdrängt und wieder in die Warteschlange eingereiht, die nach FCFS abgearbeitet wird. Dies wiederholt sich so oft, bis der Auftrag vollständig bedient ist.

Statische Prioritäten: Die Auswahl erfolgt nach einer fest vorgegebenen Priorität der Aufträge. Innerhalb einer Prioritätsklasse (mit gleichen Prioritäten) erfolgt die Auswahl nach FCFS

Dynamische Prioritäten: Die Auswahl erfolgt nach dynamischen Prioritäten, die sich in Abhängigkeit von der Zeit ändern.

Bsp. zur Kendall'schen Notation:

M/G/1 - LCFS Preemptive Resume (PR) beschreibt ein elementares Warteschlangensystem mit exponentiell verteilter Zwischenankunftszeit, beliebig verteilter Bedienzeit sowie einer einzigen Bedieneinheit. Die Bedienstrategie ist LCFS. Es gilt: Ein neu ankommender Auftrag unterbricht den gerade in Bedienung befindlichen Auftrag, um selbst bedient zu werden. Der unterbrochenen Auftrag wird erst dann weiterbedient, wenn alle nach ihm angekommenen Aufträge vollständig abgearbeitet sind.

Es gibt verschiedene Möglichkeiten zur Erweiterung der Kendall'schen Notation, z.B.

A/B/m/K/M

K gibt die Anzahl der im System vorhandenen Warteplätze an
(Systemkapazität)

M beschreibt die Anzahl der zu bedienenden Einheiten
(Quellenelemente, Population)

Bsp.: Die Struktur M/M/m/K beschreibt ein sog. Poissonsches Bedienungssystem. Die Bedienungssituation in einem derartigen System ist durch folgende Merkmale gekennzeichnet:

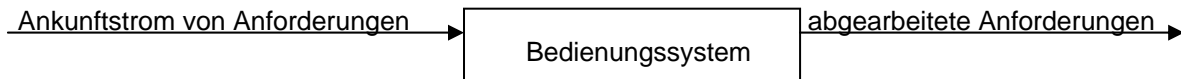
1. Der Forderungsstrom nach Bedienung wird durch einen Poisson-Prozeß beschrieben
2. Bei allen Bedieneinheiten sind die Bedienzeiten für jede Forderung nach Bedienung exponentiell verteilte unabhängige Zufallsgrößen.
3. Die Anzahl der Bedienstationen ist m.
4. Eintreffende Forderungen nach Bedienung
 - verzichten bei besetzten Bedienungsapparaten auf eine Abfertigung (Verlustsystem)
 - ordnen sich bei besetzten Bedieneinheiten in eine Warteschlange mit der Disziplin FIFO ein.

Poissonsche Bediensysteme sind Geburts-, Todesprozesse im stationären Zustand.

5.3 Bedienungssysteme

Bediensysteme sind durch folgende Komponenten bestimmt:

- (1) die Eigenschaften des Ankunftsstroms
- (2) die Anforderungen an die Bedienzeit (Bedienzeitverteilung)
- (3) die Bedienungsstrategie



1. Ankunftsprozeß

Der Ankunftsprozeß hat stationäre Zuwächse, falls die Verteilung der Anzahl der Ankünfte im Zeitraum $(t, t + \Delta t)$ nicht von t abhängt. In diesem Fall ist die mittlere Anzahl der Ankünfte (in einer ZE) die Intensität λ des Ankunftsstroms:

$$\lambda = E(A(t+1) - A(t))$$

Der Ankunftsprozeß heißt ergodisch, falls mit Wahrscheinlichkeit 1 gilt: $\lim_{t \rightarrow \infty} \frac{A(t)}{t} = \lambda$

Der Ankunftsprozeß hat unabhängige Zwischenankunftszeiten, wenn die Zeitintervalle $\{\Delta t_i\}$ zwischen je zwei Ankünften unabhängig identisch verteilt sind.

Besonders wichtig ist ein Ankunftsprozeß, bei dem die Zwischenankunftszeiten unabhängig mit der Verteilungsfunktion $F(x) = 1 - e^{-\lambda x}$ verteilt sind. Es handelt sich dabei um den Poissonprozeß mit der Intensität λ . Dieser Prozeß ist ein Markov-Prozeß mit stationären Zuwächsen.

2. Bedienzeitverteilung ("service time distribution")

Jeder Kunde fordert eine gewisse Bedienzeit vom Bedienungssystem. Der Bedienungsprozeß ist stationär, falls $\{S_n\}$ ¹⁵ eine Folge unabhängiger, identisch verteilter Zufallsvariabler ist.

3. Bedienungsstrategie

Nicht alle Anforderungen können gleichzeitig bedient werden. Die im System befindlichen Einheiten (Kunden $K_s(t)$) können in die zum Zeitpunkt t bedienten Kunden $K_b(t)$ und nicht bedienten (d.h. wartenden) Kunden $K_w(t)$ zerlegt werden: $K_s(t) = K_b(t) + K_w(t)$. Der Systemteil, der die Zuordnung der Kunden zu den Mengen $K_b(t)$ bzw. $K_w(t)$ trifft, heißt **Scheduler**. Die Vorschrift nach der der Scheduler arbeitet, ist die **Bedienungsstrategie**. Scheduler können sehr komplizierte Strategien haben und Kunden Bedienungselemente zuweisen und diese auch wieder entziehen (vernetztes Bedienungssystem). Werden Kunden nur einmal einem Bedienungselement zugewiesen, spricht man von einem einfachen Bedienungssystem.

¹⁵¹⁵ S_n ist die von n -ten Kunden geforderte Zeit für die Bedienung

5.3.1 Markov'sche Bedienungssysteme

5.3.1.1 Stationäre Verteilungen

5.3.1.1.1 M/M/m-Systeme

M / M / 1

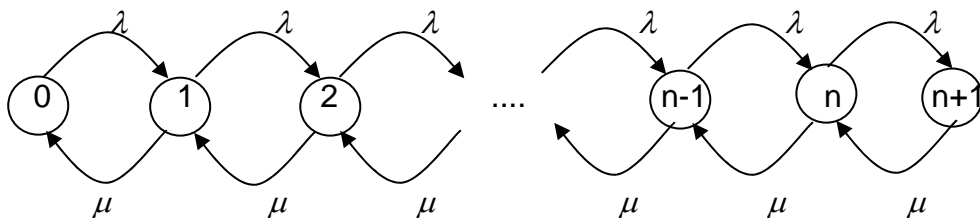
Berechnungsgrundlagen

Dieses System ist gekennzeichnet durch

$$\lambda_n(t) = \lambda \quad (n = 0, 1, 2, \dots)$$

$$\mu_n(t) = \mu \quad (n = 1, 2, \dots)$$

und dem folgenden Zustansübergangsdiagramm



Werden die Bedingungen des vorliegenden Systems in die Lösung eines Geburts-, Todesprozesses im stationären Zustand eingefügt, dann ergibt sich

$$p_n = p_0 \cdot \prod_{i=0}^{n-1} \frac{\lambda}{\mu} = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n \text{ d.h. } p_n = p_0 \cdot \left(\frac{\lambda}{\mu}\right)^n \text{ für } n \geq 0.$$

Ergodizität wird dann erreicht, wenn $\frac{\lambda}{\mu} \leq 1$ ist.

Mit $\rho = \frac{\lambda}{\mu}$ (Verkehrsintensität) ergibt sich $p_0 = 1 - \rho$ bzw. $p_n = (1 - \rho) \cdot \rho^n$

Mittlere Anzahl der Einheiten im Wartesystem: $L = \sum_{n=0}^{\infty} n \cdot p_n = (1 - \rho) \cdot \sum_{n=0}^{\infty} n \cdot \rho^n = \frac{\lambda}{\mu - \lambda}$

Mittlere Anzahl der Einheiten in der Warteschlange:

$$L^q = \sum_{n=0}^{\infty} (n - 1) \cdot p_n = L - (1 - p_0) = \frac{\rho^2}{1 - \rho}$$

Durchschnittliche Wartezeit: $W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$

Anwendungen

Die Zweigstelle einer Unternehmensberatung besitzt ein einziges Terminal, das direkt (on-line) 8 Stunden je Tag an einen zentralen Rechner angeschlossen ist. Die Ingenieure der Beratungsgesellschaft, die an verschiedenen Einsatzorten tätig sind,

fahren zu der Zweigstelle, um dort ihre Berechnungen abzuwickeln. Durchschnittlich benutzen 10 Personen während eines Tages das Terminal. Dabei verbringt durchschnittlich ein Ingenieur an dem Terminal 1/2 Stunde. So ist das Terminal nur zu 5/8 ($10 \cdot \frac{1}{2} = 5$ Stunden von 8 zur Verfügung stehenden Stunden) genutzt.

Trotzdem erhält der Zweigstellenleiter ständig Beschwerden von den Ingenieuren, daß sie zu lange auf die Benutzung des Terminals warten müssen. Es liegt ein M/M/1-System vor.

Bestimmung der mittleren Ankunftsrate und der Verkehrsintensität

$$\text{Ankunftsrate} = \frac{10 \text{ Einheiten}}{8 \cdot 60 \text{ (Min)}} = \frac{10}{480} = \frac{1}{48}$$

$$\rho = \lambda \cdot E(S) = \lambda \cdot 30 = \frac{10}{480} \cdot 30 = \frac{30}{48} = \frac{5}{8}$$

1) Wie groß ist die Wahrscheinlichkeit, daß sich zwei oder mehr Ingenieure in der Zweigstelle aufhalten?

2) Bestimme die Anzahl der Ingenieure, die sich in der Zweigstelle aufhalten

3) Bestimme die Anzahl der Ingenieure, die sich im Mittel in der Zweigstelle aufhalten.

$M / M / \infty$

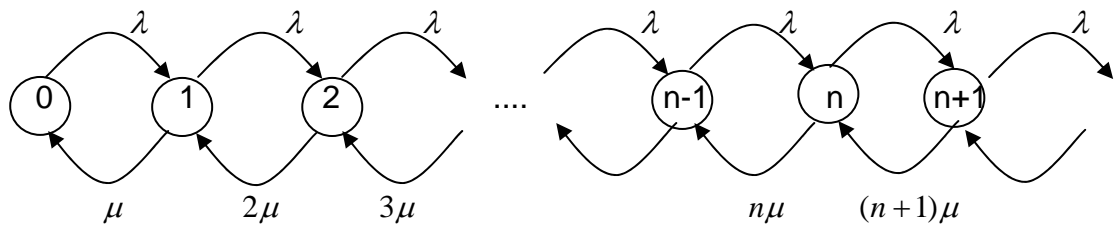
Berechnungsgrundlagen

Das System ist gekennzeichnet durch

$$\lambda_n(t) = \lambda$$

$$\mu_n(t) = n \cdot \mu$$

und das folgende Zustandsdiagramm



Die Differentialgleichungen des Geburts-, Todesprozesses für den stationären Fall führen auf die Wahrscheinlichkeiten:

$$p_n = \frac{\rho^n}{n!} \cdot e^{-\lambda/\mu}$$

Mittlere Anzahl von Einheiten im Wartesystem: $L = \frac{\lambda}{\mu}$

Durchschnittl. Wartezeit: $W = \frac{1}{\mu}$. Das Ergebnis ist einleuchtend, findet jedoch jede ankommende und zu bedienende Einheit eine Bedienungseinrichtung vor.

M/M/m**Berechnungsgrundlagen**

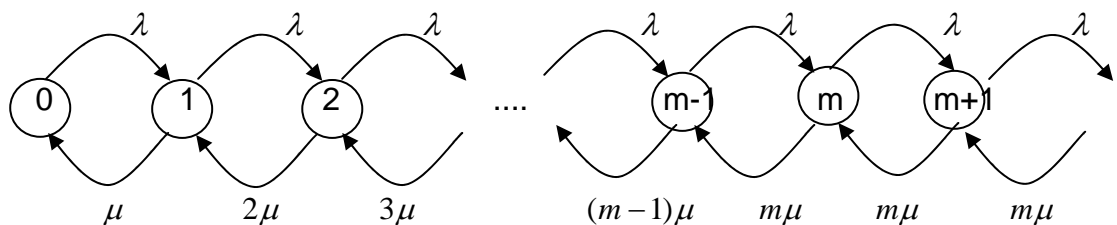
Es gibt m parallel und unabhängig voneinander arbeitende Bedienungseinrichtungen (jeweils mit der gleichen Bedienungsrate μ).

Dieses System ist gekennzeichnet durch

$$\lambda_n(t) = \lambda \quad (n = 0, 1, 2, \dots)$$

$$\mu_n(t) = \min(n\mu, m\mu) = \begin{cases} n\mu & 0 \leq n \leq m \\ m\mu & m \leq n \end{cases}$$

und dem folgenden Zustansübergangsdiagramm



Die Ergodizitätsbedingung ist erfüllt, wenn $\frac{\lambda}{m\mu} < 1$ ist.

