



Stabilität und Approximation stochastischer Optimierungsprobleme

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht am 14. Mai 2007 an der
Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von
Diplom-Mathematiker Holger Heitsch
geboren am 22. März 1973 in Berlin

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II
Prof. Dr. Wolfgang Coy

Gutachter

1. Prof. Dr. Werner Römisch (Humboldt-Universität zu Berlin)
2. PD Dr. René Henrion (WIAS Berlin/Humboldt-Universität zu Berlin)
3. Prof. Dr. Georg Ch. Pflug (Universität Wien)

Tag der Verteidigung: 9. November 2007

Dank

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Mathematik an der Humboldt-Universität zu Berlin und wurde teilweise gefördert vom BMBF (Bundesministerium für Bildung und Forschung), vom DFG-Forschungszentrum MATHEON „Mathematik für Schlüsseltechnologien“ sowie durch EDF (Electricité de France).

Besonderer Dank gilt Herrn Professor Werner Römisch, der durch zahlreiche Ideen, Diskussionen und Anregungen einen wesentlichen Anteil zum Entstehen dieser Arbeit beigetragen hat. Dafür und für das hervorragende Arbeitsklima möchte ich ihm herzlich danken.

Allen derzeitigen und ehemaligen Kollegen möchte ich für die angenehme und konstruktive Zusammenarbeit danken. Weiterhin gilt mein Dank den Mitarbeitern der Arbeitsgruppe von EDF Research & Development für wertvolle Denkanstöße und Diskussionen.

Berlin, im Mai 2007

Holger Heitsch

Inhaltsverzeichnis

1	Einleitung	1
1.1	Stochastische Optimierung	2
1.2	Gliederung der Arbeit	3
2	Das Optimierungsmodell	5
2.1	Stochastische Prozesse	5
2.1.1	Beispiel bivariater Last-Preisprozess	7
2.2	Das lineare mehrstufige Problem	8
2.2.1	Modellierung des Datenprozesses	9
3	Wahrscheinlichkeitsabstände	11
3.1	Wahrscheinlichkeitsfunktionale	11
3.1.1	Die Kostenfunktion	12
3.1.2	Eigenschaften	13
3.2	Endliche Verteilungen	14
3.2.1	Monge-Kantorovich-Funktional	15
3.2.2	Kantorovich-Rubinstein-Funktional	15
4	Stabilität	17
4.1	Mehrstufige Probleme	17
4.1.1	Generelle Voraussetzungen	18
4.1.2	Der Filtrationsabstand	20
4.1.3	Stabilität von Optimalwerten	21
4.1.4	Beispiel	28
4.1.5	Stabilität von Lösungsmengen	30
4.2	Zweistufige Probleme	36
4.2.1	Stabilität zweistufiger Kompensationsmodelle	37
5	Szenarioreduktion	39
5.1	Prinzip der optimalen Szenarioreduktion	39
5.2	Heuristische Algorithmen	42
5.2.1	Vorwärtsalgorithmus	43
5.2.2	Rückwärtsalgorithmus	43

5.3	Bestimmung der reduzierten Kosten	45
5.3.1	Dijkstra-Algorithmus	45
6	Konstruktion von Szenariobäumen	47
6.1	Die Konstruktionsalgorithmen	47
6.1.1	Konzeption	47
6.1.2	Generelle Bezeichnungen	50
6.1.3	Der Rückwärtsalgorithmus	50
6.1.4	Der Vorwärtsalgorithmus	55
6.2	Konvergenz	60
6.2.1	Konvergenz von Diskretisierungen	62
6.2.2	Konvergenz von Schätzungen	65
6.2.3	Anwendung auf die Szenariobaumkonstruktion	68
7	Anwendungen	73
7.1	Ergebnisse der Szenarioreduktion	73
7.1.1	Bewertung der Reduktionsalgorithmen	77
7.2	Anwendung der Baumkonstruktion	79
7.2.1	Szenariobäume für EDF	79
7.2.2	Konstruktion von Last/Preis-Szenarien	85
7.2.3	Szenariobäume für Buchungsklassen	89
7.2.4	Schlussbemerkung	92
	Literaturverzeichniss	95

Kapitel 1

Einleitung

Charakteristisch für eine Vielzahl von Optimierungsproblemen ist, dass getroffene Entscheidungen, je nach Entwicklung zufälliger Ereignisse, verschiedene Auswirkungen in der Zukunft nach sich ziehen können. Eine Entscheidung würde sich demnach dann als gut bewerten lassen, wenn sie im gewissen Maße unabhängig von zukünftigen Ereignissen oder zumindest mit hoher Wahrscheinlichkeit in der Zukunft zu positiven Ergebnissen führt. Denkt man an die Gewinnmaximierung eines Unternehmens innerhalb eines vorgegebenen Planungszeitraums, so könnte beispielsweise als eine gute Entscheidung eine solche angesehen werden, die mit sehr hoher Wahrscheinlichkeit einen Gewinn am Ende des Zeitraums hervorbringt. Auch der im Mittel zu erwartende Gewinn kann als ein Optimalitätskriterium für eine gute Entscheidung herangezogen werden. Beide Betrachtungsweisen lassen sich auch kombinieren. Die erste führt zu einem geringen Risiko der Entscheidung, während die zweite zu einer Ertragsmaximierung im Mittel führt. Das geeignete Werkzeug zur Modellierung liefert in beiden Fällen die stochastische Optimierung.

In der Praxis finden stochastische Überlegungen mehr und mehr den Weg bis in die Anwendungen. In der Energiewirtschaft beispielsweise sind infolge der Liberalisierung der Strommärkte die Anforderungen an das Portfoliomanagement stark gestiegen. Zunehmende Volatilitäten und höhere Risiken führten in der Vergangenheit zu einem ständig wachsenden Interesse an leistungsfähigen Optimierungsmodellen, die der neuen Situation Rechnung tragen. Stark schwankende Strompreise an den Strombörsen, verändertes Verbraucherverhalten und die Verknüpfung von Kraftwerkseinsatz und Stromhandel erfordern die Einbeziehung von Stochastik in die Optimierungsmodelle. Einfache Prognosemodelle stellen dabei nicht mehr das geeignete Werkzeug dar, um derartig komplexe Entscheidungsprozesse zu unterstützen. Diese Erkenntnis wird durch ein Vielzahl aktueller Arbeiten national und international belegt, stellvertretend seien hier die Arbeiten [26, 34, 48, 56, 85, 86] genannt.

Ein weiteres Beispiel ist die Ertragsoptimierung bei Flugbuchungen in O&D-Netzwerken eines Luftfahrtunternehmens. Das Netzwerk besteht aus einer Men-

ge von Startpunkten (Origins) und Zielen (Destinations) sowie aus einer Menge von Flügen. Ziel ist eine optimale Buchungssteuerung, die sowohl das unterschiedliche Buchungsverhalten der Fluggäste hinsichtlich Zahlungsbereitschaft und Buchungszeitraum als auch die konkurrierende Nachfrage zwischen den Flugrouten berücksichtigt. Da die Buchungsnachfrage sowie das Stornierungsverhalten mit großen Unsicherheiten behaftet sind, ist eine stochastische Modellierung sinnvoll (vgl. [59, 60]).

1.1 Stochastische Optimierung

Bei der stochastischen Modellierung von Optimierungsaufgaben werden Entscheidungen nicht allein auf Grundlage der deterministischen Daten (d.h. für den Planungszeitraum feststehender Einflussgrößen) getroffen. Stochastische Probleme sind vielmehr dadurch gekennzeichnet, dass bei der Entscheidungsfindung die Stochastik zufälliger Einflussgrößen Berücksichtigung findet. Werden die Eingangsdaten mit Zufallsvariablen abgebildet, so stellen auch die Entscheidungen des Optimierungsproblems im Allgemeinen Zufallsvariablen dar. Trotz der bestehenden Abhängigkeit der Entscheidungen von den zufälligen Eingangsdaten, soll das Optimierungsmodell stets eine eindeutige Entscheidungsgrundlage für den Beginn des Planungszeitraums liefern. Zu diesem Zweck wird häufig die Determiniertheit der sogenannten Erststufenentscheidung im Modell festgelegt. Ziel ist die Bestimmung von Entscheidungen mit hohem Ertragspotenzial zu Beginn des Planungszeitraums. Eine detaillierte Einführung in die stochastische Optimierung vor dem Hintergrund unterschiedlicher Konzepte bieten unter anderem die Arbeiten [2, 53, 54, 67, 80, 83].

In der Praxis werden zwei grundlegend verschiedene Ansätze zur Einbeziehung der Unsicherheiten in die Optimierung verfolgt. Während in der Vergangenheit oft einfachere deterministische Modelle herangezogen wurden, die sich auf die Integration von Prognosen für unsicherheitsbehaftete Daten beschränkten, werden gegenwärtig immer häufiger komplexe stochastische Modelle betrachtet, bei denen sowohl der Wahrscheinlichkeitsraum als auch die Verteilung der Zufallsgrößen einbezogen werden. Sowohl die Zufallsdaten als auch die Entscheidungen werden dann als Elemente eines abstrakten Funktionenraums modelliert und können mehrdimensionale zeitdiskrete stochastische Prozesse darstellen. Der Vorteil der stochastischen Herangehensweise besteht in der hohen Flexibilität, auf Abweichungen von erwarteten Beobachtungen und auf die damit verbundenen Risiken reagieren zu können. Die Konsequenz ist vielfach eine deutlich größere Robustheit dieser Modelle im Vergleich zu den Prognosemodellen (vgl. auch [90]).

Alle stochastischen Ansätze haben eines gemeinsam, sie benötigen Informationen über die Wahrscheinlichkeitsverteilung der Daten. Vernünftige Annahmen hinsichtlich der Verteilungseigenschaften unsicherer Daten zu treffen,

deren Realisierungen bis weit in die Zukunft reichen, stellt Entscheidungsträger vor schwierige Aufgaben. In vielen Fällen werden historischer Daten herangezogen und statistische Analysen vorgenommen, um Schlussfolgerungen bezüglich der Verteilung ziehen zu können. Schließlich wird eine Simulation des Zufalls möglich, was die Bereitstellung unterschiedlicher Realisierungen (Szenarien) erlaubt, die wiederum Eingang in das Optimierungsmodell finden.

Moderne Planungs- und Optimierungsmodelle in der Industrie, in der Finanzwirtschaft und im Handel, die Unsicherheiten und Risiken in der Planung gleichermaßen berücksichtigen, führen in der mathematischen Formulierung häufig zu hochdimensionalen mehrstufigen stochastischen Optimierungsproblemen (siehe zum Beispiel [9, 22, 26, 35, 66, 68]). Die Grundidee des mehr-

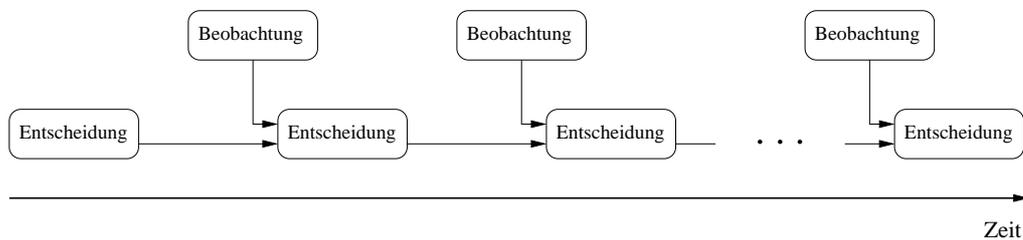


Abbildung 1.1: Illustration der zeitlichen Abfolge von Entscheidungen und Beobachtungen im mehrstufigen stochastischen Optimierungsmodell.

stufigen Problems besteht in einer zeitlichen Abfolge von Entscheidungen und Beobachtungen über einen längeren Zeitraum (siehe Abbildung 1.1). Das Wechselspiel zwischen Beobachtungen und Entscheidungen bietet die Möglichkeit, durch Kompensationsentscheidungen auf neue Ereignisse zu reagieren und negative Auswirkungen gegebenenfalls zu korrigieren. Modelle dieser Art werden auch als Recourse-Modelle bezeichnet. Eine zentrale Bedeutung hat dabei die Generierung von Szenariobäumen, was unter anderem auch durch eine Vielzahl aktueller Arbeiten belegt wird (siehe z.B. [6, 38, 47]).

1.2 Gliederung der Arbeit

Im Vordergrund dieser Arbeit steht die Modellierung von Szenariobäumen als geeignete Form der Abbildung von Unsicherheiten in zwei- und mehrstufigen (linearen) stochastischen Optimierungsproblemen. Behandelt werden sowohl algorithmische Aspekte als auch theoretische Grundlagen. Ein wesentlicher Bestandteil ist die Motivation durch aktuelle Stabilitätsresultate angepasst an die jeweils betrachteten Optimierungsmodelle.