Die Differentialgleichungen des Geburts-, Todesprozesses für den stationären Fall führen auf die Wahrscheinlichkeiten:

$$p_n = p_0 \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} \quad \text{bzw.} \quad p_n = p_0 \prod_{k=1}^n \frac{\lambda_{k-1}}{\mu_k}$$

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{k=1}^n \frac{\lambda_{k-1}}{\mu_k}}$$

Werden die Merkmale des M/M/m-Systems in diese Gleichungen eingesetzt, dann erhält man für

$$n \leq m: p_n = p_0 \prod_{k=0}^{n-1} \frac{\lambda}{(k+1)\mu} = p_0 \left(\frac{\lambda}{\mu} \right)^n \cdot \frac{1}{n!}$$

$$n > m: p_n = p_0 \prod_{k=0}^{m-1} \frac{\lambda}{(k+1)\mu} \cdot \prod_{k=m}^{n-1} \frac{\lambda}{m\mu} = p_0 \left(\frac{\lambda}{\mu} \right)^n \cdot \frac{1}{m! m^{n-m}}$$

Mit $\rho = \frac{\lambda}{m\mu} < 1$ (Verkehrsintensität je Bedienungsstelle) bzw. mit $\rho = \frac{\lambda}{\mu}$ (Verkehrsintensität) ergibt sich für

$$n \leq m: p_n = p_0 \cdot \frac{(m\rho)^n}{n!} \text{ bzw. } p_n = p_0 \cdot \frac{\rho^n}{n!}$$

$$n \geq m: p_n = p_0 \cdot \frac{\rho^n \cdot m^m}{m!} \text{ bzw. } p_n = p_0 \cdot \frac{\rho^n \cdot m^m}{m! \cdot m^{n-m}}$$

Daraus folgt

$$p(0) = \left[\sum_{n=0}^{m-1} \frac{(m \cdot \rho)^n}{n!} + \frac{(m \cdot \rho)^m}{m!} \cdot \frac{1}{1 - \rho} \right]^{-1} \text{ bzw. } p(0) = \left[\sum_{n=0}^{m-1} \frac{\rho^n}{n!} + \frac{\rho^m}{(m - \rho) \cdot (m - 1)!} \right]^{-1}$$

Mittlere Anzahl von Einheiten in der Schlange (mit $\rho = \frac{\lambda}{\mu}$):

$$L^q = \sum_{n=m+1}^{\infty} (n - m) \cdot p_n = \sum_{n=m+1}^{\infty} n \cdot p_n - m \cdot \sum_{n=m+1}^{\infty} p_n$$

$$L^q = \frac{\rho^{m+1}}{(m - 1)! (m - \rho)^2 \sum_{n=0}^{\infty} \frac{\rho^n}{n!} + (m - \rho) \cdot \rho^m}$$

Mittlere Anzahl von Einheiten im Wartesystem: $L = L^q + \rho$

Anwendungen

Die Anwendung des M/M/1-Systems wird erweitert: Ein zweites Terminal wird in der Zweigstelle aufgestellt und an den zentralen Rechner angeschlossen:

1) Wie groß ist die Wahrscheinlichkeit, dass die beiden Bedienstationen besetzt sind?

$$1 - p_0 \sum_{n=0}^{m-1} \frac{\rho^n}{n!}$$

2) Wie groß ist die Mittlere Wartezeit in der Schlange? $W^q = \frac{L^q}{\lambda}$

5.3.1.1.2 Systeme mit beschränktem Zugang

M/M/1/K

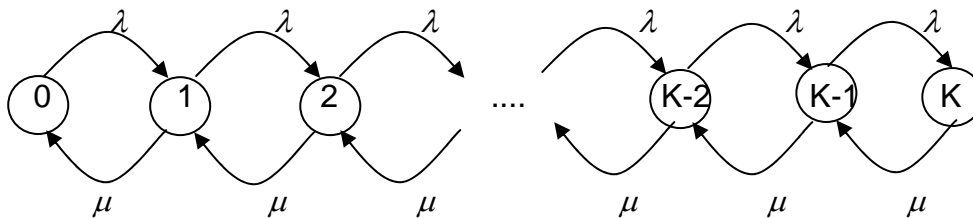
Berechnungsgrundlagen

Im Wartesystem befinden sich höchstens K zu bedienende Einheiten.
Das System ist gekennzeichnet durch

$$\lambda_n(t) = \begin{cases} \lambda & n < K \\ 0 & n \geq K \end{cases}$$

$$\mu_n(t) = \mu \quad (n = 1, 2, \dots)$$

und dem folgenden Zustansübergangsdiagramm



Werden die Bedingungen des vorliegenden Systems in die Lösung eines Geburts-, Todesprozesses im stationären Zustand eingefügt, dann ergibt sich:

$$p_n = p_0 \prod_{i=0}^{n-1} \frac{\lambda}{\mu} \quad (n \leq K)$$

$$p_n = p_0 \left(\frac{\lambda}{\mu} \right)^n \quad (n \leq K)$$

$$p_n = 0 \quad (n > K)$$

Das führt zu

$$p_0 = \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{K+1}}$$

und

$$p_n = \begin{cases} \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{K+1}} & 0 \leq n \leq K \\ 0 & n > K \end{cases}$$

bzw. falls $K = 1$:

$$p_n = \begin{cases} \frac{1}{1 + \frac{\lambda}{\mu}} & n = 0 \\ \frac{\frac{\lambda}{\mu}}{1 + \frac{\lambda}{\mu}} & n = 1 = K \\ 0 & \text{alle _ übrigen _ Fälle} \end{cases}$$

Anwendungen

In einer Schaltstation treffen Nachrichten mit einer mittleren Ankunftsrate von 240 Nachrichten in der Stunde ein. Die Länge der Nachrichten beträgt im Mittel 150 Zeichen. Die Übertragungsgeschwindigkeit auf der danach folgenden Nachrichtenübertragungsstrecke betrage 15 Zeichen/s. Die Übertragungszeit für eine Nachricht ist direkt proportional zur Nachrichtenlänge.

a) Nimm an, ein sehr großer Puffer sei vorhanden. Außerdem sei das System durch ein M/M/1-Modell definiert.

$$\text{Ankunftsrate } \lambda = 240 \frac{\text{Nachrichten}}{h} = \frac{240}{60 \cdot 60} = \frac{1}{15} \text{ Nachrichten/s}$$

$$\text{Bedienzeit } \frac{150}{15} = 10s$$

$$\rho = \frac{\frac{1}{15}}{\frac{1}{10}} = \frac{10}{15} = \frac{2}{3}$$

1. Gib die mittlere Anzahl der Nachrichten in der Schlange an!

—

2. Gib die mittlere Wartezeit der Nachrichten in der Warteschlange an!

—

3. Gib die mittlere Anzahl der Nachrichten im Wartesystem an!

—

4. Gib die mittlere Wartezeit von Nachrichten im System an! $W = \frac{1}{\mu - \lambda}$

b) Nimm an, dass der Puffer ein beschränktes Fassungsvermögen von maximal K (Nachrichten) besitzt. Das Wartesystem ist damit durch ein M/M/1/K-Modell definiert

1. Dimensioniere den Puffer so, dass die Wahrscheinlichkeit, eine ankommende Nachricht kann nicht angenommen werden, unter 5% liegt.

M/M/m/m

Jede neu hinzukommende, zu bedienende Einheit erhält eine Bedienungseinheit zugeteilt. Reicht die Anzahl der Bedienungseinheiten nicht aus, dann wird die ankommende Einheit nicht bedient und geht verloren (Verlustsystem).

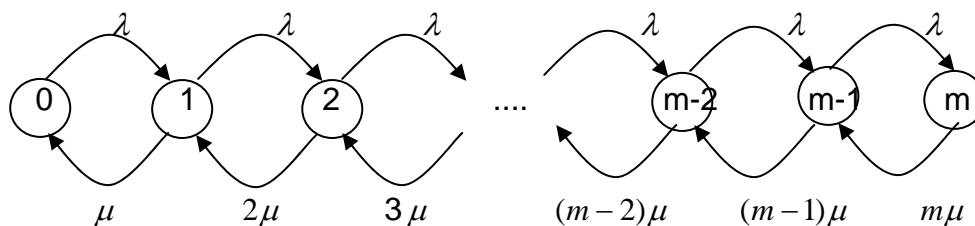
Berechnungsgrundlagen

Das System ist gekennzeichnet durch

$$\lambda_n(t) = \begin{cases} \lambda & n < m \\ 0 & n \geq m \end{cases}$$

$$\mu_n(t) = n\mu \quad (n = 1, 2, \dots)$$

und dem folgenden Zustansübergangsdiagramm



Werden die Bedingungen des vorliegenden Systems in die Lösung eines Geburts-, Todesprozesses im stationären Zustand eingefügt, dann ergibt sich (für $k \leq m$):

$$p_n = p_0 \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} = \prod_{k=0}^{n-1} \frac{\lambda}{(k+1)\mu}$$

Das führt zu:

$$p_k = p_0 \left(\frac{\lambda}{\mu} \right)^k \cdot \frac{1}{k!} \quad k \leq m$$

$$p_k = 0 \quad k > m$$

$$p_0 = \left(\sum_{n=0}^m \left(\frac{\lambda}{\mu} \right)^n \cdot \frac{1}{n!} \right)^{-1}$$

5.3.1.2 Wartezeitenverteilungen

Die Strategien FCFS, LCFS und RANDOM besitzen folgende Wartezeitenverteilungen:

für M/M/m und FCFS (mit Ankunftsintensität λ und Bedienungsintensität μ :

$$P(W \leq t \mid W > 0) = 1 - e^{-(m\mu - \lambda)t}$$

für M/M/n und LCFS:¹⁶

für M/M/m und RANDOM:

5.3.2 Allgemeine Bedienungssysteme

Bei allgemeinen Bedienungszeiten können die Zwischenankunftszeiten beliebig verteilt und auch voneinander abhängig sein. Ebenso können die geforderten Bedienzeiten beliebig verteilt sein.

¹⁶ vgl. bzw. siehe Pflug, Georg: Stochastische Modelle in der Informatik, Teubner-Verlag, S. 68

5.4 Warteschlangennetze

5.4.1 Grundlagen

Einführung

Zustandswahrscheinlichkeiten eines Warteschlangenmodells können durch Lösung des Systems der globalen Gleichgewichtsbedingungen exakt bestimmt werden. Allerdings kann die Menge der zu lösenden Gleichungen bereits bei Netzen mit einer geringen Zahl von Aufträgen und Knoten groß werden. Numerische Lösungsverfahren sind deshalb nur bei kleinen Netzen vorteilhaft verwendbar.

Für **Warteschlangennetze** existieren eine Vielzahl von Verfahren, die eine exakte Bestimmung der Leistungsgrößen unter Umgehung der globalen Wahrscheinlichkeiten ermöglichen. Erfüllen sämtliche Knoten des Warteschlangennetzes bestimmte Voraussetzungen zur Verteilung der Zwischenankunfts- und Bedienzeiten (sowie zur Warteschlangendisziplin), dann lassen sich für das Systemverhalten auf eindeutige Weise sogenannte lokale Gleichgewichtsbedingungen angeben.

Warteschlangennetze, die eine eindeutige Lösung der lokalen Gleichgewichtsbedingungen besitzen, heißen **Produktformnetze**. Solche Netze sind dadurch charakterisiert, daß sich die Lösungen für die Gleichgewichtswahrscheinlichkeiten multiplikativ aus Fakten zusammensetzen, die sich auf die Zustände der einzelnen Knoten beziehen. Die Warteschlangendisziplin ist FCFS.

Produktformnetze unterscheiden offene und geschlossene Zwischenankunfts- und Bedienzeiten.

Offene, geschlossene und gemischte Warteschlangennetze

Ein Warteschlangensystem ist offen, wenn Aufträge von außerhalb des Netzes ankommen und Aufträge dieses Netz auch verlassen können.

Bsp.: Offenes Warteschlangennetz eines Rechensystems

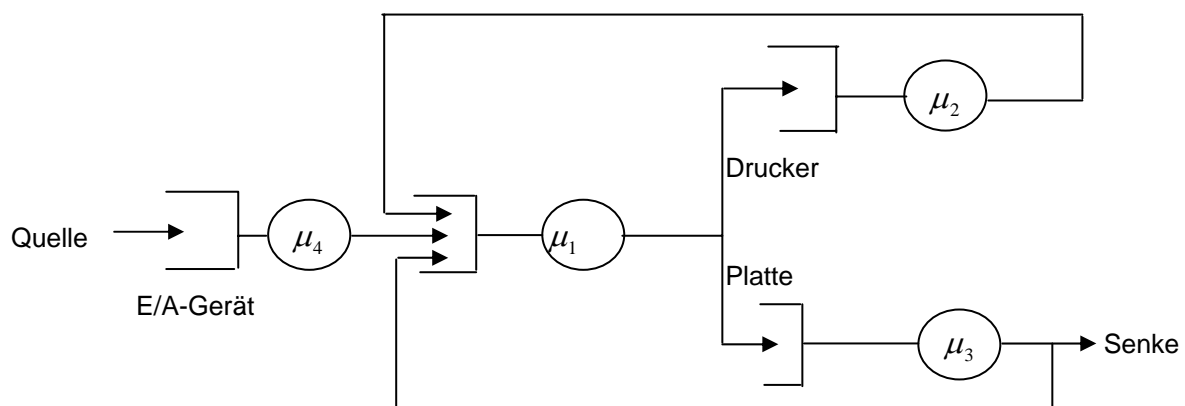


Abb.:

Ein Warteschlangennetz heißt geschlossen, wenn keine externe Auftragsankünfte und Auftragsabgänge möglich sind. Die Anzahl der Aufträge im geschlossenen Netz ist stets konstant. Man kann ein geschlossenes Netz aber auch so interpretieren, daß für einen im System verlassenden Auftrag unmittelbar ein neuer Auftrag nachrückt.

Bsp.: Geschlossenes Warteschlangennetz für ein Rechensystem (Central-Server-Modell)

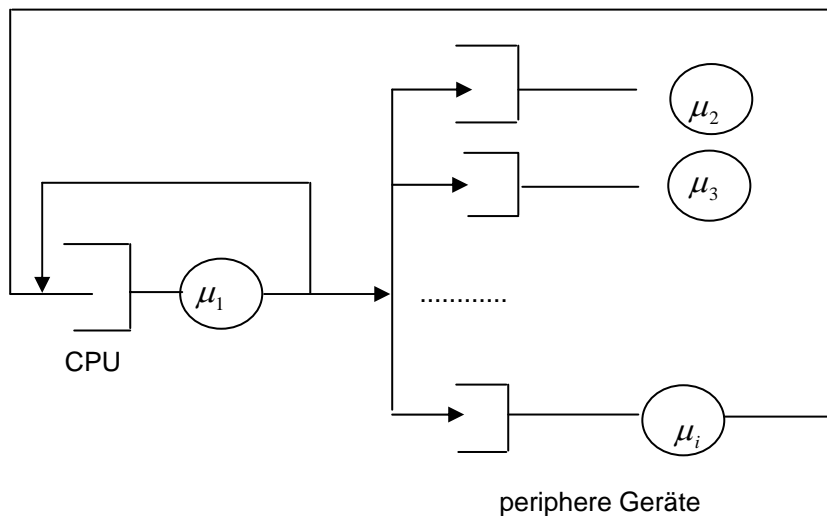


Abb.: Central-Server-Modell

Mit geschlossenen Netzen kann das Verhalten eines Rechensystems häufig besser modelliert werden (insbesondere Multiprogrammingsysteme mit festem Multiprogrammingsgrad oder interaktive Systeme mit einer geringen Anzahl von Terminals). Offene Modelle werden oft für Rechnernetze und interaktive Systeme mit vielen Terminals verwendet.

Möglich ist die Einführung verschiedener Auftragsklassen im Netz, die den unterschiedlichen Programmklassen in realen System entsprechen. Die Auftragsklassen unterscheiden sich durch unterschiedliche Bedienzeiten der einzelnen Knoten (unterschiedliche Übergangswahrscheinlichkeiten).

Erhält ein Warteschlangenmodell sowohl offene als auch geschlossene Klassen, dann wird es gemischt genannt.

Formale Beschreibung von Warteschlangennetzen

Bezeichnung Erläuterung:

N gibt die Anzahl der Knoten im Netz an

K bezeichnet in geschlossenen Netzen die konstante Anzahl der Aufträge im Netz

k_i bezeichnet die Anzahl der Aufträge im i -ten Knoten. Für geschlossenen

Netze gilt: $\sum_{i=1}^N k_i = K$

m_i ist die Anzahl der Bedieneinheiten des i -ten Knoten

μ_i ist die mittlere Bedienrate von Aufträgen im Knoten i

$1/\mu_i$ ist die mittlere Bedienzeit eines Auftrags im Knoten i

p_{ij} Ist die Wahrscheinlichkeit (WS), daß ein im Knoten i fertig bedienter Auftrag zum Knoten j wechselt. (0 steht für die Aussenwelt bei offenen Netzen).

$p_{i0} = 1 - \sum_{j=1}^N p_{ij}$ ist die Wahrscheinlichkeit, daß ein Auftrag nach Abfertigung durch Knoten i das Netz anschließend verläßt.

λ_{0i} ist die mittlere Ankunftsrate von außen beim i -ten Knoten

λ_i ist die gesamte mittlere Ankunftsrate von Aufträgen beim Knoten i .

Zur Berechnung der mittleren Ankunftsrate λ_i bei den Knoten $i = 1, \dots, N$ müssen im offenen Netz alle Ankünfte von außerhalb des Netzes und Ankünfte von allen internen Knoten summiert werden. Da im statischen Gleichgewicht die mittlere Ankunftsrate an einem Knoten dem mittleren Abgangsgrad aus diesem Knoten ist, erhält man

$$\lambda_i = \lambda_{0i} + \sum_{j=1}^N \lambda_j \cdot p_{ji} \quad \text{für } i = 1, \dots, N$$

Dieses Gleichungssystem bekommt für geschlossene Netze die Form:

$$\lambda_i = \sum_{j=1}^N \lambda_j \cdot p_{ji}$$

Die mittlere Besuchshäufigkeit (oder relative Ankunftsrate) ist: $b_i = \frac{\lambda_i}{\lambda}$

λ ist der Gesamtdurchsatz des Netzes.

5.4.2 Produktformlösungen

5.4.2.1 Jackson-Theorem für offene Netze

Wenn für alle Knoten $i=1, \dots, N$ im offenen Netz die Stabilitätsbedingung $\lambda_i < \mu_i \cdot m_i$

erfüllt ist, wobei sich die Ankunftsrate λ_i aus $\lambda_i = \lambda_{0i} + \sum_{j=1}^N \lambda_j \cdot p_{ji}$ ergeben, dann ist

die Wahrscheinlichkeit für den Gleichgewichtszustand des netzes durch das Produkt der Zustandswahrscheinlichkeiten (Randwahrscheinlichkeiten) der einzelnen Knoten gegeben: $p(k_1, k_2, \dots, k_N) = p_1(k_1) \cdot p_2(k_2) \cdot \dots \cdot p_N(k_N)$. Die einzelnen Knoten des Netzes können somit als voneinander unabhängige elementare M/M/m-FCFS-Wartesysteme mit der Ankunftsrate λ_i und der Bedienrate μ_i angesehen werden.

5.4.2.2 Gordon/Newell-Theorem für geschlossene Netze

Das Theorem von Gordon/Newell besagt, daß die Wahrscheinlichkeiten für die einzelnen Netzwerkzustände im Gleichgewicht durch folgenden Produktformsatz bestimmt werden könnten.

$$p(k_1, k_2, \dots, k_N) = \frac{1}{G(K)} \cdot \prod_{i=1}^N F_i(k_i)$$

$G(K)$ bezeichnet die Normalisierungskonstante des Netzes. Sie ergibt sich aus der Bedingung, daß sich die Wahrscheinlichkeiten aller Netzwerkzustände zu Eins summieren müssen

$$G(K) = \sum_{\sum_{i=1}^N k_i = K} \prod_{i=1}^N F_i(k_i)$$

Die $F_i(k_i)$ sind Funktionen, die den Zustandswahrscheinlichkeiten $p_i(k_i)$ des i-ten Knoten entsprechen. Es gilt

$$F_i(k_i) = \left(\frac{e_i}{\mu_i} \right)^{k_i} \cdot \frac{1}{\beta_i(k_i)}$$

Die Besuchshäufigkeiten werden aus $e_i = \sum_{j=1}^N e_j \cdot p_{ji}$ (für $i=1, \dots, N$) berechnet. Die Funktion $\beta_i(k_i)$ erhält man aus der Formel

$$\beta_i(k_i) = \begin{cases} k_i! & k_i \leq m_i \\ m_i! m_i^{k_i - m_i} & k_i \geq m_i \\ 1 & m_i = 1 \end{cases}$$

6. Simulation

6.1 Definition und Klassifizierung von Simulationsmethoden

6.1.1 Definition

„Wenn man überhaupt nicht mehr weiter kann, fängt man zu simulieren an“. Dieser Ausspruch kennzeichnet die Rolle der Simulationsverfahren. Man kann sie anwenden, wenn folgende Probleme auftreten:

- Der vorliegende Sachverhalt läßt sich überhaupt nicht oder nur schwer durch mathematische Ausdrücke beschreiben.
- Es gibt überhaupt kein oder kein einfach zu handhabendes analytisches Lösungsverfahren.
- Die Anwendung eines exakten Optimierungsverfahrens besitzt zu hohen Rechenaufwand oder zu großen Speicherbedarf

Simulationsmethoden bilden die zu untersuchende Situation unter einer Reihe konkreter Annahmen rechnerisch nach. Durch Variation dieser Voraussetzungen und wiederholte Anwendung des Verfahrens erhält man eine Menge quantitativer Aussagen. Daraus kann man Schlüsse ziehen.

Simulation bedeutet Experimentieren mit (OR-) Modellen. Berechnungsexperimente mit mathematischen Modellen dynamischer Systeme auf digitalen Rechnern werden als **digitale Simulation** bezeichnet. Damit ergibt sich die folgende Definition:

Simulation ist ein Verfahren zur Durchführung von Experimenten auf einem Digitalrechner unter Benutzung mathematischer Modelle mit dem Ziel, Aussagen über das Verhalten des realen Systems zu gewinnen.

Aus dieser Definition sind 2 wichtige Punkte abzuleiten:

1. Bei der Simulation handelt es sich nicht um eine feststehende Methode. Im Unterschied zu einem linearen Optimierungsmodell, das den relevanten Sachverhalt komplett und geschlossen darstellt, beschreibt ein Simulationsmodell im allg. den Wirkungsmechanismus eines Systems durch Abbildung einzelner Komponenten und Erfassung wechselseitiger Abhängigkeiten. Die Simulation dient der Vorhersage der Zustände von einzelnen Komponenten und des Gesamtsystems, wobei diese (End-) Zustände meistens von vielen Einflußfaktoren in Form von Wahrscheinlichkeitsverteilungen abhängen.
2. Bei der Simulation handelt es sich um eine Vorgehensweise, die dem Studium des Systemverhaltens dient, nicht aber zu einer Systemoptimierung führt. Man kann nur durch Experimentieren in der entsprechenden Richtung zu Optimallösungen gelangen.

Was genau will man mit einer Simulation erreichen?

- Dem Wissenschaftler geht es vor allem um das Überprüfen von Theorien: Er baut ein Modell nach den Vorgaben der Theorie und untersucht, ob die Vorhersagen des Modells mit der Wirklichkeit zusammenpassen-
- Der Betriebswirtschaftler hingegen sucht nach einer optimalen unternehmerischen Entscheidung durch experimentelles Durchspielen von Alternativen
- Der Ingenieur benötigt die Unterstützung beim Entwurf und der Optimierung technischer Systeme.
- Weitere Simulationszwecke sind: die Wettervorhersage, Prognosen zur Entwicklung der Umwelt oder des Marktes, das computer-basierte Training (z.B. Flugsimulator)

6.1.2 Klassifizierung

Zusammenstellung wichtiger Klassifizierungsmerkmale

(1) Einteilung nach Art des Experimentierens

Nach diesem Merkmal unterscheidet man 2 Simulationsarten:

- Monte-Carlo Simulation oder zufallsbedingte Simulation
- gezielte oder geplante Simulation

Bei der Monte-Carlo Simulation werden die den Experimenten zugrundeliegenden Bedingungen oder zumindest ein Teil von ihnen zufällig ausgewählt, bei der gezielten Simulation folgen sie einem exakt vorbestimmten Plan.

(2) Einteilung nach der Art der angestrebten Lösung

Das Verhalten von Systemen kann durch statische oder dynamische Lösungen charakterisiert werden. Eine statische Lösung ist durch einen mathematischen Zusammenhang gekennzeichnet, der objektiv gültig ist und über Experimente aufgedeckt wird, z.B. die optimale Auslegung und Fahrweise einer technischen Anlage in Abhängigkeit von bestimmten Parametern. Bei einer dynamischen Lösung will man das Verhalten des Systems in Abhängigkeit von der Zeit kennenlernen.

(3) Einteilung nach den Zeitpunkten der Berechnung

Diese Einteilung bezieht sich auf die Simulation mit dynamischer Lösung. Während bei einer statischen Lösung die Wahl des Zeitpunkts, wann irgend etwas gerechnet wird, völlig beliebig und ohne Einfluß auf das Ergebnis ist, kann man bei dynamischen Lösungen das Verhalten des Systems für beliebige Zeitpunkte im voraus berechnen oder nur dann gewisse Berechnungen durchführen, wenn das System durch äußere Einflüsse gestört wird. Man kann unterteilen in

- zeitunabhängige Simulation
- Simulation in diskreten Zeitpunkten (discret event simulation)

Bei der diskreten Ereignis-Simulation wird der Zustand eines dynamischen Systems durch zeitabhängige Zustandsvariable beschrieben. Die Zustandsvariablen ändern sich ggf. durch den Eintritt von Ereignissen an bestimmten Zeitpunkten. Je nach Vorgabe bzw. Ermittlung der diskreten Zeitpunkte (mit Hilfe einer sog. Simulationsuhr), an denen sich u.U. Zustandsvariable ändern, lassen sich 2 Arten der diskreten Ereignis-Simulation unterscheiden:

- Periodenorientierte Zeitführung (fixed-increment time advance).

Hier wird die Simulationsuhrzeit jeweils um Δt Zeiteinheiten [ZE] erhöht. Δt ist dabei je nach Problemstellung geeignet zu wählen (Minute, Stunde, Tag, etc.). Nach der Aktualisierung der Simulationsuhrzeit wird überprüft, ob irgendwelche Ereignisse während Δt eingetreten sind, die zu einer Veränderung der Zustandsvariablen führen.

- Ereignisorientierte Zeitführung (next-event time advance)

Nach der Initialisierung der Simulationsuhrzeit mit dem Wert 0 ermittelt man die Zeitpunkte, an der zukünftige Ereignisse eintreten. Anschließend wird die Simulationsuhrzeit mit dem Zeitpunkt des Eintritts des (zeitlich) ersten Ereignisses gleichgesetzt und die zugehörige Zustandsvariable aktualisiert. Der Prozeß ist solange fortzusetzen, bis eine bestimmte Abbruchbedingung eintritt.

6.2 Simulationsstudie

6.2.1 Fallbeispiel

17

In einem Betrieb mit mechanischer Teilefertigung wird ein Gußteil mehrmals im Jahr in einer Losgröße von je 500 Stück bearbeitet. Dabei bildet eine Maschinengruppe, die aus miteinander verbundenen Maschinen 1 und 2 besteht eine Kapazitätsengpaß.

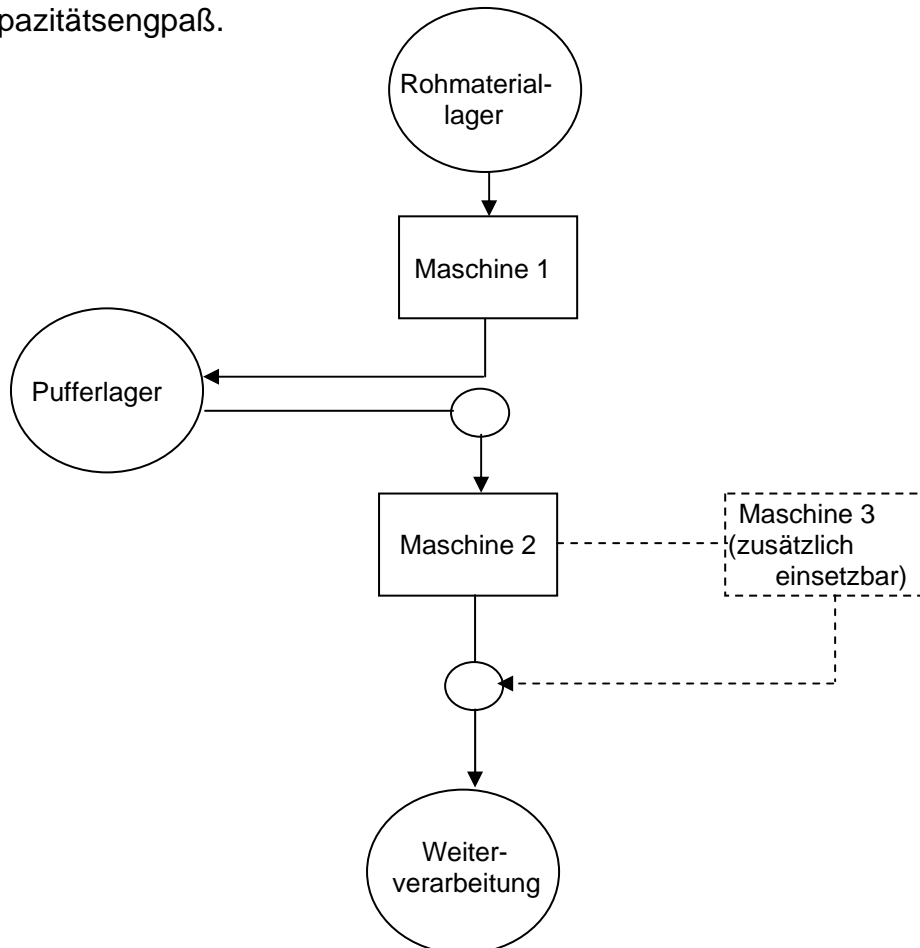


Abb.: Anordnung der Maschinengruppe und Durchlauf der Gußteile

Die Gußteile werden direkt von dem Rohmateriallager bezogen, in dem sie in ausreichender Stückzahl vorhanden sind. Jedes Teil muß zuerst die Maschine 1 durchlaufen. Anschließend gelangen die Teile in ein Pufferlager. Ist die Maschine 2 frei, dann werden sie auf dieser bearbeitet.

In der Vergangenheit erwies sich die Abstimmung zwischen den beiden Maschinen als nicht besonders gut. Gründe dafür waren die jeweils unterschiedliche Bearbeitungsdauer der Gußteile auf den beiden Maschinen, die von der Güte des Gußrohlings abhängige Unterbrechungshäufigkeit der Bearbeitungsvorgänge auf den Maschinen 1 und 2 sowie die unterschiedliche Unterbrechungsdauer.

Die mangelhafte Abstimmung führte zu langen Durchlaufzeiten in dieser Maschinengruppe.

In einer anderen Werkstatt ist eine ältere Maschine durch eine neue ersetzt worden. Die Betriebsführung steht vor der Entscheidung, diese alte Maschine 3 zu verkaufen oder sie zur Entlastung der Maschine 2 einzusetzen. Die auf Maschine 1 begonnene

¹⁷ Übernommen aus: Repetitorium: Angewandte Mathematik für den Industrial Engineer, Teil 1: Simulation in FB/IE 27/1978 Heft 1

Bearbeitung könnte dann auf dieser Maschine fortgesetzt werden, wenn die Maschine 2 belegt ist. Da die Maschine 3 wesentlich längere Bearbeitungszeiten hat als die Maschine 2, sollte die Bearbeitung des Gußteile auf Maschine 3 nur dann vorgenommen werden, wenn die Maschine 2 bereits belegt ist.

Die Betriebsleitung benötigt für ihre Entscheidung Informationen darüber, ob sich der zusätzliche Einsatz der Maschine 3 aufgrund einer Reduzierung der Durchlaufzeit für die einzelnen Lose vertreten läßt.