In den folgenden Kapiteln 2 und 3 werden die wesentlichen mathematischen Grundlagen der betrachteten Klassen von stochastischen Optimierungsmodellen sowie die Grundbegriffe und Eigenschaften verwendeter Wahrscheinlich-

keitsabstände dargestellt. Darauf aufbauend werden im Kapitel 4 grundlegende Stabilitätsaussagen bewiesen, die im mehrstufigen Fall sowohl den Abstand der Wahrscheinlichkeitsverteilungen als auch einen gewissen Filtrationsabstand zur Verifizierung von Informationsstrukturen einbeziehen. In Kapitel 6 werden nach einer Diskussion von Szenarioreduktionstechniken in Kapitel 5, die eine Erweiterung der in [36] beschriebenen Algorithmen beinhaltet, zwei verschiedene Varianten von heuristischen Verfahren zur Szenariobaumkonstruktion beschrieben, die auf einer schrittweisen Szenarioreduktion und einem damit verbundenen zeitabhängigen Clustering basieren. Das Clustering erfolgt dabei einmal rückwärts (Algorithmus 6.2) und einmal vorwärts (Algorithmus 6.5) in der Zeit. Die Stabilitätsresultate in Kapitel 4 erlauben eine fundierte Rechtfertigung der beiden Algorithmen. Der zweite Teil von Kapitel 6 widmet sich ausführlich allgemeiner Konvergenzaussagen. Schließlich werden anhand von mehreren Anwendungsbeispielen aus Energiewirtschaft und Airline-Revenue-Management in Kapitel 7, die Erweiterung der Szenarioreduktion sowie die Konstruktionsverfahren zur Generierung von Szenariobäumen hinsichtlich ihrer numerischen Eigenschaften untersucht.

Kapitel 2

Das mehrstufige stochastische Optimierungsmodell

Mehrstufige stochastische Optimierungsmodelle kommen in der Regel immer dann zum Einsatz, wenn die Optimierungsaufgabe einen Entscheidungsprozess unter dem Einfluss einer Reihe von Unsicherheiten über einen längeren Zeitraum erlaubt. Probleme dieser Art treten besonders häufig in der Finanzwirtschaft, in vielen Teilen der Industrie, in der Energiewirtschaft und in der Logistik auf. Zum derzeitigen Stand von Theorie und Lösungsmethoden für mehrstufige Optimierungsprobleme sei an dieser Stelle auf die früheren Arbeiten von Dantzig [13, 14], auf die neueren Bücher [83] und [93] sowie auf die Monografie [53] verwiesen.

2.1 Stochastische Prozesse

Im mehrstufigen Optimierungsproblem bilden sowohl die Eingangsdaten als auch die Stufenentscheidungen zeitdiskrete (multivariate) stochastische Prozesse auf einem zugrundeliegenden Wahrscheinlichkeitsraum. Bezeichnet man den Wahrscheinlichkeitsraum mit $(\Omega, \mathcal{F}, \mathbb{P})$ sowie den stochastischen Datenprozess mit $\xi = (\xi_1, \dots, \xi_T)$, so definiert ξ zu jedem Zeitpunkt t eine Zufallsvariable ξ_t , die jeweils Werte in \mathbb{R}^d annimmt. Somit beschreibt jede Komponente ξ_t eine messbare Abbildung

$$\xi_t : (\Omega, \mathcal{F}_t(\xi), \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}), \quad t = 1, \dots, T,$$

wobei $\{\mathcal{F}_t(\xi)\}_{t=1}^T$ eine Folge von σ -Algebren mit der Eigenschaft $\mathcal{F}_t(\xi) \subseteq \mathcal{F}$ ist und \mathcal{B} die Menge aller Borelmengen in \mathbb{R}^d bezeichnet.

Die Entscheidungen im Optimierungsproblem hängen wiederum vom Datenprozess ξ ab. Betrachtet man den gesamten Planungszeitraum des mehrstufigen Optimierungsproblems, so erscheint es als sinnvoll, den Entscheidungsprozess so zu gestalten, dass Entscheidungen zu einem beliebigen Zeitpunkt nur von der

jeweils verfügbaren Information abhängen. Diese Eigenschaft wird auch *Nicht-antizipativität* genannt. Bezeichnet man mit $\xi^t = (\xi_1, \dots, \xi_t)$ den Vektor aller Beobachtungen bis zum Zeitpunkt t und mit $x = (x_1, \dots, x_T)$ den gesamten Entscheidungsprozess, so bedeutet die Nichtantizipativität mathematisch, dass die Entscheidung x_t zum Zeitpunkt t nur vom Vektor ξ^t abhängen darf. Das heißt, für die Daten- und Entscheidungsprozesse des Optimierungsproblems ergibt sich eine sequenzielle Abhängigkeit der Form

$$x_1(\xi^1) \hookrightarrow x_2(x_1; \xi^2) \hookrightarrow \dots \hookrightarrow x_T(x_1, \dots, x_{T-1}; \xi^T). \quad (2.1)$$

Die Eigenschaft der Nichtantizipativität ist gleichbedeutend mit der Messbarkeit der sogenannten Stufenentscheidung x_t bezüglich der von ξ^t generierten σ -Algebra $\mathcal{F}_t(\xi)$. Damit definiert der Eingangsprozess eine gewisse *Informationsstruktur* bzw. *Filtration*, die durch folgende Inklusionseigenschaft der σ -Algebren $\mathcal{F}_t(\xi)$ gegeben ist:

$$\mathcal{F}_1(\xi) \subseteq \mathcal{F}_2(\xi) \subseteq \dots \subseteq \mathcal{F}_T(\xi). \quad (2.2)$$

Eine solche Filtrationseigenschaft beschreibt eine mit fortlaufender Zeit zunehmende Verfeinerung der σ -Algebren. Jede Verfeinerung bedeutet dabei einen Zuwachs an Information.

Um eine numerische Behandelbarkeit des Optimierungsproblems gewährleisten zu können, wird der dem Eingangsprozess zugrundeliegende Wahrscheinlichkeitsraum, auch wenn dieser im Allgemeinen nicht endlich ist, häufig durch einen solchen ersetzt. Eine entsprechende Approximation besteht aus endlich

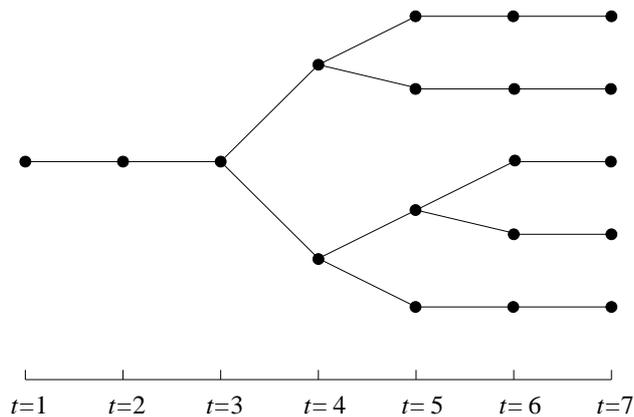


Abbildung 2.1: Illustration der Struktur eines Szenariobaums.

vielen Szenarien versehen mit gewissen Eintrittswahrscheinlichkeiten. Von elementarer Bedeutung ist dabei, dass auch die Informationsstruktur der Ausgangsdaten näherungsweise wiedergegeben wird. Die Konsequenz ist die Anordnung der Szenarien in Form eines Szenariobaums, in dem jede Verzweigung

von Szenarien einer Zunahme an Information entspricht. Abbildung 2.1 zeigt ein Beispiel eines Szenariobaums mit 5 Szenarien und einem Zeitraum unterteilt in $T = 7$ Stufen bzw. Perioden.

2.1.1 Beispiel bivariater Last-Preisprozess

In den Modellen der Energieoptimierung gehören zu den wichtigsten stochastischen Komponenten des Eingangsprozesses elektrische Last, Spotpreise, Zuflüsse und Brennstoffpreise. Exemplarisch zeigen die Abbildungen 2.2 und 2.3

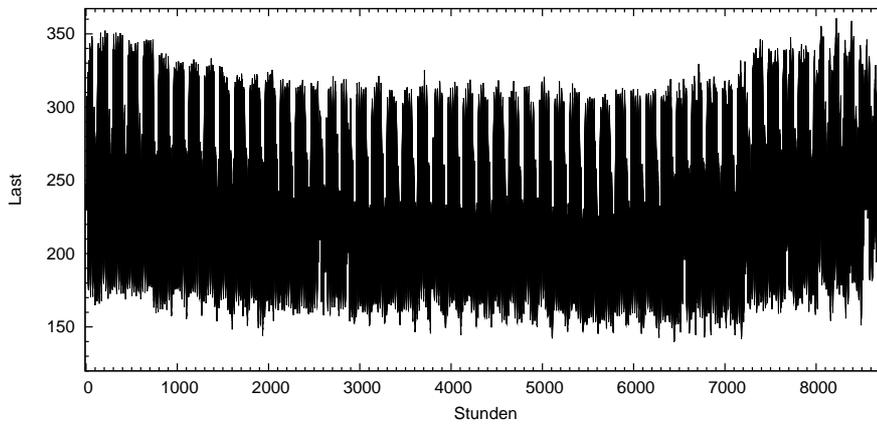


Abbildung 2.2: Typische Lastverläufe für ein Kalenderjahr.

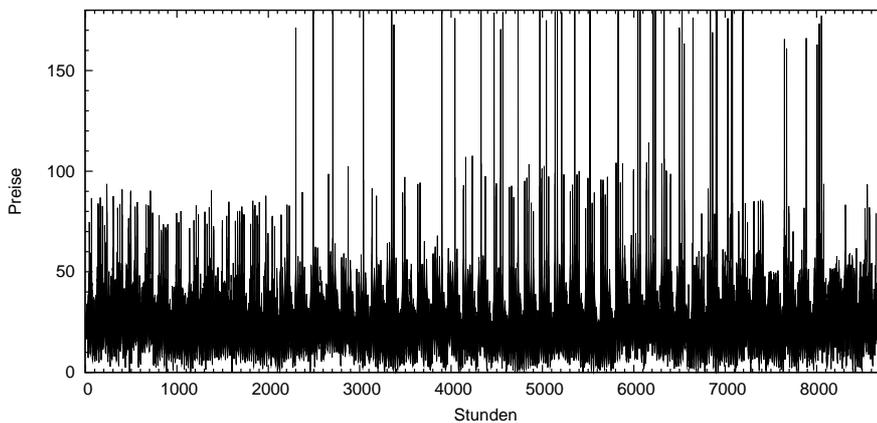


Abbildung 2.3: Typische Spotpreisverläufe für ein Kalenderjahr

charakteristische, auf historischen Daten basierende Szenarienverläufe eines bivariaten Last- und Preisprozesses innerhalb eines Kalenderjahres. Typisch für den Lastgang (obere Abbildung) sind die Tages- und Wochenzyklen sowie die saisonalen Verschiebungen im Sommer und Winter. Betrachtet man die Spot-

preisszenarien, so fallen besonders die über das gesamte Jahr verteilten außergewöhnlichen Preisspitzen auf, die das hohe Maß an Volatilität von Preisprozessen unterstreichen. Eine Modellierung solcher Prozesse erweist sich daher als besonders schwierig [5, 22, 48].

2.2 Das lineare mehrstufige Problem

Das mehrstufige (lineare) stochastische Optimierungsproblem umfasst einen (multivariaten) stochastischen Eingangsprozess $\{\xi_t\}_{t=1}^T$, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, wobei ξ_t Werte in \mathbb{R}^d annimmt sowie Stufenentscheidungen x_t , die zu \mathbb{R}^{m_t} gehören und zum Zeitpunkt t nur von den Realisierungen $\xi^t = (\xi_1, \dots, \xi_t)$ abhängen. Diese Nichtantizipativitätseigenschaft ist äquivalent zur Messbarkeit der Entscheidung x_t bezüglich der σ -Algebra $\mathcal{F}_t(\xi) \subseteq \mathcal{F}$, welche von (ξ_1, \dots, ξ_t) erzeugt wird (vgl. Abschnitt 2.1). Hinzu kommt die Annahme, dass zum Zeitpunkt $t = 1$ die entsprechenden Eingangsgrößen bekannt sein mögen, d.h. als deterministisch vorausgesetzt werden können. Das bedeutet, $\mathcal{F}_1(\xi)$ ist die triviale σ -Algebra $\mathcal{F}_1 = \{\emptyset, \Omega\}$. Diese Annahme führt zu einer ebenfalls deterministischen und damit wohlbestimmten Erststufenentscheidung x_1 . Das lineare mehrstufige stochastische Optimierungsproblem lässt sich so in folgender Form formulieren:

$$\min \left\{ \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle \right] \left| \begin{array}{l} x_t \in X_t, t = 1, \dots, T, A_{1,0}x_1 = h_1(\xi_1), \\ x_t \text{ ist } \mathcal{F}_t(\xi)\text{-messbar, } t = 1, \dots, T, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, \dots, T \end{array} \right. \right\}. \quad (2.3)$$

Die Restriktionsmengen $X_t \subseteq \mathbb{R}^{m_t}$ sind dabei polyedrische Kegel, die Koeffizienten der Zielfunktion $b_t(\xi_t)$ und die rechten Seiten $h_t(\xi_t)$ gehören zu \mathbb{R}^{m_t} bzw. \mathbb{R}^{n_t} . Die sogenannten *Recourse*-Matrizen $A_{t,0}$ und die sogenannten *Technologie*-Matrizen $A_{t,1}(\xi_t)$ sind (n_t, m_t) - bzw. (n_t, m_{t-1}) -Matrizen. Die Kosten $b_t(\cdot)$, die Technologie-Matrizen $A_{t,1}(\cdot)$ sowie die rechten Seiten $h_t(\cdot)$ dürfen affin linear vom Zufallsvektor ξ_t abhängen, was insbesondere die Spezialfälle nur stochastische rechte Seiten bzw. nur stochastische rechte Seiten und Kosten einschließt.

Die erste und dritte Gruppe der Restriktionen in (2.3) sind standard für stochastische Optimierungsprobleme. Die Bedingungen sind punktweise mit Wahrscheinlichkeit 1 zu erfüllen. Anders verhält es sich mit der zweiten Gruppe von Restriktionen. Die Messbarkeitsrestriktionen oder auch *Informationsrestriktionen* stellen funktionale und nicht-punktweise Bedingungen dar, sobald $T > 2$ ist und $\mathcal{F}_2 \subsetneq \mathcal{F}_t \subseteq \mathcal{F}$ für ein t mit $2 < t \leq T$. Das gleichzeitige Vorhandensein dieser zwei unterschiedlichen Typen von Restriktionen sorgt für die besondere Kompliziertheit von mehrstufigen stochastischen Optimierungsproblemen. Das wirkt sich sowohl auf theoretische als auch algorithmischen Fragestellungen aus. Untersuchungen in [88] und [89] kommen zu dem Ergebnis, dass mehrstufige

stochastische Probleme im Allgemeinen eine deutlich höhere Komplexität im Vergleich zu zweistufigen Modellen besitzen.

Für viele Anwendungen wächst die Bedeutung der Minimierung von Risiken neben der Ertrags- bzw. Kostenoptimierung. So existieren eine Reihe von Ansätze, Risikofunktionale in das stochastische Optimierungsmodell einzubeziehen. Besonders angepasste Risikofunktionale (vgl. [21] und [68]) erlauben eine Formulierung von Risiko-orientierten Modellen als mehrstufige stochastisches Problem der Form (2.3). Die traditionelle Risikobeschreibung über den *Value-at-Risk* (*VaR*) führt in vielen Modellen zu einer Formulierung als stochastisches Optimierungsproblem mit Wahrscheinlichkeitsrestriktionen (siehe [29] und [44]). In der aktuellen Arbeit [42] sind einige Anmerkungen zum VaR in der Optimierung zu finden.

2.2.1 Modellierung des Datenprozesses

Die verbreitetste Form der Modellierung mehrstufiger stochastischer Optimierungsprobleme besteht in der Approximation des stochastischen Datenprozesses $\xi = \{\xi_t\}_{t=1}^T$ durch endlich viele Szenarien. Die Szenarien werden dabei so strukturiert, dass sie einen Szenariobaum bilden (siehe Abschnitt 2.1). Auf diese Weise gelangt man ausgehend von Problem (2.3) zu oft sehr großen linearen Optimierungsproblemen, die mit Standardtechniken der linearen Optimierung gelöst werden können. Zu den Standardtechniken gehören beispielsweise Dekompositionsmethoden, die besondere Strukturen des Modells ausnutzen und damit eine effiziente Lösung ermöglichen. Für eine genauere Studie sei hier auf [83, Chapter 3] verwiesen.

Gegenwärtig werden verschiedene Ansätze zur Generierung von Szenariobäumen für mehrstufige stochastische Optimierungsprobleme verfolgt, die auf unterschiedlichen Prinzipien basieren. Eine Übersicht dazu ist in [18] zu finden. Die wichtigsten Konzepte sind

- (i) die sogenannten Bound-Based-Konstruktionen (vgl. [6, 20, 28, 57]),
- (ii) die Monte-Carlo- und Quasi-Monte-Carlo-Methoden (siehe [8, 87, 89] bzw. [64, 65]),
- (iii) die EVPI-Sampling-Methoden innerhalb von Dekompositionsalgorithmen (vgl. [11, 16, 45, 52]),
- (iv) die Momenten-Methoden (siehe [50, 51]) sowie
- (v) die Approximationen bzgl. geeigneter Wahrscheinlichkeitsmetriken (siehe [33, 37, 46, 47, 66]).

Die Bound-Base-Methoden (i) basieren auf der Konstruktion von diskreten Wahrscheinlichkeitsmaßen, die unter gewissen Voraussetzungen an das Modell

und den Eingangsprozess, zu oberen und unteren Schranken des Problems führen, dabei aber gleichzeitig eine numerische Behandelbarkeit gewährleisten. Geeignete Verfeinerungsstrategien ermöglichen eine schrittweise Verbesserung der Approximationen. Die aktuelle Arbeit [6] bzw. Monographie [57] bieten darüber hinaus Konvergenzresultate im Fall linearer Modelle mit stochastischen rechten Seiten bzw. konvexer Modelle, deren Stochastizität durch gewisse lineare block-diagonale autoregressive Prozesse mit kompaktem Träger beschrieben werden kann.

Die Monte-Carlo-Methoden in (ii) verwenden bedingte Sampling-Algorithmen und (Pseudo-)Zufallszahlengeneratoren für bedingte Verteilungen. Konsistenzeigenschaften wurden in [87] gezeigt, Untersuchungen zur Komplexität findet man in [88]. Die Quasi-Monte-Carlo-Methoden in [64, 65] wurden für konvexe Modelle entwickelt, deren stochastische Eingangsprozesse sich als Zeitreihen mit unabhängigen und gleichverteilten Erzeugern darstellen lassen. Während die allgemeine Theorie von *epi-konvergenten* Diskretisierungen in [64] auf das bedingte Sampling übertragbar ist, wurde in [65] ein generelles Verfahren zur Konstruktion von Szenariobäumen für die auf Zeitreihen basierenden Modelle entwickelt, welches auf der Approximation jeder der unabhängigen gleichverteilten Zufallsvariablen mit Quasi-Monte-Carlo-Methoden beruht (siehe [62]). Der Grund für die Verwendung von Quasi-Monte-Carlo-Methoden sind die guten Konvergenzeigenschaften und die Tatsache, Integrale großer Dimensionen in moderater Zeit berechnen zu können. Dagegen ist die Erzeugung von zufälligen Stichproben generell schwierig (vgl. [62, Seite 7]).

Die Momenten-Methoden (iii) basieren auf der Approximation von Wahrscheinlichkeitsverteilungen auf die Art, dass gewisse Momenten-Bedingungen erfüllt werden. Allerdings sind solche Bedingungen in der Regel für die Konsistenz der Approximation allein nicht hinreichend.

Die Methoden in (iv) beruhen auf der Integration von Sampling-Algorithmen in das Lösungsverfahren. Anstelle der Verwendung einer festvorgegebenen Approximation erfolgt das Generieren, Streichen oder Modifizieren von Szenarien innerhalb des verwendeten Dekompositionsalgorithmus zur Lösung des stochastischen Optimierungsproblems.

Die Herangehensweise in (v) ist die in dieser Arbeit verfolgte Strategie zur Konstruktion von Szenariobäumen. Sie basiert auf Wahrscheinlichkeitsmetriken, die hinsichtlich der Stabilität von mehrstufigen Modellen relevant sind. Die Arbeiten [33, 47, 66] beschäftigen sich mit den Fortet-Mourier- und Wasserstein-Metriken, die eine wesentliche Rolle bei zwei- und mehrstufigen stochastischen Optimierungsproblemen spielen. Diese Metriken stehen im Mittelpunkt von Kapitel 3 dieser Arbeit. Die Stabilitätsresultate werden in Kapitel 4 bewiesen.

Schließlich sei an dieser Stelle auch auf die Arbeiten [18, 55] verwiesen, die sich mit der Frage nach der Bewertung von Szenariobäumen auseinandersetzen.

Kapitel 3

Verteilungen und Wahrscheinlichkeitsabstände

Befasst man sich im Rahmen stochastischer Fragestellungen mit Störungstheorie, so benötigt man ein Maß zur Beurteilung von Abständen zwischen Wahrscheinlichkeitsverteilungen, welches Nähe von Szenarien zueinander genauso wie deren Eintrittswahrscheinlichkeiten berücksichtigt. Seit dem frühen Beginn der Stabilitätsuntersuchungen stochastischer Optimierungsprobleme stehen zu diesem Zweck gewisse Wahrscheinlichkeitsfunktionale im Mittelpunkt der Betrachtungen.

3.1 Wahrscheinlichkeitsfunktionale

Die Beschreibung der verwendeten Wahrscheinlichkeitsabstände beginnt mit den Monge-Kantorovich- sowie den Kantorovich-Rubinstein-Funktionalen. Es sei $\mathcal{P}(\Xi)$ die Menge aller Borel-Maße auf einer abgeschlossenen Teilmenge Ξ in \mathbb{R}^s . Mit der Notation in [69, 71] sind auf der Menge $\mathcal{P}(\Xi) \times \mathcal{P}(\Xi)$ das Monge-Kantorovich-Funktional durch

$$\hat{\mu}_c(P, Q) := \inf \left\{ \int_{\Xi \times \Xi} c(\xi, \tilde{\xi}) \eta(d\xi, d\tilde{\xi}) : \right. \\ \left. \eta \in \mathcal{P}(\Xi \times \Xi), \pi_1 \eta = P, \pi_2 \eta = Q \right\} \quad (3.1)$$

und das Kantorovich-Rubinstein-Funktional durch

$$\overset{\circ}{\mu}_c(P, Q) := \inf \left\{ \int_{\Xi \times \Xi} c(\xi, \tilde{\xi}) \eta(d\xi, d\tilde{\xi}) : \right. \\ \left. \eta \in \mathcal{M}(\Xi \times \Xi), \pi_1 \eta - \pi_2 \eta = P - Q \right\} \quad (3.2)$$

definiert. Die Menge $\mathcal{M}(\Xi \times \Xi)$ sei die Menge aller endlichen Maße auf $\Xi \times \Xi$ und π_1 bzw. π_2 bezeichnen die jeweilige Projektion auf die erste und zweite

Komponente von η . Die Kostenfunktion c in beiden Beschreibungen (3.1) und (3.2) sei gegeben. Sie definiert im Allgemeinen einen Pseudoabstand der Szenarien im Raum \mathbb{R}^s . Der wesentliche Unterschied beider Abstände liegt in der Möglichkeit der Wahl der zugelassenen Maße in der Restriktionmenge. Während beim Monge-Kantorovich-Funktional nur über Wahrscheinlichkeitsmaße minimiert wird, deren Marginal-Verteilungen genau P und Q ergeben, werden beim Kantorovich-Rubinstein-Funktional alle endlichen Maße zugelassen, deren Marginalverteilungsdifferenz der Differenz von P und Q entspricht.

3.1.1 Die Kostenfunktion

Die Kostenfunktion c bietet die Möglichkeit, durch spezielle Wahl, einen dem Problem angepassten Abstand zu definieren. Einige grundlegende Eigenschaften werden für alle zulässigen Kostenfunktionen gefordert.

Es wird im Folgenden angenommen, dass die Kostenfunktion $c : \Xi \times \Xi \rightarrow \mathbb{R}$ eine nichtnegative, symmetrische und stetige Abbildung definiert, die darüberhinaus folgende Eigenschaften erfüllt:

- (i) $c(\xi, \tilde{\xi}) = 0$ gilt genau dann, wenn $\xi = \tilde{\xi}$,
- (ii) $\sup_{\delta \rightarrow 0} \left\{ c(\xi, \tilde{\xi}) : \xi, \tilde{\xi} \in B, \|\xi - \tilde{\xi}\| \leq \delta \right\} = 0$ für jedes beschränkte $B \subseteq \Xi$,
- (iii) es existiert eine Abschätzung der Form

$$c(\xi, \tilde{\xi}) \leq \lambda_c(\xi) + \lambda_c(\tilde{\xi}) \quad \text{für alle } \xi, \tilde{\xi} \in \Xi$$

mit einer messbaren Abbildung $\lambda_c : \Xi \rightarrow \mathbb{R}_+$, die beschränkten Mengen aus Ξ auf beschränkte Mengen von \mathbb{R}_+ abbildet.

Mit diesen Eigenschaften sind beide Funktionale (3.1) und (3.2) endlich, falls die Wahrscheinlichkeitsmaße P und Q zu folgender Menge gehören:

$$\mathcal{P}_c(\Xi) := \left\{ Q \in \mathcal{P}(\Xi) : \int_{\Xi} \lambda_c(\xi) Q(d\xi) < \infty \right\}. \quad (3.3)$$

Ein wichtiges Beispiel für Kostenfunktionen, die die geforderten Eigenschaften erfüllen, sind Funktionen des Typs

$$c_r(\xi, \tilde{\xi}) := \max\{1, \|\xi - \xi_0\|^{r-1}, \|\tilde{\xi} - \xi_0\|^{r-1}\} \|\xi - \tilde{\xi}\| \quad (\xi, \tilde{\xi} \in \Xi), \quad (3.4)$$

wobei $r \geq 1$ gelte und $\xi_0 \in \Xi$ ein fixiertes Element aus der Trägermenge ist. In diesem Fall setzt man

$$\lambda_{c_r}(\xi) := 2\|\xi - \xi_0\| \max\{1, \|\xi - \xi_0\|^{r-1}\}$$

und man erhält als Menge $\mathcal{P}_{c_r}(\Xi)$ die Menge aller Wahrscheinlichkeitsmaße, die absolute Momente bis zur Ordnung r besitzen. Zur Vereinfachung der Schreibweise sei

$$\mathcal{P}_r(\Xi) := \mathcal{P}_{c_r}(\Xi), \quad \overset{\circ}{\mu}_r := \overset{\circ}{\mu}_{c_r} \quad \text{und} \quad \hat{\mu}_r := \hat{\mu}_{c_r}.$$

Neben der Kostenfunktion (3.4) ist in Hinblick auf die Stabilität mehrstufiger stochastischer Optimierungsprobleme eine weitere Kostenfunktion von Bedeutung. Sie ist über die Norm definiert und besitzt die Form

$$c(\xi, \tilde{\xi}) := \|\xi - \tilde{\xi}\|^r, \quad (3.5)$$

wobei erneut $r \geq 1$ ist und r dieselbe Rolle wie in (3.4) übernimmt. Als Konsequenz ergibt sich für $\mathcal{P}_c(\Xi)$ auch hier die Menge aller Wahrscheinlichkeitsmaße mit absoluten Momenten bis zur Ordnung r . Die Kosten (3.5) führen zu den sogenannten *L_r -minimalen Metriken* ℓ_r (siehe [72]) und besitzen die Gestalt

$$\ell_r(P, Q) := \inf \left\{ \int_{\Xi \times \Xi} \|\xi - \tilde{\xi}\|^r \eta(d\xi, d\tilde{\xi}) : \right. \\ \left. \eta \in \mathcal{P}(\Xi \times \Xi), \pi_1 \eta = P, \pi_2 \eta = Q \right\}^{\frac{1}{r}}. \quad (3.6)$$

Die Metriken ℓ_r werden auch *Wasserstein-Metriken* der Ordnung r genannt (vgl. [30]) und es gilt $\ell_r(P, Q) = (\hat{\mu}_c(P, Q))^{\frac{1}{r}}$, wenn c die Form (3.5) besitzt.