	Bearbeitungsdauer			Unterbrechung		Unterbrechungsdauer		
	Häufig- keit	Dauer (min)	Zug. Würfelz.	Häufigkeit	Zugeordn. Würfelzahl	Häufig- keit	Dauer (min)	Zug. Würfelz.
Maschine 1	1/6	3	1	1/6	6	2/6	2	1,2
	4/6	4	2,3,4,5			3/6	3	3,4,5
	1/6	5	6			1/6	4	6
Maschine 2	1/6	4	1	3/6	4,5,6	1/6	1	1
	4/6	5	2,3,4,5			3/6	2	2,3,4
	1/6	6	6			2/6	3	5,6
Maschine 3	1/6	6	1	1/6	6	3/6	2	1,2,3
	4/6	7	2,3,4,5			3/6	3	4,5,6
	1/6	8	6					

Abb.: Daten des Fallbeispiels

6.2.2 Manuelle Simulation

Der Grundgedanke der Simulation besteht darin „Prozesse der Realität durchzuspielen“. Im Fallbeispiel muß man bspw. die Bearbeitung der 500 Gußteile in der Maschinengruppe simulieren. Dabei muß für jedes Gußteil festgelegt werden:

- Wie lange dauert die Bearbeitung auf der Maschine 1?
- Kommt es zu einer Unterbrechung im Laufe der Bearbeitung?
- Wie lange dauert diese Unterbrechung gegebenenfalls?

Danach wird untersucht, ob das Gußteil die Maschine 2 oder die Maschine 3 belegt bzw. ob es in das Pufferlager gelangt. Für die weitere Bearbeitung ist es nötig, im Fall der Unterbrechung die Unterbrechungsdauer auf Maschine 2 bzw. 3 zu ermitteln. Im Fallbeispiel sind auftretende Häufigkeiten der Bearbeitungsdauer, des Unterbrechungsfalls und der Unterbrechungsdauer Vielfache eines Sechstels. Es ist daher möglich den Unterbrechungsfall und auch seine Dauer durch „Würfeln“ zu bestimmen, da jede Würfelzahl mit einer Wahrscheinlichkeit von einem Sechstel auftritt.

1	2	3	4, 5	6, 7	8, 9	10	11	12	13
Anfangszeitpunkt	Nr. des Gußteils	Nr. Maschine	Bearbeitung W. ¹⁸ D. ¹⁹	Unterbrechungsfall W. ²⁰ j/n	Unterbrechung W. ²¹ D. ²²	Belegungs-dauer ²³	Zeitp. des Beleg. ²⁴	Nr. Gußt. im Puffl.	Nr. bearb. Gußt.
0	1	1	6 5	5 nein	- -	5	5	-	-
5	1 2	2 1	6 6 3 4	5 ja 6 ja	2 2 3 3	8 7	13 12	- -	1
12	2 3	3 1	1 6 5 4	6 ja 4 nein	1 2 - -	8 4	20 16	- -	1, 2
16	3 4	2 1	4 5 3 4	5 ja 3 nein	2 2 - -	7 4	23 20	- -	1-3
20	4 5	3 1	6 8 1 3	2 nein 4 nein	- - - -	8 3	28 23	- -	1-4
23	5 6	2 1	2 5 1 3	2 nein 2 nein	- - - -	5 3	28 26	- 6	1-5
26	7	1	2 4	5 nein	- -	4	30	-	
28	6	2	6 6	4 ja	5 3	9	37	-	1-6
30	7 8	3 1	2 7 2 4	2 nein 3 nein	- - - -	7 4	37 34	- 8	1-7
34	9	1	5 4	4 nein	- -	4	38	-	
37	8	2	2 5	5 ja	4 2	7	44	-	1-8
38	9 10	3 1	2 7 5 4	3 nein 4 nein	- - - -	7 4	45 42	- 10	1-9
42	11	1	6 5	6 ja	2 2	7	49	-	
44	10	2	1 4	1 nein	- -	4	48	-	1-10
49	11 12	2 1	4 5 6 5	4 ja 1 nein	5 3 - -	8 5	57 54	- -	1-11
54	12 13	3 1	4 7 3 4	1 nein 4 nein	- - - -	7 4	1.01 58	- -	1-12
58	13 14	2 1	2 5 1 3	6 ja 3 nein	3 2 - -	7 3	1.05 1.01	- -	1-13

Abb.: Manuelle Simulation des Fallbeispiels

Die folgende Darstellung zeigt die manuelle Simulation während der ersten Stunde:

¹⁸ Würfelzahl

¹⁹ Dauer (min)

²⁰ Würfelzahl

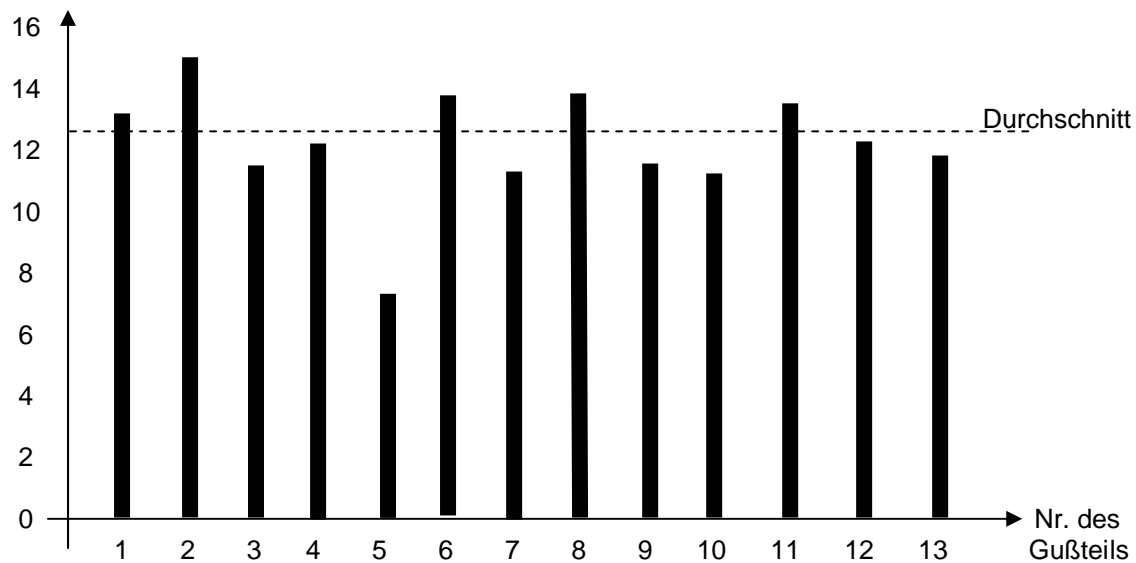
²¹ Würfelzahl

²² Dauer (min)

²³ Bearbeitung (Bearbeitung und Unterbrechung in Min.)

²⁴ Zeitpunkt des Belegungsendes

Durchlaufzeiten der Gußteile (Min)



Nach 60 Minuten sind 11 Gußteile bearbeitet. Die Durchlaufzeiten in dieser Maschinengruppe liegen zwischen 8 und 15 Minuten, bei einem Durchschnitt von rund 12 Minuten.

Zur Bestimmung der Bearbeitungsdauer für ein Los Gußteile, muß die Simulation für alle 500 Gußteile durchgeführt werden. Das wird zweckmäßigerweise mit einem Rechnerprogramm erledigt. Die Simulation dieses Beispiels mit einem Rechnerprogramm zeigt, daß sich dem Einsatz der Maschine 3 die Durchlaufzeit eines Loses von rund 49 Stunden auf rund 37.5 Stunden senken läßt.

6.2.3 Allgemeine Vorgehensweise bei einer Simulationsstudie

Eine Simulationsstudie kann in folgende Arbeitsschritte gegliedert werden:

- 1) Problemformulierung
- 2) Problemanalyse
- 3) Modellbildung
- 4) Validierung des Modells
- 5) Umsetzung der Simulationsergebnisse in die Realität

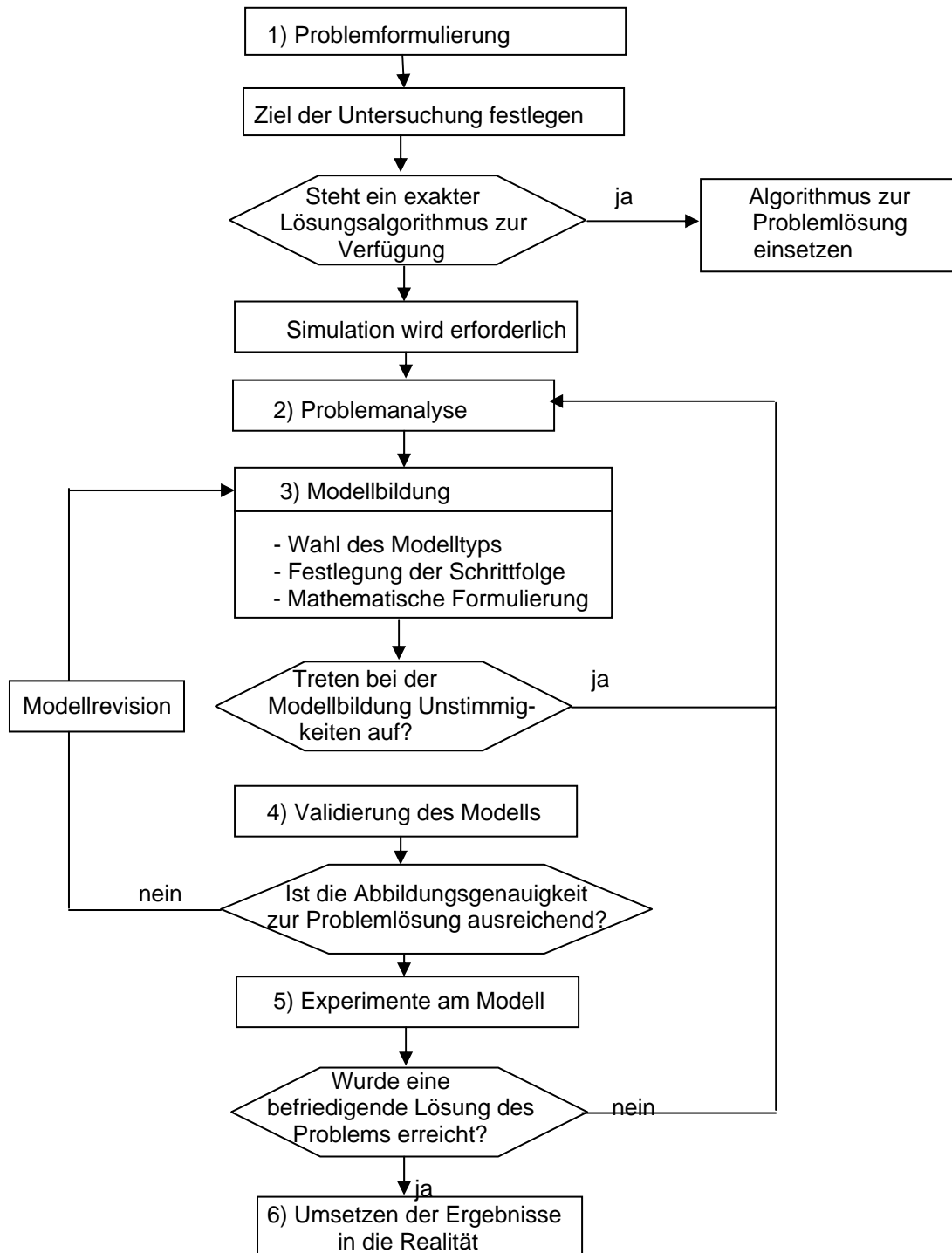


Abb.: Ablauf der Simulationsstudie

6.2.4 Werkzeuge der Simulation

Die Implementierung eines Simulationsmodells als Programm ist Voraussetzung für eine Simulationsstudie. Das Programm kann in einer Universalsprache oder Simulationssprache implementiert werden. Eine Modellierungsphilosophie folgt entweder der Ereignisorientierung oder der Prozessorientierung oder sie unterstützt beide Ansätze.

Bsp. für Simulationssprachen sind:

- GPSS (allgemeines Konzept zur Simulation diskreter und kontinuierlicher Systeme)
- SIMSCRIPT (eignet sich für die Simulation diskreter Systeme)
- SIMULA (für die Simulation diskreter Systeme)

6.3 Modelltypen der digitalen Simulation

6.3.1 Determinierte Modelle

Ein Modell heißt determiniert, wenn durch Eingabedaten und Anfangszustände der Zustandsvariablen bereits eindeutig die Abläufe im Modell und die Ergebnisse festgelegt sind.

Bsp.: Erneuerungsproblem

Aufgabenstellung:

Die Ausbringung eines chemischen Prozesses hängt stark von der verwendeten Lauge ab. Wenn die Lauge nicht erneuert wird, nimmt die Ausbringung täglich um 5% ab, jeweils bezogen auf die Ausbringung des Vortages. Wenn die Lauge erneuert wird, fällt die Produktion einen Tag lang aus. Nach Erneuerung der Lauge wird wieder die volle Ausbringung erreicht. Nach wieviel Tagen soll die Lauge erneuert werden, wenn die durchschnittliche Ausbringung möglichst groß sein soll?

Mathematische Formulierung der Aufgabe:

Ausbringung am Tag t : q^{t-1} für $t=1,2,\dots$ ($q=1-0.05=0.95$)

Ausbringung bis zum Tag T : $\sum_{t=1}^T q^{t-1} = \frac{1-q^T}{1-q}$

Durchschnittliche Ausbringung in einer Periode, d.h. in $T+1$ Tagen: $f(T) = \frac{1-q^T}{1-q}$

Im Maximum muß $f'(T) = \frac{1}{1-q} \cdot \frac{-(T+1) \cdot \ln q - (1-q^T)}{(T+1)^2} = 0$ sein. Daraus ergibt sich für T_{opt}

die transzendente Gleichung: $T = \frac{1}{-\ln q} \cdot \left(\frac{1}{q^T} - 1 \right) - 1$ mit der Lösung $T_{\text{opt}} = 5.93$.

Nachbildung des Prozesses durch eine determinierte Simulation

Man tut so, als würde man die Erneuerung der Lauge nach ein, zwei, drei usw. Tage vornehmen. Man berechnet und vergleicht die zugehörigen durchschnittlichen Ausprägungen. Wenn die durchschnittliche Ausbringung abnimmt ist darauf zu achten, daß eine Produktionsperiode aus den Tagen besteht, an denen tatsächlich produziert wird, zuzüglich dem einen Tag, an dem die Lauge

erneuert wird. Es ist also: $\text{durchschnittliche_Ausbringung} = \frac{\text{Summe_der_Ausbringungen}}{\text{Produktionstage} + 1}$

Tag t	Ausbringung am Tag t	Ausbringung bis Tag T	Durchschnittliche Ausbringung
1	1.0000	1.0000	0.500
2	0.9500	1.9500	0.6500
3	0.9025	2.8525	0.7131
4	0.8574	3.7099	0.7420
5	0.8145	4.5244	0.7541
6	0.7738	5.2982	0.7569
7	0.7351	6.0333	0.7542
8	0.6983	6.7316	0.7480

Abb.: Beispiel zu determinierten Simulation

Die Erneuerung der Lauge wird zweckmäßigerweise nach jeweils 6 Produktionstagen vorgenommen.

6.3.2 Stochastische Modelle

6.3.2.1 Beschreibung und Vorgehensweise

Die stochastische Simulation (auch simuliertes Stichprobenverfahren genannt) wird dann eingesetzt, wenn gewisse Parameter eines Modells als auch der Eingangsgrößen nicht fest vorgegeben, sondern zufallsbedingt sind.

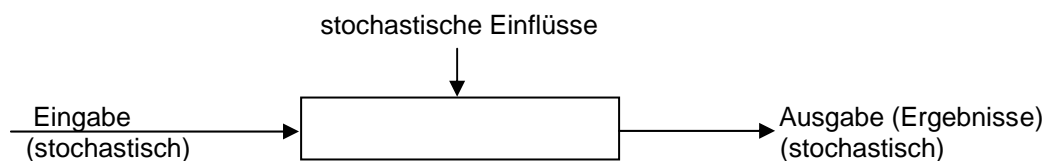


Abb.: Stochastische Modelle

Der Zufall kann an zwei Stellen des Systems eingreifen:

- Die Parameterwerte des Systems streuen
- Die Eingangsgrößen sind zufällige (stochastische) Prozesse

Sowohl bei stochastischen Parametern als auch bei stochastischen Eingangsfunktionen ergeben sich für die Zeitfunktionen der Ausgangsgrößen des Systems zufällige Zeitfunktionen²⁵. Die Gesamtheit (das Ensemble) der Zeitfunktionen bildet einen stochastischen Prozeß $X(t)$. Die Aufgabe kann die Bestimmung der statischen Eigenschaften der Ausgangsgröße $Y(t)$ sein. Dazu erzeugt man eine Stichprobe der Ausgangsgröße, indem man eine Reihe von Realisierungen für die stochastischen Eingangsgrößen bestimmt und mit jeder der Realisierungen einen Simulationslauf durchführt.

Die Stichprobe der Ausgangsgröße wird ausgewertet, indem man bspw. Schätzwerte für den Mittelwert und die Standardabweichung oder die Häufigkeitsverteilung der Ausgangsgröße zu jedem Zeitpunkt t ermittelt. Im einfachsten Fall sind nur statistische Kenngrößen der Ausgangsfunktion gefragt,

²⁵ Unter einer zufälligen Zeitfunktion versteht man eine Zeitfunktion $x(t)$ aus einer Schar möglicher Realisierungen.

6.3.2.2 Stochastische Methoden in der Simulation diskreter Systeme

Aufgabenstellung²⁶: Ein Sessellift kann alle 30s zwei Personen befördern. Die Bedienrate ist demnach $M = 2$. Die Wartekapazität K ist ebenfalls 2.

Wie groß ist die Warteschlangenlänge?

Wie hoch ist die Auslastung des Lifts?

1. Modellannahmen und mathematische Beschreibung der Kenngrößen

Modellannahmen

$X(n)$... zufällige Anzahl von Personen in der Warteschlange vor dem Lift zur Zeit $n\Delta t$

$A(n)$... zufällige Anzahl von Personen, die im Intervall $[n\Delta t, (n+1)\Delta t]$

$M(n)$... zufällige Anzahl von Personen, die zur Zeit $n\Delta t$ (im Takt n) den Lift belegen

$\{X(n), n = 0, 1, 2, \dots\}$... stochastischer Prozeß (Kette) mit dem Zustandsraum (Wertebereich) $E = \{0, 1, 2, 3, \dots, K, K^*\}$. Im Bsp. hat $X(n)$ die Zustände: 0, 1, 2, 2*

K^* ... Länge der Warteschlange ist K und zwischen n und $n+1$ ist ein Überlauf passiert.

Kenngrößen:

$p_k(n) = P(X(n)), k = 0, 1, 2, \dots, K^* (k \in E)$ beschreibt die Warteschlangenlängenverteilung im Takt n .

Die Warteschlange wird durch folgende Kenngrößen charakterisiert:

- mittlere Warteschlangenlänge im Takt n : $E(X(n)) = \sum_{k \in E} kP(X(n) = k)$
- Verlustwahrscheinlichkeit im Takt n : $p_{K^*}(n) = P(X(n) = K^*)$

Der Lift wird durch folgende Kenngrößen charakterisiert:

- mittlere Anzahl besetzter Liftplätze im Takt: $E(M(n)) = \sum_{k=0}^{M-1} kP(M(n) = k)$
- Auslastung des Lifts im Takt n : $U(n) = \frac{E(M(n))}{M}$

Zur Lösung der Aufgabe gibt es grundsätzlich 2 Methoden:

- Simulation des Systemverhaltens und Schätzung der Warteschlangenlängenverteilung $p_k(n)$ und der Systemgrößen anhand der simulierten Daten.
- Mathematische Berechnung der Warteschlangenverteilung und der Systemkenngößen mit Methoden der Wahrscheinlichkeitsverteilung

2. Lösung des Problems durch Simulation

Nachbilden des Systems im Zeitraum $[0, T\Delta t (n = 0, 1, \dots, T)]$ auf dem Rechner und Schätzung der gewünschten Wahrscheinlichkeit $p_k(n)$ durch die relative Häufigkeit bzw. Schätzung der Systemkenngößen

²⁶ vgl. 5.1.2.4

Die relative Häufigkeit ergibt sich aus Anzahl der Simulationsläufe, bei den im Takt n die Warteschlangenlänge gleich k ist, dividiert durch die Anzahl N der Simulationsläufe.

Über ein Rechnerprogramm wurde für $N = 100$ und $n = 3$ ermittelt:

```

C:\WINNT\system32\cmd.exe
Microsoft Windows 2000 [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.

D:\dok\or\ss05\progr\pr63310>java LiftSimul
Warteschlangenlaengenverteilung:
P<X(3)=0> = 0,12
P<X(3)=1> = 0,48
P<X(3)=2> = 0,31
P<X(3)= 2*> = 0,09
mittlere Warteschlangenlaenge: 1,28
Liftauslastung :0,64

D:\dok\or\ss05\progr\pr63310>

```

Abb.:

Experimentelle Untersuchung des Systemverhaltens

Die Simulation ermöglicht Untersuchungen, ob aus der Anfangsphase eine stabile Phase abgeleitet werden kann bzw. ob Stationarität bzw. Ergodizität gegeben ist. So kann über Simulationsstudien ($N = 10000$) für verschiedene Anfangszustände des Systems für $n = 0, 1, 2, 3, \dots, 50$ gezeigt werden:

- Das Systemverhalten ist ab $n \approx 40$ stationär, d.h. es ist $p_k(n) \approx p_k^*$
- $\lim_{n \rightarrow \infty} p_k(n) = p_k^*$ ist unabhängig von der Anfangsverteilung
- $p_k^*, k \in E$ ist eine ergodische Verteilung. Das System nennt man in diesem Fall ergodisch²⁷ bzw. stabil.

3. Lösung des Problems mit Hilfe der Wahrscheinlichkeitsverteilung, ergodische Markovketten

Die Lösung erfolgt durch Modellierung der zufälligen Warteschlangenlänge als homogene Markovkette mit Berechnung der Übergangswahrscheinlichkeiten. Nach Überprüfung, ob ergodische Markovketten vorliegen, kann die Berechnung der ergodischen Wahrscheinlichkeiten p_k^* und der Systemgrößen in der stabilen Phase erfolgen

4. Vergleich: Simulation und Wahrscheinlichkeitsrechnung

Simulation	Wahrscheinlichkeitsrechnung
Einfach, Grundkenntnisse der Statistik reichen aus	Theorie relativ schwierig, Vereinfachende Modellannahmen nötig
Systemverhalten nur mit großem Aufwand und ungenau abschätzbar	Systemverhalten schnell und genau berechenbar.

²⁷ vgl. 5.2.1.4

6.3.2.3 Die Erzeugung von Zufallszahlen

Zur Lösung stochastischer Simulationsprobleme und auch bei der Anwendung von Monte-Carlo-Methoden werden Zufallszahlen benötigt. Die Erzeugung von Zufallszahlen geschieht mit Hilfe von Zufallszahlengeneratoren, an die u.a. folgende Anforderungen zu stellen sind:

- Die Zufallszahlen sollen gegebenenfalls reproduzierbar sein, um verschiedene Simulationsalternativen miteinander vergleichen zu können.
- Eine Zufallszahlenfolge soll eine möglichst große Periodenlänge aufweisen, d.h. die Zufallszahlenfolge darf sich erst nach einer großen Anzahl von Zufallsgrößen wiederholen
- Die erzeugten Zufallszahlen müssen gute statistische Eigenschaften aufweisen

Da Simulationsexperimente in den meisten Fällen mit Hilfe eines Rechners durchgeführt werden, kommen zwei Anforderungen hinzu:

- die Generierungszeit, die zur Erzeugung einer Zufallszahl benötigt wird, soll möglichst klein sein
- Der Speicherplatzbedarf für die Zufallszahlenfolge bzw. für das Generatorprogramm soll möglichst klein sein

Bei den Methoden zur Erzeugung von Zufallszahlen (Zufallszahlengeneratoren) kann man zwei große Klassen unterscheiden: Generatoren, die echte Zufallszahlen erzeugen und die Pseudozufallszahlen.

1. Echte Zufallszahlen

Zu den Generatoren für echte Zufallszahlen zählen z.B. das Würfeln, das Werfen einer Münze, Roulette und Lottozahlen-Ziehungsgeräte. Weitere geeignete Quellen von Zufallszahlen findet man unter anderem im Zerfall radioaktiver Substanzen, in der Höhenstrahlung und bei Rauschvorgängen in elektronischen Leitungen. Der Vorteil dieser Generatoren besteht darin, daß man echte Zufallszahlen erhält, die die notwendigen statistischen Eigenschaften aufweisen. Nachteile ergeben sich allerdings, wenn innerhalb kurzer Zeit sehr viele oder für Kontrollzwecke reproduzierbare Zufallszahlen benötigt werden. Aus diesem Grunde werden Generatoren für echte Zufallszahlen nur bei weniger umfangreichen Simulationsproblemen angewendet.

Zufallszahlentabellen sind in der einschlägigen Literatur in umfangreichem Maße vorhanden. Sie sind in der Regel aufgrund eines echten Zufallsmechanismus entstanden, unterliegen also keinem mathematischen Bildungsgesetz. Da sie echte Zufallszahlen sind, weisen sie gute statistische Eigenschaften auf und sind gleichzeitig reproduzierbar. Ein wesentlicher Nachteil ergibt sich, wenn in kurzer Zeit sehr viele Zufallszahlen benötigt werden, da der Zugriff auf die Zufallszahlen einen hohen Zeitaufwand erfordert. Deshalb werden Tabellen von Zufallszahlen ebenfalls nur bei Problemen angewandt, bei denen nur wenige Zufallszahlen benötigt werden.

2. Pseudozufallszahlen

Aus den Nachteilen, die bei der Verwendung von echten Zufallsgeneratoren bzw. bei der Verwendung von Zufallszahlentafeln auftreten, ergeben sich die Gründe, Zufallszahlen nach mathematischen Bildungsgesetzen zu generieren. Denn zur Erzeugung von Zufallszahlen mit Hilfe von Algorithmen bedarf es gerade bei Verwendung eines Computers nur kurzer Generierungszeiten und geringer Speicherkapazitäten. Auch sind die so erzeugten Zufallszahlen leicht reproduzierbar. Die mathematische Erzeugungsmethode ist allerdings kein echter Zufallsprozeß, man spricht daher auch von Pseudozufallszahlen, da die von ihr gelieferte Zufallszahlenfolge nach einem bestimmten Zyklus wieder ihren Ausgangswert erreicht oder in einer konstanten Folge z.B. von Nullen entartet. Die Tatsache, daß die Zufallszahlen periodisch und "unecht" sind, ist jedoch so lange unerheblich, als die

Zahlen die statistischen Tests zur Kontrolle der Zufälligkeit bestehen und ihr Zyklus so lang ist, daß in einem Simulationsexperiment nicht mehr als eine Periode benötigt wird.
Bekannte Methoden zur Erzeugung (0,1)-verteilter Zufallszahlen sind:

Zufallszahlen aus Ziffern in transzendenten Zahlen

Zu den transzendenten Zahlen gehören z.B. die Zahl Pi ($\pi = 3,141592654\dots$) und die Eulersche Zahl ($e = 2,71828\dots$). Man kann die Ziffern in transzendenten Zahlen deshalb als Zufallszahlen verwenden, weil die Zahlen e und π hinter dem Komma unendlich viele Stellen haben, die -so wird vermutet- einer Gleichverteilung unterliegen. So war z.B. bereits 1962 die Zahl π auf mehr als 100.000 Stellen bekannt, ohne daß sich nennenswerte Abweichungen von der Gleichverteilung ergaben. Diese Methode hat aber aufgrund des damit verbundenen Rechenaufwandes keine praktische Bedeutung.

Die Mid-Square-Methode

Die Mid-Square-Methode von J. v. Neumann ist ein sehr bekanntes Verfahren zur Erzeugung von Zufallszahlen. Hier wird als Startwert x_1 eine Zahl von mindestens vier Stellen quadriert. Das Ergebnis ist eine maximal achtstellige Zahl. Von dieser Zahl schneidet man die ersten und letzten zwei Ziffern ab und erhält die erste Pseudozufallszahl x_2 , die zur Berechnung der zweiten Zufallszahl wieder quadriert wird. Sollte nach dem Quadrieren nur eine weniger als achtstellige Zahl herauskommen, so wird die Zahl durch führende Nullen ergänzt. Zum besseren Verständnis dient folgendes Beispiel:

$$\begin{aligned}x_1 &= 3123, \\x_2 &= 3123^2 = 09753129, \\x_3 &= 7531^2 = 56715961, \\x_4 &= 7159^2 = 51251281, \text{ usw.}\end{aligned}$$

Durch Transformation (z.B. man denkt sich die Zahlen als "hinterm Komma" stehend) erhält man (0,1)-verteilte Zufallszahlen. Diese Methode ist zwar recht einfach konzipiert und schnell auszuführen, bringt aber schwerwiegende Mängel mit sich, wie z.B. das Auftreten einer kurzen Periodenlänge der Zufallszahlenfolge. Beispielsweise führt der Startwert 3600 zu folgendem Zyklus: 3600, 9600, 1600, 5600, 3600, usw. Der Startwert 2500 führt sogar zu einem Zyklus, der nur eine Zahl umfaßt. Oder der Startwert 1944 liefert nach 13 Rechenschritten das Ergebnis 0. Da das Auftreten dieser Mängel eher die Regel als die Ausnahme ist, hat die Mid-Square-Methode nur noch historische Bedeutung.