3.1.2 Eigenschaften

Die folgenden dualen Darstellungen und Charakterisierungen der Monge-Kantorovich- und Kantorovich-Rubinstein-Funktionale sind von besonderem Interesse. Die folgenden Resultate werden in [73, Section 2], [71, Section 4.3] und [69, Theorems 6.2.5 and 6.3.1] bewiesen:

Proposition 3.1 *Für alle Wahrscheinlichkeitsmaße $P, Q \in \mathcal{P}_c(\Xi)$ besitzt das Kantorovich-Rubinstein-Funktional $\overset{\circ}{\mu}_c$ in (3.2) die duale Darstellung*

$$\overset{\circ}{\mu}_c(P, Q) = \sup_{f \in \mathcal{F}_c} \left| \int_{\Xi} f(\xi) P(d\xi) - \int_{\Xi} f(\xi) Q(d\xi) \right|, \quad (3.7)$$

wobei \mathcal{F}_c die Klasse aller Funktionen $f : \Xi \rightarrow \mathbb{R}$ bezeichnet, für die die Eigenschaft $f(\xi) - f(\tilde{\xi}) \leq c(\xi, \tilde{\xi})$ für alle $\xi, \tilde{\xi} \in \Xi$ gilt.

Proposition 3.2 *Es sei Ξ kompakt. Jedes Kantorovich-Rubinstein-Funktional mit einer Kostenfunktion c ist gleich dem Monge-Kantorovich-Funktional mit reduzierten Kosten \hat{c} . Das heißt, es gilt*

$$\overset{\circ}{\mu}_c(P, Q) = \overset{\circ}{\mu}_{\hat{c}}(P, Q) = \hat{\mu}_{\hat{c}}(P, Q) \leq \hat{\mu}_c(P, Q), \quad (3.8)$$

wobei die reell-wertige Funktion \hat{c} auf $\Xi \times \Xi$ durch

$$\hat{c}(\xi, \tilde{\xi}) := \inf \left\{ \sum_{i=1}^{n-1} c(\xi_i, \xi_{i+1}) : n \in \mathbb{N}, \xi_i \in \Xi, \xi_1 = \xi, \xi_n = \tilde{\xi} \right\} \quad (3.9)$$

definiert ist. Die Funktion \hat{c} ist eine Metrik auf Ξ mit der Eigenschaft $\hat{c} \leq c$ und ist identisch mit c , falls c die Dreiecksungleichung erfüllt.

Proposition 3.3 *Das Kantorovich-Rubinstein-Funktional $\hat{\mu}_r$ definiert eine Metrik auf der Menge $\mathcal{P}_r(\Xi)$ und die Abschätzung*

$$\left| \int_{\Xi} \|\xi\|^r P(d\xi) - \int_{\Xi} \|\xi\|^r Q(d\xi) \right| \leq r \hat{\mu}_r(P, Q) \quad (3.10)$$

ist gültig für alle $P, Q \in \mathcal{P}_r(\Xi)$. Die Konvergenz einer Folge (P_n) von Wahrscheinlichkeitsmaßen gegen einen Grenzwert P in diesem metrischen Raum ist äquivalent zu jeder der folgenden Bedingungen:

- (a) $(\hat{\mu}_r(P_n, P))$ konvergiert gegen 0 für $n \rightarrow \infty$,
- (b) (P_n) konvergiert schwach gegen P und die absoluten Momente der Ordnung r von P_n konvergieren gegen die von P .

Der Abstand $\hat{\mu}_r$ wird auch *Fortet-Mourier-Metrik* der Ordnung r genannt. Solche Metriken und deren Eigenschaften wurden als erstes in [27] studiert und ursprünglich über ihre dualen Darstellungen (3.7) beschrieben. Die Kompaktheitsforderung in Proposition 3.2 ist keine Einschränkung, wenn das Resultat auf Wahrscheinlichkeitsverteilungen mit endlicher Trägermenge angewendet wird. Die Bedeutung für diese Arbeit liegt in der Tatsache begründet, dass sowohl die Monge-Kantorovich-Funktionale, hier insbesondere die L_r -minimalen Metriken, als auch die Kantorovich-Rubinstein-Funktionale in Stabilitätsresultaten (siehe Kapitel 4) Verwendung finden. Darüber hinaus erlauben die Monge-Kantorovich-Funktionale, über ihre Formulierung als Transportprobleme ganz spezielle Darstellungen (siehe Kapitel 5).

3.2 Endliche Verteilungen

In den meisten praktischen Anwendungen verwendet man endliche Approximationen. In dem Fall ist die Trägermenge der Verteilung endlich, wodurch sich die Formeln zur Beschreibung der Wahrscheinlichkeitsabstände vereinfachen. Im Folgenden seien P und Q endliche Maße gegeben durch ihre Szenarien

$$\text{supp}(P) = \{\xi^1, \dots, \xi^N\} \quad \text{und} \quad \text{supp}(Q) = \{\tilde{\xi}^1, \dots, \tilde{\xi}^M\}$$

sowie den zugehörigen Wahrscheinlichkeiten

$$p_i = P(\{\xi^i\}), \quad i = 1, \dots, N \quad \text{und} \quad q_j = Q(\{\tilde{\xi}^j\}), \quad j = 1, \dots, M.$$

3.2.1 Monge-Kantorovich-Funktional

Das Monge-Kantorovich-Funktional (3.1) nimmt im diskreten Fall die Form

$$\hat{\mu}_c(P, Q) = \inf \left\{ \sum_{i=1}^N \sum_{j=1}^M c(\xi^i, \tilde{\xi}^j) \eta_{ij} \mid \eta_{ij} \geq 0, \sum_{j=1}^M \eta_{ij} = p_i, \sum_{i=1}^N \eta_{ij} = q_j \right\} \quad (3.11)$$

an. Der Monge-Kantorovich-Abstand zweier diskreter Maße ist somit der Optimalwert eines klassischen linearen (*Massen-Transportproblems*). Der dem Problem zugrunde liegende gerichtete Graph zur Lösung des Problems ist in Abbildung 3.1 dargestellt.

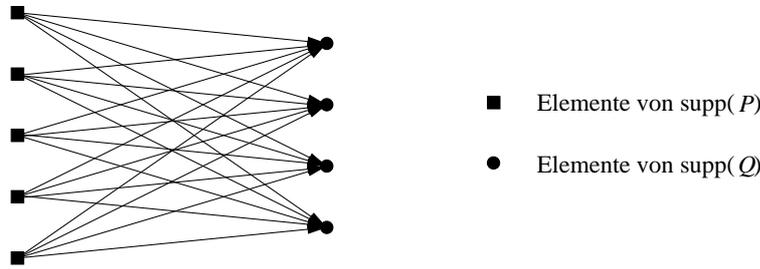


Abbildung 3.1: Das Transportproblem besteht in der Bestimmung eines kostenminimalen Weges für die Wahrscheinlichkeitsgewichte von P nach Q , wobei nur direkte Verbindungen zwischen Atomen von P und Q erlaubt sind.

Die L_r -minimalen bzw. Wasserstein-Metriken ergeben sich auch hier als Spezialfall des Monge-Kantorovich-Abstands entsprechend, wenn für die Kostenfunktion die jeweilige Potenz des Normabstands eingesetzt wird (vgl. Abschnitt 3.1.1).

3.2.2 Kantorovich-Rubinstein-Funktional

Zur Beschreibung des Kantorovich-Rubinstein-Abstands seien die Szenarien von Q durch $\xi^{N+j} := \tilde{\xi}^j$ für $j = 1, \dots, M$ identifiziert. Damit nimmt das Funktional (3.2) im diskreten Fall die folgende Form an:

$$\begin{aligned} \overset{\circ}{\mu}_c(P, Q) = \inf \left\{ \sum_{i,j=1}^{N+M} c(\xi^i, \xi^j) \eta_{ij} \mid \right. \\ \left. \eta_{ij} \geq 0, \sum_{j=1}^{N+M} \eta_{ij} - \sum_{j=1}^{N+M} \eta_{ji} = P(\{\xi^i\}) - Q(\{\xi^i\}) \right\}. \quad (3.12) \end{aligned}$$

Der Kantorovich-Rubinstein-Abstand zweier diskreter Maße ist damit der Optimalwert eines Netzwerkfluss-Problems vom Typ *Minimum-Cost-Flow*. Die Ver-

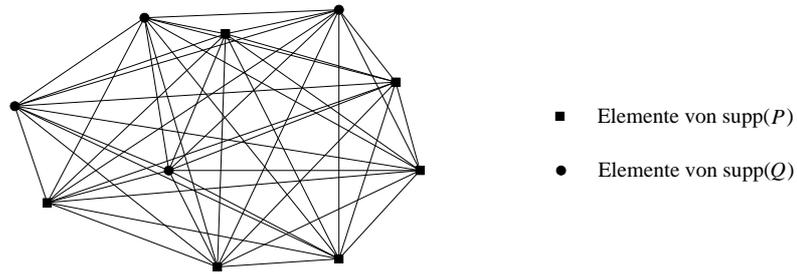


Abbildung 3.2: Das Minimum-Cost-Flow-Problem besteht in der Bestimmung eines kostenminimalen Weges für die Wahrscheinlichkeitsgewichte von P nach Q , wobei alle Verbindungen zwischen Atomen aus P und Q erlaubt sind.

anschaulichung des Netzwerkgraphens zur Lösung dieses Problemtyps ist in Abbildung 3.2 dargestellt. Der Unterschied zum Transportproblem (3.11) besteht neben den veränderten Flussbedingungen in der zusätzlichen Freiheit bei der Bestimmung kostenoptimaler Wege im Netzwerk. Beim Minimum-Cost-Flow-Problem ist jeder Weg zur Verbindung zweier Knoten zugelassen.

Kapitel 4

Stabilität

Stabilitätsaussagen geben Antwort auf die so wichtige Frage nach der Abhängigkeit der Optimalwerte bzw. der optimalen Entscheidungen von Störungen innerhalb der Problemdaten. In der stochastischen Optimierung sind insbesondere die Auswirkungen des Übergangs zu Näherungen hinsichtlich des stochastischen Eingangsdatenprozesses von Bedeutung. In diesem Kapitel werden allgemeine Aussagen für mehrstufige Optimierungsprobleme hergeleitet und daran anschließend spezielle Ergebnisse für zweistufige Probleme diskutiert. Einige der Aussagen sind auch in den aktuellen Arbeiten [41, 39] und [40] zu finden.

4.1 Mehrstufige Probleme

Unter einfachen Zusatzvoraussetzungen kann das lineare mehrstufige stochastische Optimierungsproblem (2.3) in Abschnitt 2.2 äquivalent als Minimumproblem für die deterministische Erststufenentscheidung x_1 umformuliert werden (siehe z.B. [83, Chapter 1] oder [23, 76]). Es besitzt dann die Gestalt

$$\min \left\{ \mathbb{E}[f(x_1, \xi)] = \int_{\Xi} f(x_1, \xi) P(d\xi) : x_1 \in X_1 \right\}, \quad (4.1)$$

wobei Ξ eine abgeschlossene Teilmenge im \mathbb{R}^{Td} ist, die den Träger der Wahrscheinlichkeitsverteilung P von ξ enthält. Des Weiteren ist f ein Integrand auf $\mathbb{R}^{m_1} \times \Xi$, der durch folgende dynamische Rekursion definiert wird:

$$\begin{aligned} f(x_1, \xi) &:= \Phi_1(x_1, \xi^1) = \langle b_1(\xi_1), x_1 \rangle + \Phi_2(x_1, \xi^2), \\ \Phi_t(x_1, \dots, x_{t-1}, \xi^t) &:= \inf \left\{ \langle b_t(\xi_t), x_t \rangle + \mathbb{E} [\Phi_{t+1}(x_1, \dots, x_t, \xi^{t+1}) | \mathcal{F}_t(\xi)] : \right. \\ &\quad \left. x_t \in X_t, x_t \text{ ist } \mathcal{F}_t(\xi)\text{-messbar}, \right. \\ &\quad \left. A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \right\} \\ &\quad (t = 2, \dots, T) \end{aligned}$$

$$\Phi_{T+1}(x_1, \dots, x_T, \xi^{T+1}) := 0.$$

Für diese Darstellung des Optimierungsproblems wurden quantitative Stabilitätsresultate bzgl. der Fortet-Mourier-Metriken in [70, 78] und bzgl. der L_r -minimalen Metriken zuvor in [79] im Fall $T = 2$ bewiesen.

Im Fall $T > 2$ hängt der Integrand f von bedingten Erwartungen bzgl. $\mathcal{F}_t(\xi)$ und damit in nichtlinearer Weise vom zugrundeliegenden Wahrscheinlichkeitsmaß P ab. Das hat zur Folge, dass bestehende Methoden zur Untersuchung des stochastischen Optimierungsproblems der Form (4.1) hinsichtlich Stabilität, wie sie in [70, 78] zu finden sind, nicht einfach auf mehrstufige Modelle für $T > 2$ übertragbar sind.

Ein Ausweg besteht in der Betrachtung mehrstufiger stochastischer Optimierungsprobleme als allgemeine Optimierungsprobleme in geeigneten unendlich-dimensionalen Funktionenräumen (siehe dazu auch [63] und [76]). Passend sind hier Banach-Räume vom Typ

$$L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \quad \text{mit} \quad 1 \leq r' \leq \infty, \quad m = \sum_{t=1}^T m_t,$$

versehen mit den Normen

$$\begin{aligned} \|x\|_{r'} &:= \left(\sum_{t=1}^T \mathbb{E}[\|x_t\|^{r'}] \right)^{\frac{1}{r'}} \quad \text{für } 1 \leq r' < \infty \quad \text{und} \\ \|x\|_{\infty} &:= \max_{t=1, \dots, T} \text{ess sup } \|x_t\| \quad \text{für } r' = \infty. \end{aligned}$$

Die Wahl von r' hängt sowohl von der Struktur des mehrstufigen stochastischen Problems als auch von den Eigenschaften des stochastischen Eingangsprozesses ξ ab. Ausschlaggebend ist dabei, welche der Parameter des Problems zufällig sind, d.h. wirklich von ξ abhängen.

4.1.1 Generelle Voraussetzungen

Um das Optimierungsmodell wohldefinieren zu können, wird angenommen, dass der Eingangsprozess ξ für ein $r \geq 1$ zu $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit $s := Td$ gehört. Passende Normen sind dann durch

$$\begin{aligned} \|\xi\|_r &:= \left(\sum_{t=1}^T \mathbb{E}[\|\xi_t\|^r] \right)^{\frac{1}{r}} \quad \text{für } 1 \leq r < \infty \quad \text{und} \\ \|\xi\|_{\infty} &:= \max_{t=1, \dots, T} \text{ess sup } \|\xi_t\| \quad \text{für } r = \infty. \end{aligned}$$

definiert. Zwischen den Parametern r und r' besteht je nach Problemstellung eine kanonische Abhängigkeit. Eine geeignete Wahl wäre

$$r' := \begin{cases} \frac{r}{r-1} & , \text{ die Kosten sind zufällig,} \\ r & , \text{ die rechten Seiten sind zufällig,} \\ r = 2 & , \text{ die Kosten und die rechten Seiten sind zufällig,} \\ \infty & , \text{ alle Technologie-Matrizen sind zufällig und } r = T. \end{cases} \quad (4.2)$$

Der Wert r gibt an, bis zu welcher Ordnung absolute Momente für den Eingangsprozess ξ existieren. Die Bestimmung der Werte r' nach (4.2) sichert dann, dass die Zielfunktion des Optimierungsproblems wohldefiniert und endlich ist. Die Forderung nach der Existenz von gewissen Momenten des Eingangsprozesses ist auch vom zweistufigen Fall her bekannt (vgl. Kapitel 2 in [83]). Die Mindestforderung ist $r \geq 1$, falls entweder die rechten Seiten oder die Kosten stochastisch sind. Die Flexibilität im Fall dass ξ Momente der Ordnung $r > 1$ besitzt, kann in der Abschwächung der Bedingung (A3) (siehe nächste Seite) genutzt werden, indem r' in Abhängigkeit von r so klein wie möglich gewählt wird. Ist das gegebene stochastische Problem komplett zufällig (d.h. Kosten sowie rechte Seiten und alle Technologie-Matrizen sind stochastisch), so benötigt man $r \geq T$ und es bleibt keine Flexibilität in der Wahl von r' .

Mit F wird im Folgenden die Zielfunktion des Optimierungsproblems bezeichnet. Das heißt, es sei F eine Abbildung

$$F : L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \rightarrow \overline{\mathbb{R}},$$

die durch

$$F(\xi, x) := \begin{cases} \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle \right] & , x \in \mathcal{X}(\xi), \\ +\infty & , \text{sonst,} \end{cases} \quad (4.3)$$

definiert wird, wobei

$$\mathcal{X}(\xi) := \{x = (x_1, \dots, x_T) \in \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t}) \mid x_1 \in \mathcal{X}_1(\xi_1), x_t \in \mathcal{X}_t(x_{t-1}; \xi_t), t = 2, \dots, T\}$$

die Menge aller zulässigen Elemente von (2.3) bezeichnet und ferner

$$\begin{aligned} \mathcal{X}_1(\xi_1) &:= \{x_1 \in X_1 \mid A_{1,0}x_1 = h_1(\xi_1)\}, \\ \mathcal{X}_t(x_{t-1}; \xi_t) &:= \{x_t \in \mathbb{R}^{m_t} \mid x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\} \end{aligned}$$

die jeweiligen Zulässigkeitsbereiche der ersten bzw. t -ten Stufe ($t = 2, \dots, T$) bezeichnen. Damit lässt sich das stochastische Optimierungsproblem (2.3) kurz in folgender Form schreiben:

$$\min\{F(\xi, x) \mid x \in \mathcal{X}(\xi)\}. \quad (4.4)$$

Mit $v(\xi)$ sei der Optimalwert des Problems bezeichnet und für ein beliebiges nichtnegatives α sei

$$l_\alpha(F(\xi, \cdot)) := \{x \in \mathcal{X}(\xi) \mid F(\xi, x) \leq v(\xi) + \alpha\}$$

die α -Levelmenge des stochastischen Optimierungsproblems (4.4). Für $\alpha = 0$ stimmt die Levelmenge mit der Lösungsmenge von (4.4) überein, die mit

$$S(\xi) := l_0(F(\xi, \cdot)) \quad (4.5)$$

bezeichnet wird.

Um Stabilitätsaussagen formulieren zu können, müssen einige weitere Forderungen an das Problem gestellt werden. Folgende Bedingungen seien für das stochastische Optimierungsproblem (4.4) erfüllt:

- (A1) $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$, d.h., $\int_{\Omega} \|\xi(\omega)\|^r d\mathbb{P}(\omega) < \infty$.
- (A2) Es existiert ein $\delta > 0$ derart, dass für alle Prozesse $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit $\|\tilde{\xi} - \xi\|_r \leq \delta$, für alle $t = 2, \dots, T$ und $x_1 \in \mathcal{X}_1(\tilde{\xi}_1)$, $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \dots, t-1$, zulässige, das heißt $\mathcal{F}_t(\tilde{\xi})$ -messbare Entscheidungen $x_t \in \mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ existieren (*Relatively-Complete-Recourse* lokal in ξ).
- (A3) Die Optimalwerte $v(\tilde{\xi})$ von (4.4) mit Eingangsprozessen $\tilde{\xi}$ sind endlich für alle $\tilde{\xi}$ in einer Umgebung des Originalprozesses ξ und die Zielfunktion F ist *lokal gleichmäßig levelbeschränkt* in ξ , d.h. für ein $\alpha > 0$ existiert ein $\delta > 0$ und eine beschränkte Teilmenge B von $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ mit der Eigenschaft, dass $l_\alpha(F(\tilde{\xi}, \cdot))$ für alle $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit $\|\tilde{\xi} - \xi\|_r \leq \delta$ in B enthalten ist.

4.1.2 Der Filtrationsabstand

Bei mehrstufigen stochastischen Optimierungsproblemen ist nicht allein die Verteilungsinformation des Eingangsprozesses wesentlich sondern auch die induzierte Informationsstruktur. Die Informationsstruktur gibt Auskunft über den bedingten Informationsfluss innerhalb des gesamten Optimierungszeitraums und beeinflusst die Lösung des Problems in entscheidender Weise.

Um die Struktur von Informationen in stochastischen Optimierungsmodellen zu quantifizieren, lässt sich ein *Filtrationsabstand* hinsichtlich zweier verschiedener Eingangsprozesse und im direkten Bezug zum Optimierungsproblem definieren. Es seien ξ und $\tilde{\xi}$ gegeben ($\tilde{\xi}$ kann zum Beispiel eine Näherung von ξ sein). Dann definiert

$$D_f(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0, \alpha]} D_{f, \varepsilon}(\xi, \tilde{\xi}), \quad (4.6)$$

einen Filtrationsabstand zwischen ξ und $\tilde{\xi}$, wobei $D_{f, \varepsilon}(\xi, \tilde{\xi})$ einen ε -Filtrationsabstand der Form

$$D_{f, \varepsilon}(\xi, \tilde{\xi}) := \inf_{\substack{x \in l_\varepsilon(F(\xi, \cdot)) \\ \tilde{x} \in l_\varepsilon(F(\tilde{\xi}, \cdot))}} \sum_{t=2}^{T-1} \max \left\{ \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'} \right\} \quad (4.7)$$

bezeichnet. Das Infimum in (4.7) wird bezüglich aller Elemente der jeweiligen ε -Levelmengen gebildet, d.h. bezüglich zulässiger und nahe optimaler Entscheidungen des Original- bzw. Näherungsproblems. Die Bezeichnung Filtrationsabstand für das Funktional D_f wird hier verwendet, obwohl die Dreiecksungleichung im Allgemeinen nicht gilt.

Ein vergleichbarer Abstand zwischen σ -Algebren wurden bereits in [3] verwendet. Dieser erzeugt eine spezielle Topologie auf der Menge von σ -Algebren. Unter Verwendung der Resultate aus den Arbeiten [82] und [61] kann er in der Form

$$d_B(\mathcal{B}, \mathcal{B}') := \sup_{f \in \Phi} \|\mathbb{E}[f|\mathcal{B}] - \mathbb{E}[f|\mathcal{B}']\|_1 \quad (4.8)$$

dargestellt werden und ist für alle Unteralgebren \mathcal{B} und \mathcal{B}' einer σ -Algebra \mathcal{F} definiert. Die Menge Φ umfasst alle \mathcal{F} -messbaren Funktionen f mit der Eigenschaft $\|f(\omega)\| \leq 1$ für alle $\omega \in \Omega$.

Nach [49] konvergiert eine Filtration genau dann, wenn jede σ -Algebra zu jeder Zeitstufe in Bezug auf den Abstand d_B konvergiert. Damit kann die Konvergenz von Filtrationen auf der Basis der Summe von Abständen von σ -Algebren charakterisiert werden. Der Filtrationsabstand (4.6) kann in diesem Sinne verwendet werden, da er sich durch die Summe von Abständen der Form (4.8) beschränken lässt. Für ξ und $\tilde{\xi}$ aus $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit $\|\tilde{\xi} - \xi\|_r \leq \delta$ erhält man unmittelbar die Abschätzung

$$\begin{aligned} D_f(\xi, \tilde{\xi}) &\leq \sup_{x \in B} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\tilde{\xi})]\|_{r'} \\ &\leq C \sup_{\|x\|_{r'} \leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\tilde{\xi})]\|_{r'}, \end{aligned} \quad (4.9)$$

wobei ein $\delta > 0$ und eine beschränkte Menge B in $L_{r'}$ entsprechend der Bedingungen (A2) bzw. (A3) gewählt seien. Die Konstante $C > 0$ sei so gewählt, dass $\|x\|_{r'} \leq C$ für alle $x \in B$.

Andere Abstände zwischen Filtrationen und σ -Algebren, wurden direkt im Zusammenhang mit stochastischen Optimierungsproblemen entworfen (vgl. z.B. [12]). Eine umfangreiche Studie zu neuen Resultaten und Anwendungen von Informationsabständen findet man in [1].

4.1.3 Stabilität von Optimalwerten

Das erste zentrale Resultat bezieht sich auf die Stabilität von Optimalwerten im stochastischen Optimierungsproblem (2.3) und ist ebenfalls in der aktuellen Arbeit [41] zu finden.

Theorem 4.1 *Es seien die Bedingungen (A1), (A2) und (A3) erfüllt und $\mathcal{X}_1(\xi_1)$ sei darüber hinaus nichtleer und beschränkt. Dann existieren positive Konstanten L und δ , so dass die Abschätzung*

$$|v(\xi) - v(\tilde{\xi})| \leq L \left(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi}) \right) \quad (4.10)$$

für alle $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit der Eigenschaft $\|\tilde{\xi} - \xi\|_r \leq \delta$ gilt.

Beweis: Sei $M_t : \mathbb{R}^{n_t} \Rightarrow \mathbb{R}^{m_t}$ die mengenwertige Abbildung

$$u \mapsto \{x \in \mathbb{R}^{m_t} \mid A_{t,0}x = u, x \in X_t\} \quad \text{für } t = 2, \dots, T.$$

Die Abbildungen M_t besitzen einen polyedrischen Graphen und sind deshalb auf ihren Definitionsbereichen $\text{def } M_t \subseteq \mathbb{R}^{m_t}$ Lipschitz-stetig in Bezug auf den Hausdorff-Abstand [77, Beispiel 9.35]. Folglich existieren positive Konstanten l_t , so dass

$$\sup_{x \in M_t(\bar{u})} d(x, M_t(\tilde{u})) \leq l_t \|\bar{u} - \tilde{u}\| \quad (4.11)$$

für alle $\bar{u}, \tilde{u} \in \text{def } M_t$ gilt, wobei $d(x, C)$ den Abstand von x zur nichtleeren Menge C in \mathbb{R}^{m_t} bezeichnet.

Sei $\alpha > 0$ und $\delta > 0$ entsprechend der Bedingungen (A2) und (A3) fixiert. Weiterhin sei $\varepsilon \in (0, \alpha]$ beliebig und ein $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ so gewählt, dass $\|\tilde{\xi} - \xi\|_r < \delta$ und $v(\tilde{\xi}) \in \mathbb{R}$. Für ein beliebiges Element $\bar{x} \in l_\varepsilon(F(\xi, \cdot))$ werden nun rekursiv Elemente \tilde{x}_t aus $L_{r'}(\Omega, \mathcal{F}_t(\tilde{\xi}), \mathbb{P}; \mathbb{R}^{m_t})$ ($t = 1, \dots, T$) konstruiert, die einerseits die Nebenbedingungen für $\tilde{\xi}$ erfüllen und für die andererseits Abschätzungen für

$$\|\mathbb{E}[\tilde{x}_t \mid \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\|, \quad t = 2, \dots, T,$$

unter Verwendung von (4.11) möglich sind. Dazu seien $\tilde{x}_1 := \bar{x}_1$ und $\hat{L}_1 := 1$ gesetzt und für $1 < t \leq T$ rekursiv Vektoren $\bar{u}_t := h_t(\xi_t) - A_{t,1}(\xi_t)\bar{x}_{t-1}$ und $\tilde{u}_t := h_t(\tilde{\xi}_t) - A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1}$ bestimmt. Damit sind die mengenwertige Abbildungen, definiert durch

$$\omega \mapsto M_t(\tilde{u}_t(\omega)) \quad \text{und} \quad \omega \mapsto \arg \min_{x \in M_t(\tilde{u}_t(\omega))} \|\mathbb{E}[\tilde{x}_t \mid \mathcal{F}_t(\tilde{\xi})](\omega) - x\|,$$

wegen der Messbarkeit von \tilde{x}_{t-1} bzgl. $\mathcal{F}_{t-1}(\tilde{\xi})$ sowie bekannter Messbarkeitsresultate für mengenwertige Abbildungen (vgl. zum Beispiel [77, Theorem 14.36]), messbar bzgl. $\mathcal{F}_t(\tilde{\xi})$. Darüber hinaus ist das Bild der Abbildung $\omega \mapsto M_t(\tilde{u}_t(\omega))$ aufgrund von (A2) stets nichtleer, wodurch [77, Theorem 14.37] anwendbar ist, was die Existenz einer $\mathcal{F}_t(\tilde{\xi})$ -messbaren Selektion \tilde{x}_t innerhalb der zweiten Abbildung sichert. Weil $\mathbb{E}[\tilde{x}_t \mid \mathcal{F}_t(\tilde{\xi})]$ zu $M_t(\mathbb{E}[\tilde{u}_t \mid \mathcal{F}_t(\tilde{\xi})])$ gehört, gilt für ein solches \tilde{x}_t wegen (4.11) die Abschätzung

$$\begin{aligned} \|\mathbb{E}[\tilde{x}_t \mid \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq l_t \|\mathbb{E}[\tilde{u}_t \mid \mathcal{F}_t(\tilde{\xi})] - \tilde{u}_t\| \\ &\leq l_t \left(\|\mathbb{E}[h_t(\xi_t) \mid \mathcal{F}_t(\tilde{\xi})] - h_t(\tilde{\xi}_t)\| \right. \\ &\quad \left. + \|\mathbb{E}[A_{t,1}(\xi_t)\bar{x}_{t-1} \mid \mathcal{F}_t(\tilde{\xi})] - A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1}\| \right) \\ &\leq l_t \left(\|\mathbb{E}[h_t(\xi_t) \mid \mathcal{F}_t(\tilde{\xi})] - h_t(\tilde{\xi}_t)\| \right. \\ &\quad \left. + \|\mathbb{E}[A_{t,1}(\xi_t)\bar{x}_{t-1} - A_{t,1}(\tilde{\xi}_t)\bar{x}_{t-1} \mid \mathcal{F}_t(\tilde{\xi})]\| \right. \\ &\quad \left. + \|\mathbb{E}[A_{t,1}(\tilde{\xi}_t)\bar{x}_{t-1} \mid \mathcal{F}_t(\tilde{\xi})] - A_{t,1}(\tilde{\xi}_t)\mathbb{E}[\bar{x}_{t-1} \mid \mathcal{F}_{t-1}(\tilde{\xi})]\| \right. \\ &\quad \left. + \|A_{t,1}(\tilde{\xi}_t)\mathbb{E}[\bar{x}_{t-1} \mid \mathcal{F}_{t-1}(\tilde{\xi})] - A_{t,1}(\tilde{\xi}_t)\tilde{x}_{t-1}\| \right). \quad (4.12) \end{aligned}$$

Wenn das gegebene Problem nicht komplett stochastisch ist, treten einige der Terme in der Abschätzung nicht auf. Hängen die Technologiematrizen nicht vom Zufall ab, entfällt die zweite Zeile in der letzten Ungleichung und die letzten zwei Zeilen vereinfachen sich. Sind die rechten Seiten des Problem nicht vom Zufall abhängig, so verschwindet die erste Zeile der letzten Ungleichung.

Je nach Fall wird die Ungleichungskette wie folgt fortgesetzt. Im Fall, dass sowohl rechte Seiten als auch die Technologiematrizen stochastisch sind, folgt mit der Jensen-Ungleichung sowie der Eigenschaft, dass $h_t(\cdot)$ und $A_{t,1}(\cdot)$ jeweils affin linear vom Zufall abhängen

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq l_t K_t \left(\|\mathbb{E}[\xi_t - \tilde{\xi}_t | \mathcal{F}_t(\tilde{\xi})]\| + \mathbb{E}[\|\xi_t - \tilde{\xi}_t\| \|\bar{x}_{t-1}\| | \mathcal{F}_t(\tilde{\xi})] \right. \\ &\quad + \max\{1, \|\tilde{\xi}_t\|\} \|\mathbb{E}[\bar{x}_{t-1} - \mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] | \mathcal{F}_t(\tilde{\xi})]\| \\ &\quad \left. + \max\{1, \|\tilde{\xi}_t\|\} \|\mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] - \tilde{x}_{t-1}\| \right), \end{aligned}$$

wobei K_t eine positive Konstante ist. Erneute Anwendung der Jensen-Ungleichung und die Eigenschaft $\|\xi_\tau\| \leq C \|\xi^t\|$, $\tau = 2, \dots, t$, ergibt rekursiv

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq \\ &\hat{L}_t \left(\sum_{\tau=2}^t \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\} \mathbb{E}[(1 + \|\bar{x}_{\tau-1}\|) \|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{\tau=2}^{t-1} \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right) \quad (4.13) \end{aligned}$$

mit positiven Konstanten \hat{L}_t für $t = 2, \dots, T$. Falls die Technologiematrizen nicht stochastisch sind, aber die rechten Seiten, ergibt sich aus (4.12)

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq l_t K_t \left(\|\mathbb{E}[\xi_t - \tilde{\xi}_t | \mathcal{F}_t(\tilde{\xi})]\| \right. \\ &\quad + \|\mathbb{E}[\bar{x}_{t-1} - \mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] | \mathcal{F}_t(\tilde{\xi})]\| \\ &\quad \left. + \|\mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] - \tilde{x}_{t-1}\| \right), \end{aligned}$$

so dass für nicht-stochastische Technologiematrizen rekursiv die Abschätzung

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq \hat{L}_t \left(\sum_{\tau=2}^t \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right) \quad (4.14) \end{aligned}$$

folgt. Noch weiter vereinfacht sich die Abschätzung, falls nur die Kosten im Modell stochastisch sind. Die Ungleichung (4.12) liefert in dem Fall

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq l_t K_t \left(\|\mathbb{E}[\bar{x}_{t-1} - \mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] | \mathcal{F}_t(\tilde{\xi})]\| \right. \\ &\quad \left. + \|\mathbb{E}[\bar{x}_{t-1} | \mathcal{F}_{t-1}(\tilde{\xi})] - \tilde{x}_{t-1}\| \right). \end{aligned}$$

Durch rekursives Anwenden der letzten Beziehung ergibt sich damit für den Fall, dass weder Technologiematrizen noch die rechten Seiten stochastisch sind

$$\|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| \leq \hat{L}_t \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})]. \quad (4.15)$$

Mit den gewonnenen Abschätzungen kann durch Einsetzen von \bar{x} und \tilde{x} in die Zielfunktion, fallweise eine obere Schranke für die Differenz der Optimalwerte

$$v(\tilde{\xi}) - v(\xi) \leq F(\tilde{\xi}, \tilde{x}) - F(\xi, \bar{x}) + \varepsilon.$$

im Folgenden bestimmt werden.

Beginnend mit dem einfachsten Fall, in dem nur die rechten Seiten zufällig sind, ergibt sich unter Verwendung von (4.3) und (4.14)

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq \sum_{t=2}^T \mathbb{E}[\langle b_t, \mathbb{E}[\tilde{x}_t - \bar{x}_t | \mathcal{F}_t(\tilde{\xi})] \rangle] + \varepsilon \\ &\leq \sum_{t=2}^T \|b_t\| \mathbb{E}[\|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\tilde{\xi})]\|] + \varepsilon \\ &\leq \hat{L} \sum_{t=2}^T \mathbb{E} \left[\sum_{\tau=2}^t \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right] + \varepsilon \\ &\leq \hat{L} T \mathbb{E} \left[\sum_{t=2}^T \|\xi_t - \tilde{\xi}_t\| + \sum_{\tau=2}^{T-1} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| \right] + \varepsilon \\ &\leq L \left(\|\xi - \tilde{\xi}\|_r + \sum_{\tau=2}^{T-1} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\|_r \right) + \varepsilon, \end{aligned}$$

wobei $\hat{L} := \max_{t=1, \dots, T} \hat{L}_t \|b_t\|$ und $r \geq 1$. Sind die Kosten zufällig, so erhält man

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq F(\tilde{\xi}, \tilde{x}) - F(\tilde{\xi}, \bar{x}) + F(\tilde{\xi}, \bar{x}) - F(\xi, \bar{x}) + \varepsilon \\ &\leq \mathbb{E} \left[\sum_{t=2}^T \langle b_t(\tilde{\xi}_t), \mathbb{E}[\tilde{x}_t - \bar{x}_t | \mathcal{F}_t(\tilde{\xi})] \rangle \right] \\ &\quad + \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\tilde{\xi}_t) - b_t(\xi_t), \bar{x}_t \rangle \right] + \varepsilon \\ &\leq \hat{K} \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}_t\|\} \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\tilde{\xi})]\| \right. \\ &\quad \left. + \sum_{t=1}^T \|\tilde{\xi}_t - \xi_t\| \|\bar{x}_t\| \right] + \varepsilon \end{aligned}$$

mit einer positiven Konstanten \hat{K} . Durch Anwendung der Hölder'schen Ungleichung im zweiten Summanden des letzten Ausdrucks ergibt sich

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq \hat{K} \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}_t\|\} \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\tilde{\xi})]\| \right] \\ &\quad + \hat{K} \|\tilde{\xi} - \xi\|_r \|\bar{x}\|_{r'} + \varepsilon, \end{aligned} \quad (4.16)$$

wobei $r' = \frac{r}{r-1}$ gilt. Unter Ausnutzung der Beschränktheit von $\|\bar{x}\|_{r'}$ gemäß (A3) und unter Verwendung von (4.14) kann die Ungleichungskette im Fall stochastischer Kosten und stochastischer rechter Seiten (d.h. $r = r' = 2$) wie folgt fortgesetzt werden:

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq \hat{L} \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}_t\|\} \left(\sum_{\tau=2}^t \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \right. \\ &\quad \left. \left. + \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right) \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_2 + \varepsilon \\ &\leq \hat{L} T \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}\|\} \mathbb{E}[\|\xi_t - \tilde{\xi}_t\| | \mathcal{F}_t(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{t=2}^T \max\{1, \|\tilde{\xi}\|\} \mathbb{E}[\|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\| | \mathcal{F}_{t+1}(\tilde{\xi})] \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_2 + \varepsilon \\ &\leq L \left(\|\tilde{\xi} - \xi\|_2 + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_2 \right) + \varepsilon, \end{aligned}$$

wobei die Cauchy-Schwarz'sche Ungleichung und die Tatsache, dass $\tilde{\xi}$ in einer beschränkten Menge in L_2 variiert, in der Abschätzung verwendet wurden. Die positive Konstante L hängt hier von ξ ab. Sind nur die Kosten stochastisch, so vereinfacht sich die letzte Ungleichungskette, indem statt (4.14) die Beziehung (4.15) verwendet wird. In dem Fall sei $r' = \frac{r}{r-1}$ und es ergibt sich

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq \hat{L} \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}_t\|\} \sum_{\tau=2}^{t-1} \mathbb{E}[\|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_r + \varepsilon \\ &\leq \hat{L} T \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}\|\} \mathbb{E}[\|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\| | \mathcal{F}_{t+1}(\tilde{\xi})] \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_r + \varepsilon \end{aligned}$$

$$\leq L \left(\|\tilde{\xi} - \xi\|_r + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'} \right) + \varepsilon,$$

wobei hier die Hölder'sche Ungleichung anstelle der Cauchy-Schwarz'schen Ungleichung sowie die Beschränktheit von $\tilde{\xi}$ in L_r verwendet wurden.