Das lineare Kongruenzverfahren

Am gebräuchlichsten sind die sogenannten Kongruenzgeneratoren, von denen hier das lineare Kongruenzverfahren (Lehmer-Generator) vorgestellt wird. Das **Bildungsgesetz** lautet:

$$x_{i+1} = (a \cdot x_i + c) \bmod m$$

Die neue Pseudozufallszahl x_{i+1} entsteht durch Multiplikation der alten Pseudozufallszahl x_i mit der Konstanten a , der Addition von c und durch anschließende Division dieses Terms durch eine Konstante m , wobei der Divisionsrest die neue Zufallszahl x_{i+1} darstellt. Modulo m ($\bmod m$) bedeutet, daß aus $a \cdot x_i + c$ der Divisionsrest von m gebildet wird. Ist $c=0$ spricht man von der **multiplikativen Kongruenzmethode**. Man erhält (0,1)-gleichverteilte Zufallszahlen durch: $z_i = x_i / m$.

Mit den Parametern $x_0 = 7$ (Startwert); $a = 21$, $c = 3$ und $m = 17$ errechnen sich folgende (0,1)-gleichverteilte Zufallszahlen z_i :

i	$x_{i+1} = (a \cdot x_i + c) \bmod m$	$z_i = \frac{x_i}{m}$
1	$(21 \cdot 7 + 3) \bmod 17 = 14$	0.823529
2	$(21 \cdot 14 + 3) \bmod 17 = 8$	0.470588
3	$(21 \cdot 8 + 3) \bmod 17 = 1$	0.058824
4	$(21 \cdot 1 + 3) \bmod 17 = 7$	0.411765
5	$(21 \cdot 7 + 3) \bmod 17 = 14$	0.823529
6	$(21 \cdot 14 + 3) \bmod 17 = 8$	0.470588

Die Zahlenfolge ist periodisch, denn schon nach vier Schritten taucht wieder 7 als Rest auf, der als Startwert Verwendung fand. So stellt die Zahlenfolge von nun an eine Wiederholung der früheren Werte dar.

Je nach Wahl der Parameter a , c , m können sehr unterschiedliche Zahlenfolgen entstehen. Geeignete Werte für a , c , m wurden durch zahlentheoretische Überlegungen aber auch in jüngerer Zeit durch umfangreiche Computereperimente gefunden. Rechenzentren oder Computerhersteller bieten im allgemeinen Programme zur Zufallszahlenerzeugung an, in denen die Parameter so gewählt sind, daß die Periodenlänge und weitere Eigenschaften der Zufallszahlen den meisten Ansprüchen gerecht werden.

Statistische Tests für die Qualität von Pseudozufallszahlen: Pseudozufallszahlen sind nur dann für **Simulationen** verwertbar, wenn sie sich statistisch nicht von echten Zufallszahlen unterscheiden. Um dieser Bedingung gerecht zu werden, müssen bestimmte Kriterien erfüllt werden, die die "gute Qualität" der Zufallszahlenfolge unter Beweis stellen. Solche Kriterien sind:

- Die Zufallszahlen müssen der gewünschten Verteilung genügen, also z.B. gleichverteilt sein
- Die Zufallszahlen müssen unabhängig voneinander sein
- Es darf keine Regelmäßigkeit innerhalb der Zahlenfolge existieren
- Die Zufälligkeit des Auftretens der einzelnen Zufallszahlen muß gegeben sein

Das Eintreten dieser Kriterien kann man u.a. mit Hilfe von **statistischen Tests** überprüfen.

6.3.2.4 Verteilungen

6.3.2.4.1 Mathematische Grundlagen

Der Begriff der Zufallsvariable kann anhand von **Zufallsexperimenten** mit endlichem Ereignisraum erläutert werden. Ein endlicher Ereignisraum ist gegeben, wenn eine bestimmte Anzahl von n Elementarereignissen vorliegt. Betrachtet man beispielsweise die geworfene Augenzahl beim Würfeln, so zeigt sich hier, wie auch in vielen anderen Fällen, daß die **Elementarereignisse** selbst aus Zahlenwerten bestehen. Es handelt sich in diesem Fall um die Werte 1, 2, 3, 4, 5 und 6. Auch in allen anderen Fällen können den Elementarereignissen sinnvoll Zahlenwerte zugeordnet werden. Vor der Durchführung des Zufallsexperimentes kann nicht gesagt werden, welcher Wert konkret eintritt. Derartige Variablen, wie hier die geworfene Augenzahl beim Würfeln, deren Werte vom Zufall abhängen, werden **Zufallsvariable** (Zufallsveränderliche) genannt und im allgemeinen mit Großbuchstaben wie X , Y , oder Z bezeichnet. Ein einzelner Wert, den eine Zufallsvariable X annimmt, nennt man Realisation oder auch Ausprägung der Zufallsvariablen und wird mit x bezeichnet. Wahrscheinlichkeiten für den Verlauf des Experiments werden in der Form $P(\text{Aussage_über_}X)$ spezifiziert, z.B.:

Zufallsvariable X erhält den Wert eines Würfels bei einem Wurf: $P(X=1) = 1/6$

Zur (grafischen) Darstellung verwendet man bei Zufallsvariablen (meist) die (reelle) **Verteilungsfunktion** $F(x)$, der $P(X \leq x)$ zugeordnet ist. Dies ist notwendigerweise eine monoton wachsende Funktion, die bei diskreten Zufallsvariablen Treppenform hat. Liegt eine stetige Zufallsvariable X vor, so gibt $F(x)$ die Wahrscheinlichkeit dafür an, daß ein Element höchstens den Wert x annimmt. Die Verteilungsfunktion einer stetigen Zufallsvariablen ist keine Treppenfunktion wie im diskreten Falle, sondern hat einen stetigen Verlauf. Eigenschaften von Verteilungsfunktionen sind:

- F ist monoton wachsend, d.h. $x < y \Rightarrow F(x) \leq F(y)$
- $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$
- F ist rechtsseitig stetig

Damit erhält man für die Wahrscheinlichkeit, dass x einen Wert um Intervall $(a, b]$ liefert: $P(a < X \leq b) = F(b) - F(a)$

Zufallsvariable können miteinander verknüpft werden, z.B. addiert, multipliziert, usw. Hat man 2 oder mehr Zufallsvariablen X , Y , so werden innere Abhängigkeiten aus der gemeinsamen Verteilungsfunktion $F_{x,y}(x,y)$ deutlich, der $P(X \leq x, Y \leq y)$ zugeordnet ist.

Die Ableitung der Verteilungsfunktion wird als Wahrscheinlichkeitsdichte oder auch **Dichtefunktion** $f(x)$ bezeichnet und entspricht der Wahrscheinlichkeitsfunktion im diskreten Falle. Bei stetigen Zufallsvariablen interpretiert man Wahrscheinlichkeiten als Fläche unter der Dichtefunktion.

$$F(b) - F(a) = \int_a^b f(x) dx$$

Wichtige Parameter von Verteilungen sind Erwartungswert $E(X) = \sum x_i p_i$ (diskrete Werte x_i mit Wahrscheinlichkeiten p_i) bzw. $E(X) = \int x f(x) dx$ (stetig mit Dichte f) und Varianz (Streuung) $E(X^2)^{28}$. Der Erwartungswert ist ein linearer Operator, d.h. aus $Z = aX + bY$ ergibt sich $aE(X) + bE(Y)$.

Wichtig für Simulationsexperimente ist der Begriff der Stichprobe:

Der Stichprobenmittelwert ist definiert als $\bar{x} = \frac{1}{n} \div \sum_{i=1}^n x_i$. Die Varianz σ^2 einer Stichprobe ist

definiert als $\sigma^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$.

Falls Simulationsexperimente unter Verwendung eines Zufallszahlengenerators durchgeführt werden, erhält man als Ergebnis jedes einzelnen Experiments einen Wert einer (oder mehrerer) Zufallsvariablen. Für zuverlässige Resultate ist es daher nötig, mehrere Experimente durchzuführen und den Mittelwert und die Varianz von Stichproben zu bestimmen.

6.3.2.4.2 Gleichverteilung

Bei der Gleichverteilung U auf dem Intervall $[0,1)$ entstehen nur Werte aus diesem Intervall, wobei jeder Punkt gleichwahrscheinlich ist, d.h. $F(x) = x$ für $x \in [0,1)$.

Als Dichte f ergibt sich $f(x) = 1$ für $x \in [0,1)$ bzw. $f(x) = 0$ für $x \notin [0,1)$.

Zufallszahlengeneratoren liefern normalerweise Stichproben von U . Oft braucht man jedoch nicht U , sondern gleichverteilte Zahlen in einem Intervall $[a,b)$:

Verteilungsfunktion und Dichte

$$F_{U[a,b)}(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x < b \\ 1 & \text{für } x \geq b \end{cases}$$

$$f_{U[a,b)}(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{1}{b-a} & \text{für } a \leq x < b \\ 0 & \text{für } x \geq b \end{cases}$$

Erwartungswert und Varianz

$$E(U_{[a,b)}) = \int_a^b x f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{b^2 - a^2}{2 \cdot (b-a)} = \frac{b+a}{2}$$

$$\text{Var}(U_{[a,b)}) = E(U_{[a,b)}^2) - E^2(U_{[a,b)}) = \int_a^b x^2 \frac{1}{b-a} dx - E^2(U_{[a,b)}) = \frac{b^3 - a^3}{3 \cdot (b-a)} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

²⁸ Allgemein nennt man $E(X^n)$ das n -te Moment der Verteilung.

Zur Erzeugung von Zufallszahlen Y , die $U_{[a,b]}$ verteilt sind, reicht es auf einer U -verteilten Variable X die Transformation $Y=a+(b-a)X$ anzuwenden. Für alle y mit $a \leq y \leq b$ gilt: $P(Y \leq y) = P(a + (b-a)X \leq y) = P(X \leq (y-a)/(b-a)) = (y-a)/(b-a)$

6.3.2.4.3 Exponentialverteilung und verwandte Verteilungen

Die **Exponentialverteilung** gibt die Wahrscheinlichkeit dafür an, dass zwischen dem Eintreffen zweier aufeinanderfolgender Ereignisse höchstens x Zeiteinheiten vergehen. Zur Bestimmung des Abstands, in dem "Modell-Kunden" durch die Quelle erzeugt werden, wird die Zwischenankunftszeit meist exponentiell verteilt gewählt. Die Begründung dafür liefern sog. Zählprozesse.

Zählprozesse: Ein Zählprozeß ist eine Familie $\{X(t) | t \text{ ist reell und positiv}\}$ von Zufallsvariablen mit den folgenden 2 Eigenschaften: $X(t)$ ist aus dem Bereich der natürlichen Zahlen und $X(t_1) \leq X(t_2)$ für $t_1 \leq t_2$.

Besonders wichtige Zählprozesse sind die Poisson-Prozesse. Dies sind Zählprozesse mit den folgenden vier Eigenschaften:

- Die Inkremente sind unabhängig voneinander, d.h. für beliebige Zeitpunkte $t_1 < t_2 \leq t_3 < t_4$ sind stets die Zufallsvariablen $X(t_2) - X(t_1)$ und $X(t_4) - X(t_3)$ voneinander unabhängig.
- Die Inkremente sind stationär, d.h. die Verteilung von $X(t_2) - X(t_1)$ hängt nur von dem Abstand $t_2 - t_1$ ab, nicht aber von dem Wert t_1 (bzw. t_2).

$$\lim_{h \rightarrow 0} \frac{P(X(h) \geq 1)}{h} = 0, \text{ d.h. es kommen keine Ankünfte in Gruppen vor.}$$

$$\text{- Es ist } X(0) = 0$$

Bzgl. der Verteilungsfunktion F der Zwischenankunftszeiten eines Poisson-Prozesses $X(t)$ mit $0 < t_1 < t_2$ gilt:

$$P(X(t_2) = 0) = P(X(t_1) = 0 \wedge X(t_2) - X(t_1) = 0) = P(X(t_1) - X(0) = 0 \wedge (X(t_2) - X(t_1) = 0) = P(X(t_1) - X(0) = 0) \cdot (P(X(t_2) - X(t_1) = 0) = P(X(t_1) = 0) \cdot P(X(t_2 - t_1) = 0)$$

Es gibt nur wenige Funktionen, die für alle $x, y > 0$ die Eigenschaft $f(x+y) = f(x) + f(y)$ haben: Entweder ist f konstant oder aber $f(x) = e^{cx}$. Mit den Randbedingungen $f(x) \leq 1$, $f(0) = 1$, $\lim_{x \rightarrow \infty} f(x) = 0$ ergibt sich $c < 0$. Also gilt für geeignetes $\lambda > 0$ (Ankunftsrate

des Poisson-Prozesses): $P(X(t) = 0) = e^{-\lambda t}$ und damit $P(X(t) > 0) = 1 - e^{-\lambda t}$. Damit kann für die Verteilung des Zeitabstands zwischen zwei aufeinanderfolgenden Ankünften bis zur jeweils nächsten Ankunft (von einem beliebigen Zeitpunkt t_0 aus) angegeben werden: $P_{t_0}(T \leq t) = P(X(t_0) < P(t_{0+t})) = P(0 < X(t)) = 1 - e^{-\lambda t}$

Eine Zufallsvariable X mit der Verteilungsfunktion $F(X) = 1 - e^{-\lambda x}$ heißt exponentialverteilt, ihre Dichte ergibt sich aus $f(x) = \lambda \cdot e^{-\lambda x}$.

Erwartungswert und Varianz:

$$E(X) = 1/\lambda$$

$$\text{Var}(X) = 1/\lambda^2$$

Erzeugung exponentialverteilter Zufallszahlen (aus gleichverteilten Zufallszahlen mit dem Prinzip der Inversion²⁹):

²⁹ prinzipiell bei allen Verteilungen anwendbar

$$F^{-1}(y) = -\ln(1 - y) / \lambda$$

Ausfallrate: $\frac{f(t)}{1 - F(t)} = \frac{\lambda \cdot e^{-\lambda t}}{e^{-\lambda t}} = \lambda$

Liegen mehrere unabhängig verteilte Poisson-Prozesse mit Raten $\lambda_1, \lambda_2, \dots, \lambda_n$ vor, so können durch Zählen der Ereignisse aller Prozesse diese zu einem Prozeß zusammengefasst werden. Dieser durch Überlagerung entstehender Prozeß ist wieder ein Prozeß, nun mit der Rate $\sum \lambda_i$

Erlang-k-Verteilung

Durch Aussummieren von k unabhängigen Exponentialverteilungen gleicher Rate λ erhält man die **Erlang-k-Verteilung**, die den Erwartungswert k / λ und die Varianz k / λ^2 hat.

Dichte: $f(x) = \lambda \cdot e^{-\lambda x} (\lambda x)^{k-1} / (k-1)!$

Verteilungsfunktion: $F(x) = 1 - \sum_{j=0}^{k-1} e^{-\lambda x} (\lambda x)^j / j!$

Die Exponentialverteilung besitzt die praxisferne Eigenschaft, dass sehr kurze Zeiten die höchste Wahrscheinlichkeitsverteilung besitzen. In der Praxis brauchen viele Prozessen eine Mindestzeit, d.h. kurze Zeiten sind eher unwahrscheinlich. Die Erlang-Verteilung berücksichtigt dies. Sie besitzt einen zusätzlichen Parameter k. k wird als Ordnung der Verteilung betrachtet und kann die Werte $k = 1, 2, 3, \dots$ annehmen. Zu jedem k gibt es eine Erlang-Verteilung (Erlang-k-Verteilung). Je größer k ist, desto unwahrscheinlicher werden kurze Zeiten, desto geringer wird aber auch die Standardabweichung (die Streuung der Verteilung). Praktische Bedeutung haben Erlang-Verteilungen bis zur Ordnung 3. Für $k = 1$ ist die Erlangverteilung mit der Exponentialverteilung identisch.

Geometrische Verteilung

(diskrete Verteilung, eng verwandt mit der Exponentialverteilung).

Gegeben ist ein Experiment, das mit einer Wahrscheinlichkeit p erfolgreich und mit $q = (1 - p)$ erfolglos verläuft. Gesucht ist die Wahrscheinlichkeit P(n) dafür, dass n Wiederholungen des Experiments hintereinander erfolglos bleiben und das (n+1). Experiment gelingt:

$P(n) = q^n \cdot p$

Verteilungsfunktion: $F(n) = \sum p \cdot q^n = 1 - q^{n+1}$

6.3.2.4.4 Normalverteilung und verwandte Verteilungen

Die Normalverteilung ist definiert über Dichte: $N(\mu, \sigma^2) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-1/2 \left(\frac{x-\mu}{\sigma}\right)^2}$. Wegen der Symmetrie in $x - \mu$ hat sie offensichtlich den Mittelwert μ , als Varianz ergibt sich σ^2 . σ ist die Standardabweichung von $N(\mu, \sigma^2)$.

Beliebige Linearkombinationen normalverteilter Zufallszahlen sind wiederum normalverteilt. Mit $Y \sim N(\mu, \sigma^2)$ ist $aY \sim N(a\mu, a^2\sigma^2)$. Ferner gilt bei $Y_i \sim N(\mu_i, \sigma_i^2)$ für die Summe $Y_1 + Y_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Die Bedeutung der Normalverteilung ergibt sich aus dem zentralen Grenzwertsatz: Bereits unter schwachen Voraussetzungen ist der Mittelwert $(X_1 + X_2 + X_3 + \dots + X_n)/n$ normalverteilt. Eine abgeschwächte Version dieses Satzes hat bspw. folgende Aussage:

$X_1 + X_2 + X_3 + \dots$ sind unabhängige, aber identisch verteilte Zufallszahlen mit (gleichem) Erwartungswert e und (gleicher) Varianz v . Falls S_n die Summe $X_1 + X_2 + X_3 + \dots + X_n$ zugeordnet ist, konvergiert die Verteilung mit wachsendem n von $\frac{S_n - ne}{\sqrt{nv}}$ gegen die Standard-Normalverteilung $N(0,1)$. Damit ist für große n die Verteilung der Summe S_n annähernd $N(e, v/n)$ verteilt.

Bsp.: Die Aufsummierung von 12 Gleichverteilungen $U(-1/2, 1/2)$ mit Erwartungswert $1/12$ ist annähernd $N(0, 1)$ -verteilt.

Binomialverteilung

(diskrete Verteilung, der Normalverteilung verwandt)

Sie drückt die Wahrscheinlichkeit für die Anzahl erfolgreicher Bernoulli-Experimente aus: n unabhängige Experimente, von denen jedes mit der Wahrscheinlichkeit p erfolgreich ist und mit der Wahrscheinlichkeit $q = (1 - p)$ nicht gelingt. $P_{n,p}(i)$ ist die Wahrscheinlichkeit dafür, dass genau i der Experimente gelingen. Es ergibt sich:

$$P_{n,p}(i) = \binom{n}{i} \cdot p^i \cdot q^{n-i}$$

Erwartungswert: $n \cdot p$

Varianz: $n \cdot p \cdot q$

Poissonverteilung

Sie geht aus der Binomialverteilung für große n und kleine Wahrscheinlichkeiten p hervor. Der Mittelwert ist der einzige Parameter der Verteilung ($\sigma^2 = \mu$). Es gilt

$$P(\mu, x = k) = \frac{\mu^k}{k!} \cdot e^{-\mu} \text{ für } k = 0, 1, 2, \dots \text{ Als Faustregel für die Anwendbarkeit bei}$$

mehrstufigen Zufallsexperimenten gilt $np < 10$ mit $n > 1500p$.

Große Bedeutung hat die Poisson-Verteilung bei der diskreten Simulation, denn sie beschreibt die Wahrscheinlichkeit dafür, ob genau k Ereignisse in einem festen

Intervall eintreten. Dies wird bei Warteschlangenproblemen aller Art benötigt (Ankunft von Kunden, Eintreffen von Telefongesprächen, Ankunft eines Fahrzeugs, usw.).

6.3.2.4.5 Allgemeine Methoden zur Erzeugung von Zufallszahlen

Direkte Inversion einer stetigen Zufallsvariablen

Ist die Verteilungsfunktion F einer Zufallsvariablen X leicht berechenbar und invertierbar, so kann man aus gleichverteilten Zufallsvariablen U eine F -verteilte Zufallsvariable konstruieren:

Die Zufallsvariable Y habe die Verteilungsfunktion F . F ist stetig und streng monoton. U ist gleichverteilt auf dem Intervall $[0,1)$, als $P(U \leq x)$ für $x \in [0,1)$. Dann gilt $P(F^{-1}(U) \leq t) = P(U \leq F(t))$, d.h. die Zufallsvariable $F^{-1}(U)$ hat die gleiche Verteilung wie X .

Die Transformation kann graphisch, analytisch oder über "Table-look-up" (Nachschlagen in einer Tabelle) erfolgen.

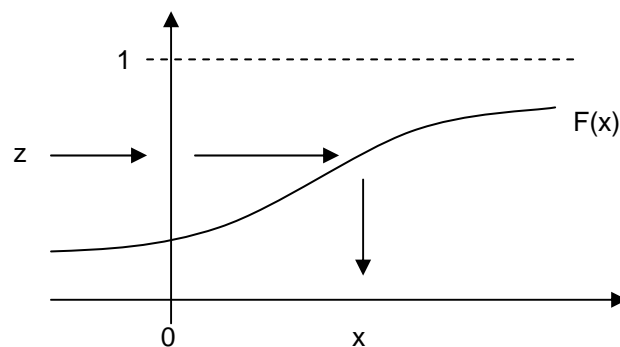


Abb.: Transformation einer $(0,1)$ -gleichverteilten Zufallszahl z in eine Zufallszahl x einer stetigen Verteilung

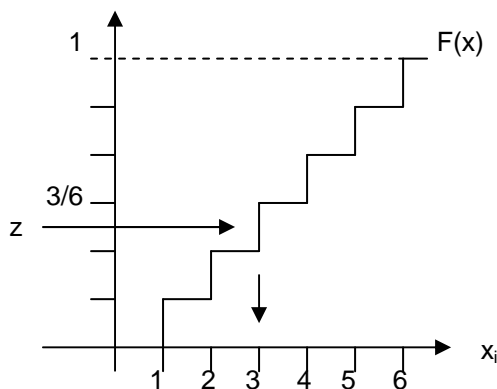


Abb.: Transformation einer $(0,1)$ -gleichverteilten Zufallszahl z in eine Zufallszahl x einer diskreten Verteilung.

Notwendig ist also ein Berechnungsverfahren für die Inversion F^{-1} der betrachteten Verteilungsfunktion. Ist diese Inverse leicht zu berechnen, dann kann man mit der Inversion Zufallszahlen mit Verteilung F erzeugen.

Bsp.: Bei der Exponentialverteilung ergibt sich aus $F(x) = y = 1 - e^{-\lambda x}$, $e^{-\lambda x} = 1 - y$,
 $x = -\log(1 - y) / \lambda$

Funktionale Approximation der Inversen

Bestimme mit numerischen Methoden eine Approximation für die Inverse F^{-1} der Verteilungsfunktion.

Bsp.: rationale Approximation mit $Y = \sqrt{-\log(1-U)^2}$ und

$$X = Y + \frac{p_0 + p_1Y + p_2Y^2 + p_3Y^3 + p_4Y^4}{q_0 + q_1Y + q_2Y^2 + q_3Y^3 + q_4Y^4} \text{ mit folgenden Werten für die } p_i \text{ und } q_i$$

$p_0 = 0.322232431088$	$q_0 = 0.099348462606$
$p_1 = -1$	$q_1 = 0.588581570495$
$p_2 = -.342242088547$	$q_2 = 0.531103462366$
$p_3 = 0.0204231210245$	$q_3 = 0.10353775285$
$p_4 = 0.00004536422101148$	$q_4 = 0.0038560700634$

Für $0.5 \leq U < 1$ approximiert die sich ergebende Funktion $X = X(U)$ die Inverse von $N(0,1)$ ³⁰ mit einer relativen Genauigkeit von etwa 6 Dezimalen.

³⁰ Eine direkte formale Anwendung der Inversionsmethode ist nicht möglich, da es keinen geschlossenen Ausdruck für die Verteilungsfunktion der Normalverteilung (und ihrer Inversen) gibt.

6.4 Diskrete Simulation

6.4.1 Modellierung und Sichtweise

Die Modellbildung zur Simulation diskreter Systeme sollte folgende Schritte umfassen:

- beteiligte Objekte identifizieren
- Systemzustand über Zustandsvariable beschreiben (statische Struktur)
- Zustandsänderung an konkreten Zeitpunkten beschreiben (dynamische Struktur).

Zustandsänderungen erfolgen durch Ereignisse. (Ereigniszeitpunkte). Die Zeitspanne zwischen zwei Ereignissen variiert fallweise. Sie lassen sich aber mit einer Wahrscheinlichkeitsverteilung beschreiben.

Bsp.: Postschalter

Beteiligte Objekte an einem Postschalter sind Kunden, die Warteschlange und der Postbeamte. Zustandsvariablen sind die Warteschlangenlänge und der Zustand Schalter belegt/nicht belegt. Zustandsänderungen ergeben sich durch die Ereignisse: Ankunft neuer Kunde, Bearbeitungsbeginn Kunde, Bearbeitungsende Kunde. Die Ankunftszeiten der Kunden am Postschalter sind zufällig; kennt man aber den Mittelwert (z.B. im Schnitt 20 Kunden pro Stunde), kann man die Ankunftszeiten über eine Exponentialverteilung beschreiben. Die Bearbeitungszeiten schwanken ebenfalls. Hier bietet sich zur Beschreibung die Normalverteilung oder eine Erlangverteilung an (sehr kurze Bearbeitungszeiten sind unwahrscheinlich).

Es gibt drei unterschiedliche Sichtweisen zur Durchführung der Simulation auf einem Rechner: Ereignisse, Aktivitäten, Prozesse.

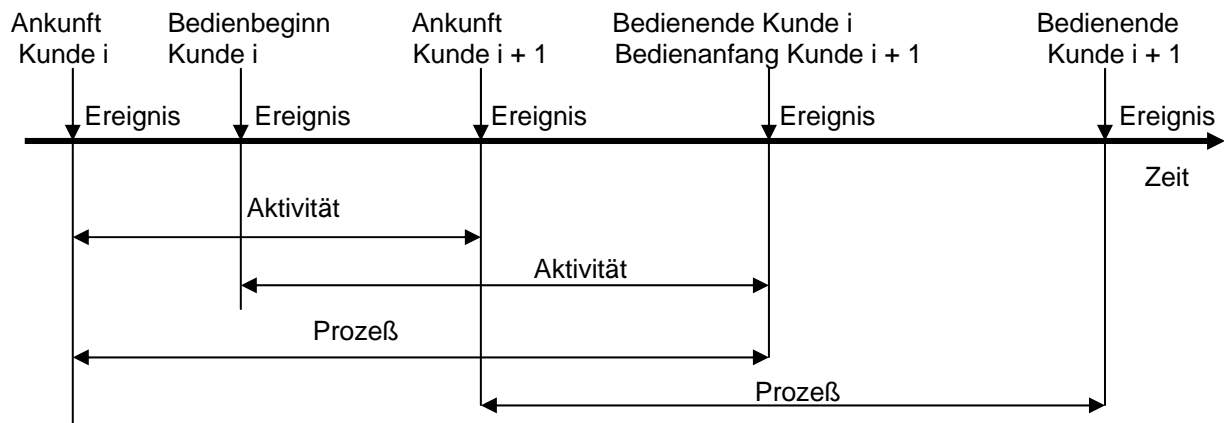


Abb.: Ereignisse, Aktivitäten und Prozesse

6.4.2 Komponenten ereignisorientierter Simulationen

1. Ereignisorientierte Simulation

Ereignisse sind für die Änderung der Zustandsvariablen verantwortlich. Der Zustand eines Systems wird durch geeignete Datenstrukturen beschrieben.

Ereignisse

Man kann Ereignisse in Ereignistypen klassifizieren, die jeweils eine spezifische Zustandsvariablen-Änderung bewirken. Diese Änderung kann pro Ereignistyp in einer

Ereignisroutine zusammengefasst werden. Das Simulationsprogramm ist dann für die zeitgerechte Ausführung der Ereignisroutine verantwortlich.

Programmtechnische Realisierung:

Im Simulationsprogramm gibt es eine globale Simulationszeit t , der jeweils der nächste Ereigniszeitpunkt t_i zugewiesen wird.

Zukünftige Ereignisse werden mit Ereigniszeitpunkt t_i und Ereignistyp ET_i in einer Ereignisliste gespeichert.

In einer Schleife wird je nach anstehendem Ereignistyp die entsprechende Ereignisroutine ausgelöst. Eine Ereignisroutine kann neue (zukünftige) Ereignisse in die Ereignisliste eintragen.

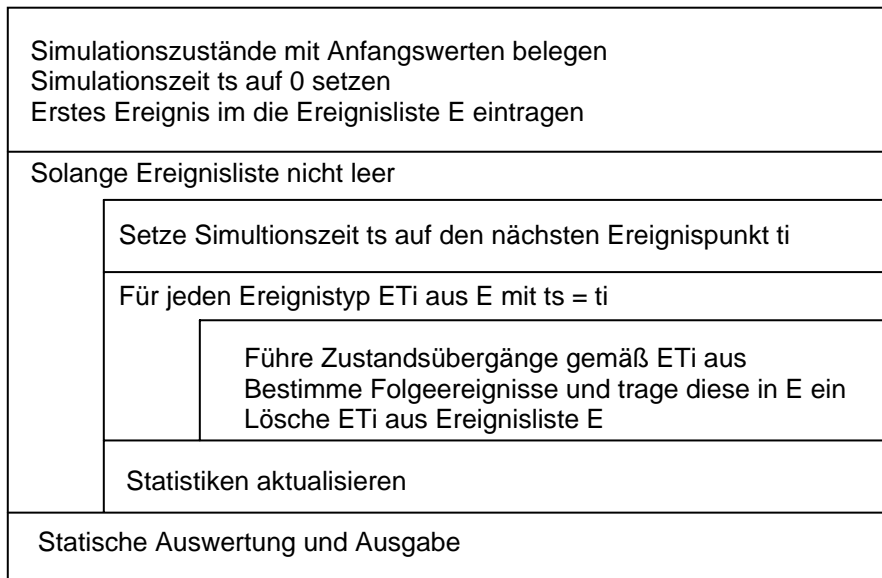


Abb.: Programmablauf der ereignisorientierten Simulation

2. Aktivitätsorientierte Simulation

Ereignistypen werden weiter gruppiert und Aktivitäten zugeordnet. Jeder Beginn und jedes Ende einer Aktivität entsprechen dabei einem Ereignis, d.h. nur hier sind Zustandsänderungen des Systems möglich. Jede Aktivität hat eine festgelegte Zeitdauer, wobei diese eventuell einer statistischen Verteilungsfunktion folgt. Jeder Aktivität ist eine Aktivitätsroutine zugeordnet, die die Zustandsänderungen am Beginn und am Ende der Aktivität beschreibt. Aktivitäten werden wiederholt angestoßen. Im Postschalterbeispiel gibt es die Aktivitäten Ankunftserzeugung und Bedienung.