Als letztes bleibt der Fall, in dem sowohl die Technologiematrizen als auch die Kosten und die rechten Seiten stochastisch sind. Mit $r = T$ und $r' = \infty$ nimmt die allgemeine Abschätzung (4.13) damit die Form

$$\begin{aligned} \|\mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t\| &\leq \hat{L}_t \left(\sum_{\tau=2}^t \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\} \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{\tau=2}^{t-1} \max\{1, \|\tilde{\xi}^t\|^{t-\tau}\} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\|_\infty \right) \end{aligned}$$

an. Einsetzen in die Ungleichung (4.16) ergibt (mit $r = T$, $r' = \infty$)

$$\begin{aligned} v(\tilde{\xi}) - v(\xi) &\leq \hat{L} \mathbb{E} \left[\sum_{t=2}^T \left(\sum_{\tau=2}^t \max\{1, \|\tilde{\xi}^t\|^{t+1-\tau}\} \mathbb{E}[\|\xi_\tau - \tilde{\xi}_\tau\| | \mathcal{F}_\tau(\tilde{\xi})] \right. \right. \\ &\quad \left. \left. + \sum_{\tau=2}^{t-1} \max\{1, \|\tilde{\xi}^t\|^{t+1-\tau}\} \|\bar{x}_\tau - \mathbb{E}[\bar{x}_\tau | \mathcal{F}_\tau(\tilde{\xi})]\|_\infty \right) \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_T + \varepsilon \\ &\leq \tilde{L} \mathbb{E} \left[\sum_{t=2}^T \max\{1, \|\tilde{\xi}^t\|^{t-1}\} \mathbb{E}[\|\xi_t - \tilde{\xi}_t\| | \mathcal{F}_t(\tilde{\xi})] \right. \\ &\quad \left. + \sum_{t=2}^{T-1} \max\{1, \|\tilde{\xi}^t\|^{t-1}\} \|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_\infty \right] \\ &\quad + \hat{K} \bar{K} \|\tilde{\xi} - \xi\|_T + \varepsilon \\ &\leq \bar{L} \mathbb{E}[\max\{1, \|\tilde{\xi}\|^T\}] \left(\|\xi - \tilde{\xi}\|_T + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_\infty \right) + \varepsilon, \end{aligned}$$

wobei \bar{L} eine positive Konstante ist und erneut die Hölder-Ungleichung verwendet wurde. Weil $\tilde{\xi}$ zu einer beschränkten Menge in L_T gehört, existiert auch in diesem Fall eine von ξ abhängende Konstante $L > 0$, so dass

$$v(\tilde{\xi}) - v(\xi) \leq L \left(\|\xi - \tilde{\xi}\|_r + \sum_{t=2}^{T-1} \|\bar{x}_t - \mathbb{E}[\bar{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'} \right) + \varepsilon \quad (4.17)$$

mit $r = T$ und $r' = \infty$ gilt.

In allen betrachteten Fällen erhält man für die Differenz der Optimalwerte eine Abschätzung der Form (4.17). Das Vertauschen der Rollen von ξ und $\tilde{\xi}$ führt analog zur Abschätzung

$$v(\xi) - v(\tilde{\xi}) \leq L \left(\|\xi - \tilde{\xi}\|_r + \sum_{t=2}^{T-1} \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'} \right) + \varepsilon. \quad (4.18)$$

Die Summen innerhalb der Ungleichungen (4.17) und (4.18) lassen sich durch

$$\sum_{t=2}^{T-1} \max \left\{ \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'} \right\}$$

beschränken. Da die Abschätzungen (4.17) und (4.18) darüber hinaus für alle $\tilde{x} \in l_\varepsilon(F(\xi, \cdot))$ bzw. für alle $\tilde{x} \in l_\varepsilon(F(\tilde{\xi}, \cdot))$ gelten, erhält man so

$$\begin{aligned} |v(\xi) - v(\tilde{\xi})| &\leq L \left(\|\xi - \tilde{\xi}\|_r + D_{f,\varepsilon}(\xi, \tilde{\xi}) \right) + \varepsilon \\ &\leq L \left(\|\xi - \tilde{\xi}\|_r + \sup_{\delta \in (0, \alpha]} D_{f,\delta}(\xi, \tilde{\xi}) \right) + \varepsilon. \end{aligned}$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, folgt aus der letzten Ungleichung die Behauptung. \square

Eine hinreichende Eigenschaft dafür, dass die Bedingung (A2) erfüllt wird, ist die sogenannte *Complete-Fixed-Recourse*-Bedingung an alle Matrizen $A_{t,0}$, d.h. die Mengen X_t sind polyedrische Kegel und $A_{t,0}X_t = \mathbb{R}^{n_t}$ gilt für alle $t = 2, \dots, T$. Die Forderung (A3) der lokalen gleichmäßigen Levelbeschränktheit der Zielfunktion F ist standard in der Störungstheorie von Optimierungsproblemen (siehe zum Beispiel [77, Theorem 1.17]). Die Endlichkeitsforderung der Optimalwerte ist nötig, da sie nicht allein aus der Levelbeschränktheit von F für alle relevanten Paare (r, r') folgt.

Im Spezialfall, in dem Ω endlich ist, oder im Fall $1 < r' < \infty$ ist die Existenz von Lösungen des Optimierungsproblems (4.4) und damit auch die Endlichkeit von $v(\xi)$ aufgrund der Kompaktheit bzw. schwachen Folgenkompaktheit von $l_\alpha(F(\xi, \cdot))$ in reflexiven Banachräumen und aufgrund der Linearität der Zielfunktion gesichert (siehe Proposition 4.3 in Abschnitt 4.1.5). Der Filtrationsabstand nimmt dann die Form

$$D_f(\xi, \tilde{\xi}) = \inf \left\{ \sum_{t=2}^{T-1} D_t(\xi, \tilde{\xi}) : x \in S(\xi), \tilde{x} \in S(\tilde{\xi}) \right\} \quad (4.19)$$

an, wobei $D_t(\xi, \tilde{\xi})$ die Gestalt

$$D_t(\xi, \tilde{\xi}) := \max \left\{ \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'} \right\} \quad (4.20)$$

für $t = 2, \dots, T - 1$ besitzt.

Eine weitere Bemerkung betrifft die Wahl von $r = r' = 2$ im Fall stochastischer Kosten und stochastischer rechter Seiten. Mehr Flexibilität bei der Wahl von r und r' kann in diesem Fall erzielt werden, wenn die Kosten $b_t(\xi_t)$ endliche Momente der Ordnung $\hat{r} \geq 1$ besitzen. In dem Fall lässt sich der Beweis leicht für den Fall, dass $r := \hat{r}$ und $r' := \frac{\hat{r}}{\hat{r}-1}$ gewählt wird, modifizieren. Zusätzlich muss dann aber auch die Zugehörigkeit von $h_t(\xi_t)$ zu $L_{r'}$ gesichert sein.

Offen ist auch die Wahl des Wahrscheinlichkeitsraums. Theorem 4.1 gilt für jeden Wahrscheinlichkeitsraum, auf dem sich sowohl ξ als auch $\tilde{\xi}$ realisieren lassen. Es wäre also sinnvoll, nach einem Wahrscheinlichkeitsraum zu suchen, auf dem der L_r -Abstand $\|\xi - \tilde{\xi}\|_r$ sowie die $L_{r'}$ -Abstände $\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}$ und $\|\tilde{x}_t - \mathbb{E}[\mathcal{F}_t(\xi)]\|_{r'}$ für $t = 2, \dots, T - 1$ klein werden. Eine solche Überlegung ist vernünftig, denn es gilt: Wenn mit P und Q die Wahrscheinlichkeitsverteilungen von ξ bzw. $\tilde{\xi}$ bezeichnet werden, so existiert eine optimale Lösung $\eta^* \in \mathcal{P}(\Xi \times \Xi)$ des Transportproblems (3.6) [69, Theorem 8.1.1], das heißt,

$$\ell_r^*(P, Q) = \int_{\Xi \times \Xi} \|\xi - \tilde{\xi}\|_r \eta^*(d\xi, d\tilde{\xi}),$$

wobei $\pi_1 \eta^* = P$ und $\pi_2 \eta^* = Q$ gilt. Des Weiteren existiert ein Wahrscheinlichkeitsraum $(\Omega', \mathcal{F}', \mathbb{P}')$ und ein sogenanntes *Optimal-Coupling*, d.h. ein Paar $(\xi'(\cdot), \tilde{\xi}'(\cdot))$ von Zufallsvektoren definiert auf $(\Omega', \mathcal{F}', \mathbb{P}')$ mit Werten in $\Xi \times \Xi$, so dass die gemeinsame Wahrscheinlichkeitsverteilung gerade η^* ist [69, Theorem 2.5.1]. Insbesondere gilt dann, dass der Abstand der Zufallsvektoren gemessen in $L_r(\Omega', \mathcal{F}', \mathbb{P}'; \mathbb{R}^s)$ gleich dem L_r -minimale Abstand der Wahrscheinlichkeitsverteilungen ist, d.h. es gilt

$$\ell_r(P, Q) = \|\xi'(\cdot) - \tilde{\xi}'(\cdot)\|_r.$$

Analog kann für die $L_{r'}$ -Abstände $\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}$ und $\|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'}$ vorgegangen werden.

4.1.4 Beispiel

Ein einfaches Beispiel belegt die Unverzichtbarkeit der Einbeziehung eines Informationsabstands bei der Stabilitätsuntersuchung mehrstufiger stochastischer Optimierungsprobleme hinsichtlich von Störungen des Eingangsprozesses. Insbesondere wird deutlich, dass allein die Verwendung des L_r -Abstands für das Stabilitätsresultat von Theorem 4.1 nicht genügt.

Betrachtet wird ein mehrstufiges stochastisches Optimierungsproblem zur Bestimmung einer optimalen Beschaffungsstrategie über einen gegebenen Zeitraum mit zufälligen Kosten. Die Stufenentscheidung x_t des Problems gibt dabei den Beschaffungsanteil an, der zur Deckung eines vorgegebenen Gesamtbedarfs $a > 0$, innerhalb der Zeitperiode t bezogen wird. Die stochastischen Preise seien

durch die Zufallszahlen ξ_t für $t = 1, \dots, T$ gegeben. Die Zielfunktion gibt die zu erwartenden Kosten wieder, die bei einer Beschaffungsstrategie $x = (x_1, \dots, x_T)$ entstehen. Damit lässt sich das Problem in der Form

$$\min \left\{ \mathbb{E} \left[\sum_{t=1}^T \xi_t x_t \right] \left| \begin{array}{l} (x_t, s_t) \in X_t = \mathbb{R}_+^2, \\ (x_t, s_t) \text{ ist } \mathcal{F}_t(\xi)\text{-messbar,} \\ s_t - s_{t-1} = x_t, t = 2, \dots, T, \\ s_1 = 0, s_T = a \end{array} \right. \right\}$$

formulieren. Die Stufenvariable s_t dient hier als Hilfsvariable und beschreibt den Gesamtstand der Beschaffung zum Zeitpunkt t .

Zur Illustration sei $T := 3$. Mit (P_ε) wird eine Familie von Wahrscheinlichkeitsverteilungen betrachtet, so dass P_ε in Abhängigkeit von $\varepsilon > 0$ die Verteilung des stochastischen Preisprozesses ξ_ε der Form

$$\begin{aligned} \xi_\varepsilon^1 &= (3, 2 + \varepsilon, 3) \quad \text{und} \quad p_1 := P_\varepsilon(\{\xi_\varepsilon^1\}) = 0.5, \\ \xi_\varepsilon^2 &= (3, 2, 1) \quad \text{und} \quad p_2 := P_\varepsilon(\{\xi_\varepsilon^2\}) = 0.5 \end{aligned}$$

beschreibt. Darüber hinaus sei Q die Verteilung des Grenzprozesses $\tilde{\xi}$ von (ξ_ε) für $\varepsilon \rightarrow 0$, dargestellt durch

$$\begin{aligned} \tilde{\xi}^1 &= (3, 2, 3) \quad \text{und} \quad q_1 := Q(\{\tilde{\xi}^1\}) = 0.5, \\ \tilde{\xi}^2 &= (3, 2, 1) \quad \text{und} \quad q_2 := Q(\{\tilde{\xi}^2\}) = 0.5. \end{aligned}$$

Für jedes $\varepsilon > 0$ definieren die stochastischen Prozesse ξ_ε und $\tilde{\xi}$ jeweils unterschiedliche Baumstrukturen (siehe Abbildung 4.1). Das ist gleichbedeutend damit, dass sich die Filtrationen, die durch die Prozesse ξ_ε bzw. $\tilde{\xi}$ erzeugt werden, unterscheiden. Das hat zur Folge, dass auch wenn ε beliebig klein gewählt

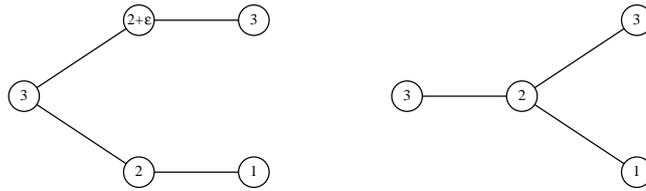


Abbildung 4.1: Szenariobäume für P_ε mit $\varepsilon > 0$ (links) und Q .

wird, die Optimalwerte der Probleme mit Eingangsprozessen ξ_ε nicht gegen den Optimalwert des Problems mit Eingangsprozess $\tilde{\xi}$ konvergiert. Es ergibt sich

$$v(\xi_\varepsilon) = \frac{3 + \varepsilon}{2} a \quad \text{und} \quad v(\tilde{\xi}) = 2a$$

für die Optimalwerte bei gleichzeitigen optimalen Entscheidungen

$$x^1 = (0, a, 0) \quad \text{und} \quad x^2 = (0, 0, a) \quad \text{bzw.} \quad \tilde{x}^1 = \tilde{x}^2 = (0, a, 0).$$

Auf der anderen Seite konvergiert die Verteilung P_ε gegen die Verteilung Q bezüglich ℓ_1 bzw. ξ_ε gegen $\tilde{\xi}$ in L_1 . Denn es gilt

$$\ell_1(P_\varepsilon, Q) = \|\xi_\varepsilon - \tilde{\xi}\|_1 = \frac{\varepsilon}{2} \rightarrow 0 \quad (\varepsilon \rightarrow 0).$$

Folglich ist das 3-stufige stochastische Optimierungsmodell zur Beschaffungsoptimierung nicht stabil in Bezug auf den L_1 -Abstand (bzw. ℓ_1). Die Stabilitätsabschätzung für die Differenz der Optimalwerte $|v(\xi_\varepsilon) - v(\tilde{\xi})|$ in Theorem 4.1 bleibt dennoch gültig. Der Grund dafür, dass die Optimalwerte im Gegensatz zu den Verteilungen nicht konvergieren, liegt im nicht verschwindenden Filtrationsabstand $D_f(\xi_\varepsilon, \tilde{\xi})$ der beiden betrachteten Prozesse ξ_ε und $\tilde{\xi}$ begründet. Es gilt für alle $\varepsilon > 0$ mit der Darstellung (4.19) und (4.20) sowie $r' = \infty$:

$$\begin{aligned} D_f(\xi_\varepsilon, \tilde{\xi}) &= \inf \left\{ D_2(\xi_\varepsilon, \tilde{\xi}) : x \in S(\xi_\varepsilon), \tilde{x} \in S(\tilde{\xi}) \right\} \\ &= \inf \left\{ \|x_2 - \mathbb{E}[x_2 | \mathcal{F}_2(\tilde{\xi})]\|_\infty : x \in S(\xi_\varepsilon) \right\} \\ &= \max \left\{ \left| a - \frac{a}{2} \right|, \left| 0 - \frac{a}{2} \right| \right\} = \frac{a}{2}, \end{aligned}$$

wobei ausgenutzt wurde, dass stets $\mathcal{F}_2(\tilde{\xi}) \subset \mathcal{F}_2(\xi_\varepsilon)$ gilt. Die Stabilitätsungleichung (4.10) ist damit zum Beispiel mit der Konstanten $L = 1$ erfüllt.

4.1.5 Stabilität von Lösungsmengen

Das erste Resultat betrifft die Stabilität von Erststufenentscheidungen mehrstufiger stochastischer Optimierungsprobleme. Mit derselben Technik wie im Beweis von [78, Theorem 9] kann die folgende Stabilitätsaussage für die Lösungsmengen $S_1(\xi)$ des Erststufenproblems (4.1) nachgewiesen werden.

Proposition 4.2 *Es seien die Voraussetzungen von Theorem 4.1 erfüllt. Dann gilt für die Erststufenlösungen $S_1(\xi)$ bzw. $S_1(\tilde{\xi})$*

$$\sup_{x \in S_1(\tilde{\xi})} d(x, S_1(\xi)) \leq \Psi_\xi^{-1} \left(\hat{L}(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi})) \right), \quad (4.21)$$

wobei mit

$$\Psi_\xi(\tau) := \inf \left\{ \mathbb{E}[f(x_1, \xi)] - v(\xi) : d(x_1, S_1(\xi)) \geq \tau, x_1 \in X_1 \right\}$$

und $\Psi_\xi^{-1}(\alpha) := \sup \{ \tau \in \mathbb{R}_+ : \Psi_\xi(\tau) \leq \alpha \}$ ($\alpha \in \mathbb{R}_+$) die Zuwachsfunktion des Originalproblems (4.1) in der Nähe der Lösungsmenge $S_1(\xi)$ bezeichnet wird.

Das folgende Resultat zur Charakterisierung von Lösungsmengen des mehrstufigen stochastischen Optimierungsproblems (4.4) lässt sich in voller Allgemeinheit nur unter weiteren Voraussetzungen herleiten. Folgende Bedingungen seien zusätzlich zu den Bedingungen (A1) bis (A3) formuliert:

- (A4) Es existiert ein zulässiges Element z aus $\times_{t=1}^T L_{\hat{r}}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_t})$ ($\frac{1}{r} + \frac{1}{\hat{r}} = 1$) des dualen stochastischen Optimierungsproblems zu (4.4), d.h. es gilt

$$\begin{aligned} A_{t,0}^* z_t + A_{t+1,1}^*(\xi_{t+1}) z_{t+1} - b_t(\xi_t) &\in X_t^*, \quad t = 1, \dots, T-1, \\ A_{T,0}^* z_T - b_T(\xi_T) &\in X_T^*, \end{aligned} \quad (4.22)$$

wobei mit X_t^* die Polarkegel der polyedrischen Kegel X_t ($t = 1, \dots, T$) bezeichnet sind. Die Notation $*$ bei Matrizen steht für das Transponieren der jeweiligen Matrix.

- (A5) Im Fall $r' = 1$ existiert für alle $c \geq 0$ eine Funktion $g \in L_1(\Omega, \mathcal{F}, \mathbb{P})$ mit der Eigenschaft

$$\sum_{t=1}^T \langle b_t(\xi_t(\omega)), x_t \rangle \geq c|x| - g(\omega)$$

für alle $x \in \mathbb{R}^m$ mit $x_t \in X_t$ für $t = 1, \dots, T$, $A_{1,0}x_1 = h_1(\xi_1)$ und $A_{t,0}x_t + A_{t,1}(\xi_t(\omega))x_{t-1} = h_t(\xi_t(\omega))$ für $t = 2, \dots, T$ sowie für \mathbb{P} -fast alle $\omega \in \Omega$.

Um den Satz von Weierstraß zur Existenz von Minima anwenden zu können, benötigt man eine Topologie \mathcal{T} auf $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, so dass die Levelmengen $l_\alpha(F(\xi, \cdot))$ kompakt bezüglich \mathcal{T} sind. Die Norm-Topologie ist für unendlichdimensionale Optimierungsprobleme in L_p -Räumen zu stark und daher nicht geeignet. Einen Ausweg liefern die schwachen Topologien $\sigma(L_p, L_q)$ in den Räumen $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, wobei $p \in [1, \infty]$ und $\frac{1}{p} + \frac{1}{q} = 1$. Diese Topologien besitzen die Hausdorff-Eigenschaft und werden durch eine Basis erzeugt, die aus den Mengen

$$\mathcal{O} = \left\{ x \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : \left| \mathbb{E} \left[\sum_{t=1}^T \langle x_t - x_t^0, y_t^i \rangle \right] \right| < \varepsilon, i = 1, \dots, n \right\},$$

für alle $x^0 \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, $n \in \mathbb{N}$, $\varepsilon > 0$ und $y^i \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ sowie $i = 1, \dots, n$, besteht.

Für $p \in [1, \infty)$ lässt sich die schwache Topologie $\sigma(L_p, L_q)$ in der Form $\sigma(E, E^*)$ mit einem Banach-Raum E und seinem topologisch dualen E^* darstellen. Ist $p = \infty$, so wird die schwache Topologie $\sigma(L_\infty, L_1)$ auf dem Banach-Raum $L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ auch Schwach*-Topologie genannt, da diese von der Form $\sigma(E^*, E)$ ist. Wenn Ω endlich ist, so stimmen die schwache Topologie und die Norm-Topologie überein. Ist der Raum $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ unendlichdimensional, so ist die zugehörige schwache Topologie $\sigma(L_p, L_q)$ nicht metrisierbar. Für $p \in [1, \infty)$ sind nach dem Eberlein-Šmulian-Theorem Teilmengen von $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ genau dann (relativ) schwach kompakt, wenn sie (relativ) schwach folgenkompakt sind. Für $p = \infty$ geht die letzte Eigenschaft im Allgemeinen verloren.

Eine weitere Eigenschaft, die hier erwähnt sei, betrifft die Metrisierbarkeit auf kompakten Teilmengen. Es gilt, dass wenn eine Teilmenge B von $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ kompakt bezüglich der schwachen Topologie $\sigma(L_p, L_q)$ ist, so ist deren Einschränkung auf B immer dann metrisierbar, wenn $L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ separabel ist. Diese und weitere Aussagen können in [24, Kapitel 3 und 4] nachvollzogen werden.

Die folgende Aussage bzgl. der Existenz von Lösungen in (4.4) und deren Charakterisierung ist der aktuellen Arbeit [39, Theorem 2.1] entnommen.

Proposition 4.3 *Es seien die Aussagen (A1) – (A5) für ein Paar (r, r') mit der Bedingung (4.2) erfüllt. Dann ist die Lösungsmenge $S(\xi)$ von (4.4) nicht-leer, konvex und kompakt bzgl. der schwachen Topologie $\sigma(L_{r'}, L_q)$ ($\frac{1}{r'} + \frac{1}{q} = 1$). (Die Bedingungen (A4) und (A5) werden hier nur für den Fall $r' \in \{1, \infty\}$ benötigt.)*

Beweis: Der Integrand $f : \Omega \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ sei definiert durch

$$f(\omega, x) := \begin{cases} \sum_{t=1}^T \langle b_t(\xi_t(\omega)), x_t \rangle & , x_1 \in \mathcal{X}_1(\xi_1), x_t \in \mathcal{X}_t(x_{t-1}, \xi_t(\omega)), t = 2, \dots, T, \\ +\infty & , \text{sonst.} \end{cases}$$

Dann ist f ein proper-normal-konvexer Integrand (vgl. [75] und [77, Kapitel 14]). Es sei $(\omega, x) \in \Omega \times \mathbb{R}^m$ so gewählt, dass $x_1 \in \mathcal{X}_1(\xi_1)$ und $x_t \in \mathcal{X}_t(x_{t-1}, \xi_t(\omega))$ für $t = 2, \dots, T$. Dann existiert wegen (A4) ein Element $z \in \times_{t=1}^T L_{\hat{r}}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_t})$, welches die Eigenschaft (4.22) erfüllt. Folglich existiert für jedes $t = 1, \dots, T$ ein $x_t^*(\omega) \in X_t^*$, so dass

$$\begin{aligned} b_t(\xi_t(\omega)) &= A_{t,0}^* z_t(\omega) + A_{t+1,1}^*(\xi_{t+1}(\omega)) z_{t+1}(\omega) - x_t^*(\omega) \quad (t = 1, \dots, T-1), \\ b_T(\xi_T(\omega)) &= A_{T,0}^* z_T(\omega) - x_T^*(\omega). \end{aligned}$$

Einsetzen der obigen Darstellung von $b_t(\xi_t(\omega))$ in den Integranden f ergibt mit $F(\xi, x) = \mathbb{E}[f(\omega, x)]$ die Abschätzung

$$\begin{aligned} f(\omega, x) &= \sum_{t=1}^{T-1} \langle A_{t,0}^* z_t(\omega) + A_{t+1,1}^*(\xi_{t+1}(\omega)) z_{t+1}(\omega) - x_t^*(\omega), x_t \rangle \\ &\quad + \langle A_{T,0}^* z_T(\omega) - x_T^*(\omega), x_T \rangle \\ &\geq \sum_{t=1}^{T-1} \langle A_{t,0}^* z_t(\omega) + A_{t+1,1}^*(\xi_{t+1}(\omega)) z_{t+1}(\omega), x_t \rangle + \langle A_{T,0}^* z_T(\omega), x_T \rangle \\ &= \sum_{t=1}^T \langle z_t(\omega), A_{t,0} x_t \rangle + \sum_{t=1}^{T-1} \langle z_{t+1}(\omega), A_{t+1,1}(\xi_{t+1}(\omega)) x_t \rangle \\ &= \sum_{t=1}^T \langle z_t(\omega), h_t(\xi_t(\omega)) \rangle. \end{aligned}$$

Folglich gilt

$$f(\omega, x) \geq g(\omega), \quad \text{wobei} \quad g := \sum_{t=1}^T \langle z_t, h_t(\xi_t) \rangle \in L_1(\Omega, \mathcal{F}, \mathbb{P}).$$

Für den konjugiert-normal-konvexen Integranden $f^* : \Omega \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$, der durch

$$f^*(\omega, y) := \sup_{x \in \mathbb{R}^m} \{ \langle y, x \rangle - f(\omega, x) \}$$

gegeben ist, wird damit impliziert, dass die Abschätzung $f^*(\omega, 0) \leq -g(\omega)$ gilt. Folglich ist die Voraussetzung von [75, Corollary 3D] erfüllt und es kann geschlossen werden