Programmtechnische Realisierung

Jede Aktivität besitzt und setzt eine eigene Uhr.

Bei Beginn und am Ende einer Aktivität können Zustandsänderungen durchgeführt werden (Aktivitätsroutine).

Eine Aktivitätsroutine kann neue Aktivitäten anstoßen, in dem die entsprechende Uhr gesetzt wird.

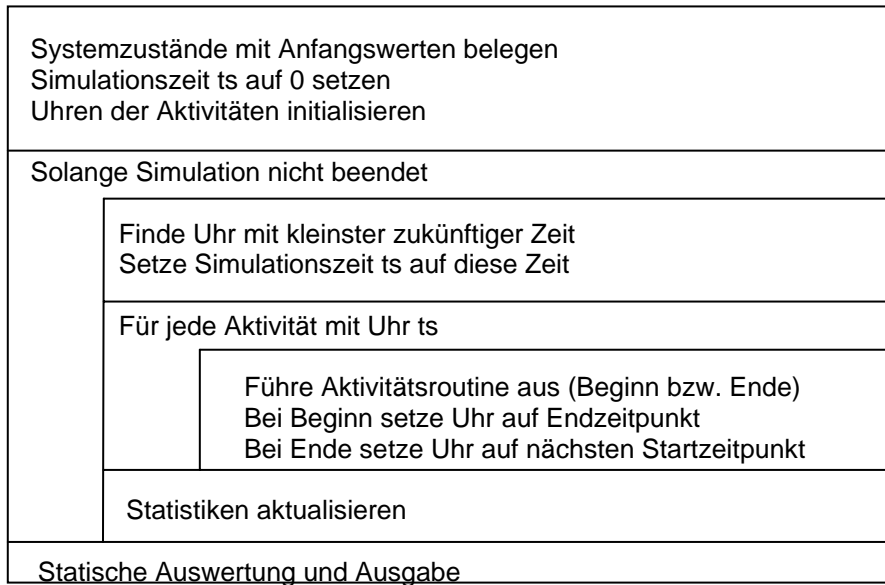


Abb.: Programmablauf der aktivitätsorientierten Simulation

3. Prozessorientierte Simulation

Eine größere Anzahl von Ereignissen wird als zusammengehörig betrachtet. Das System wird in Objekte zerlegt, die über Ereignisse miteinander interagieren.

Bsp.: Postschalter

Das Objekt Kunde kennt drei Ereignisse: "ankommen", "bedient werden", "weggehen".

Das Objekt Bediener kennt die Ereignisse: "Bedienung beginnen", "Bedienung beenden".

Das Objekt Umwelt "schickt" Kunden in das Bediensystem.

Diese dynamischen Objekte werden auch als Prozesse bezeichnet. Das zeitliche Verhalten eines Prozesses wird in einer Prozessroutine festgehalten. Das Simulationssystem muß in der Lage sein, mehrere Prozesse unabhängig von-einander zu verwalten (erzeugen, starten, beenden, unterbrechen, fortführen, ...).

Programmtechnische Realisierung

Prozesse laufen unabhängig voneinander.

Sie haben die Möglichkeit, Nachrichten untereinander auszutauschen.

Jeder Prozess läuft solange, bis eine Wartesituation eintritt. Dies kann das Warten auf eine Nachricht sein oder die Deaktivierung des Prozesses für eine festgelegte Zeitspanne.

6.4.3 Typische Modellklassen

Problemstellungen kommen aus den Bereichen: Bedienungs-/Wartesysteme, Lagerhaltungssysteme, ausfallanfällige Systeme

6.4.3.1 Bedienungs-/ Wartesysteme

Es handelt sich um Systeme mit

- Kunden / Aufträgen
- Abfertigung in Bedienstationen mit evtl. mehreren Bedieneinheiten
- Warteschlangen bei nicht sofort erfüllbaren Anforderungen

6.4.3.2 Lagerhaltungssysteme

Grundlagen

Lager dienen als Puffer zwischen Produktion (oder Auslieferung) und Verbrauch. Es gibt viele Arten, z.B. Verkaufs-, Produktions-, Zwischen-, Zentrallager. Ihnen gemeinsam ist, dass sie alle einen Ausgleich zwischen Nachfrage und Lieferfähigkeit schaffen sollen. Aufgabe des Lagerverwalters ist es, Kosten so gering wie möglich zu halten. Hierbei hat er folgende Aufgaben zu berücksichtigen:

- Hohe Deckung der Nachfrage
- wenig Kapitalbindung im Lager
- Niedrige Beschaffungs- und Lagerkosten
- keine Schwierigkeit bei der Zulieferung

Diese Ziele stehen teilweise in Konkurrenz zueinander. Es ist notwendig, die verschiedenen Ziele aufeinander abzustimmen. Der Lagerhalter wird mit folgenden Fragen konfrontiert:

- Wieviel ist zu bestellen?
- Wann ist zu bestellen?
- Wann sind die Bestände zu kontrollieren?
- Wie hoch muß der Lagerbestand sein?
- Wie weit geht der Planungshorizont?

Lagerhaltungspolitik

Die Beantwortung der vorstehenden Fragen führt zwangsläufig zur Verfolgung einer bestimmten Lagerhaltungspolitik, die durch folgende Einflußgrößen beschrieben werden kann:

die Bestellmenge q	wird häufig als Losgröße bezeichnet, wenn die Bestellmenge zu allen Bestellpunkten gleich ist.
das Bestellniveau S	beschreibt die vorgegebene Höhe auf die der Lagerbestand aufzustocken ist

der Bestellpunkt s	bezeichnet die Mengen des Lagerbestands, bei der nachbestellt werden muß
der Bestellzyklus T	beschreibt die Zeitspanne, die zwischen 2 aufeinanderfolgenden Bestellungen vergeht.
die Beschaffungszeit τ	gibt an, wie viel Zeit zwischen Bestellung und Verfügbarkeit der Ware vergeht.

Bsp. für verschiedene Lagerhaltungspolitik

(T,q)-Politik	In festen Zeitintervallen (Bestellzyklus T) wird eine konstante Menge (Losgröße) q bestellt.
(s,q)-Politik	Jeweils zu Zeitpunkten, in denen der Lagerbestand den Bestellpunkt unterschritten hat, wird eine Menge q nachbestellt
(s,S)-Politik	Die Bestellung wird ausgelöst, wenn der Bestand den Bestellpunkt s unterschreitet. Der Lagerbestand wird bis zum Bestellniveau S aufgestockt.

Lagerhaltungssimulation

1. Simulation einer (T,q)-Lagerhaltungspolitik

Die Simulation soll folgende Bedingungen berücksichtigen:

- normalverteilter Bedarf mit $\mu = 100$ und $s = 20$
- Lagerkosten: 0.15 Cent je Stück
- Bestellkosten: 20.- Euro je Bestellung
- Fehlmengenkosten: 2.50 Euro je Stück
- Rabatt: 2% bei Bestellung über 2500.- Euro
- Preis: 7.50 Euro je Stück
- Lieferzeit: 1 Monat

Lösung mit Hilfe eines Java-Programms. Der Bedarf wird für jeden Monat mit Hilfe von normalverteilten Zufallszahlen ($\mu = 100$ und $s = 20$) ermittelt. Zu Beginn gebe es einen Restbestand von 3 Stück.

Falls man verschiedene (T,q)-Politiken durchsimuliert³¹, ergeben sich folgende Ergebnisse:

Politik	Kosten
1,100	381,17
2,200	340,19
3,300	385,00
4,400	265,21
5,500	343,01
6,600	418,51

Am günstigsten ist im vorliegenden Modell eine (4,400)-Politik. Hier wird auch die Rabattgrenze erreicht

2. Simulation einer Lagerhaltungspolitik (S,s) für ein Produkt

Die Simulation soll folgende Bedingungen berücksichtigen:

- Betrachtungszeitraum: n Monate (120 Monate)

³¹ vgl. pr63360, LagerSimul.java

- Kundenbestellungen fallen in exponentiell verteilten Abständen, im Mittel 0.1 Monate, mit folgenden Wahrscheinlichkeiten an:

Stück	Wahrscheinlichkeit
1	1/6
2	1/3
3	1/3
4	1/6

- Kundenbestellungen werden, so weit wie möglich, sofort befriedigt, ein Rest, sobald vom Großhändler geliefert wird.
- Was nicht geliefert werden kann, erzeugt einen negativen Lagerbestand.
- Zu Beginn eines jeden Monats werden z Stück beim Großhändler bestellt gemäß einer Strategie (S,s) :

$$z = \begin{cases} S - q & q < s \\ 0 & q \geq s \end{cases}$$

- Die Kosten dafür sind $K + i \cdot z$
 K : Festkosten (setup cost)
 i : Stückkosten

- Die Lieferung erfolgt nach einer Zeit, die gleichverteilt zwischen 0.5 und 1 Monat ist.
- $q(t)$: Lagerbestand, auch negativ bei unbefriedigten Bestellungen:

$$q^+(t)$$

$q^-(t)$: unbefriedigte Bestellungen, noch an die Kunden zu liefern

- Lagerkosten (holding cost) $c_L = 1$ [GE] je Stück und Monat
- Festkosten der Lagerung werden nicht berücksichtigt, da bei allen Strategien gleich.
- Ausfallkosten bei nicht möglicher Lieferung $c_F = 5$ [GE] je Stück und Monat
- Gesamte Lagerkosten und Ausfallkosten je Monat im Mittel:

Mittlerer physischer Bestand: $\bar{q}^+ = \frac{1}{n} \int_0^n q^+(t) dt$

Mittlere nicht befriedigte Bestellungen: $\bar{q}^- = \frac{1}{n} \int_0^n q^-(t) dt$

mittlere monatliche Lagerkosten: $c_L \cdot \bar{q}^+$

mittlere Ausfallkosten: $c_F \cdot \bar{q}^-$

- 9 Strategien

s	20	20	20	20	40	40	40	60	60
S	40	60	80	100	60	80	100	80	100

- Ereignisse:

Lieferung vom Großhändler (order arrival)	1
Kunde bestellt (demand)	2
Simulationseende nach n Monaten (end simulation)	3
Bestandsprüfung am Monatsanfang	4

3 Verteilungen für Zufallszahlen werden gebraucht:

- Größe der Bestellungen durch den Kunden: Mit $best_1 = [0, 1/6), best_2 = [1/6, 1/2), best_3 = [1/2, 5/6), best_4 = [5/6, 1]$ und $U(0,1)$ (0,1)-gleichverteilte Zufallszahl wird die Kundenbestellung $d = i$ für $U(0,1) \in best_i$ genommen. Verzögerungen der Lieferung durch den Großhändler sind (a,b)-gleichverteilt mit $a = 0.5, b = 1$: $a + (b - a)U(0,1)$. Die Zeit zwischen Bestellungen von Kunden: exponentiell verteilt, im Mittel 0.1 Monate.

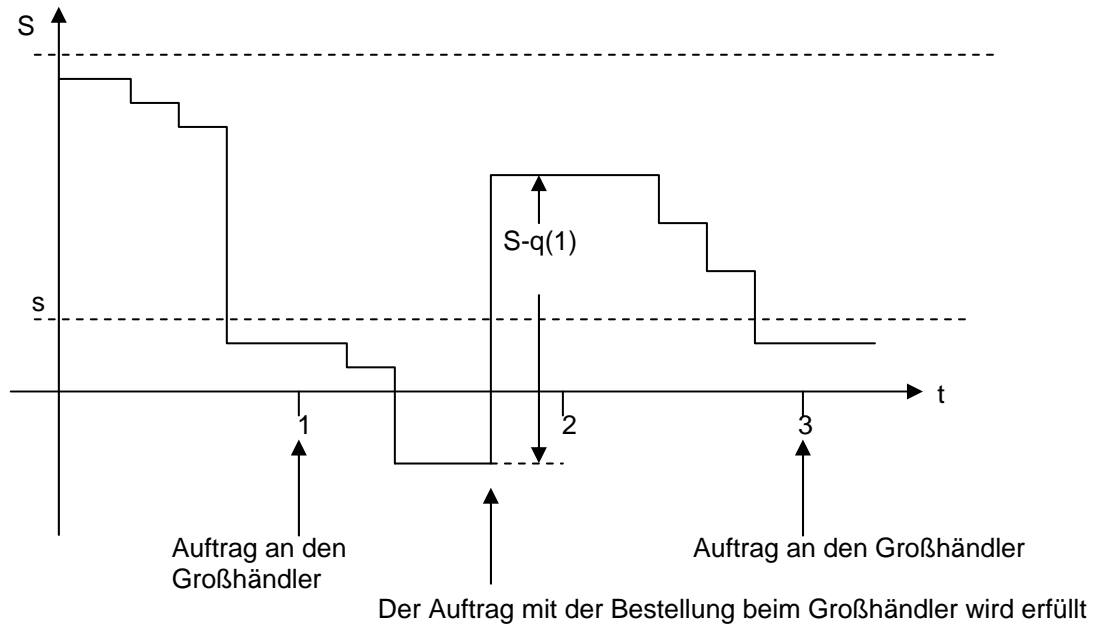


Abb.:

Ausgabe des Programms:

C:\Eingabeaufforderung

D:\dok\or\ss05\progr\pr63360>LagerSimulS

Single-product inventory system

Initial inventory level60 items

Number of demand sizes4

Distribution function of demand sizes0.1670.5000.8331.000

Mean interdemand time0.10

Delivery lag range0.50 to1.00 months

Length of the simulation120 months

K = 32.0 i = 3.0 h = 1.0 pi = 5.0

Number of policies9

Policy	Average total cost	Average ordering cost	Average holding cost	Average shortage cost
< 20, 40>	125.71	97.96	8.89	18.87
< 20, 60>	120.65	91.70	17.43	11.52
< 20, 80>	123.65	86.29	25.72	11.64
< 20,100>	126.58	81.82	36.32	8.45
< 40, 60>	126.96	101.07	24.82	1.07
< 40, 80>	126.66	90.02	34.59	2.05
< 40,100>	134.86	89.10	44.32	1.44
< 60, 80>	143.48	98.70	44.72	0.06
< 60,100>	147.94	94.87	53.05	0.03

Abb.: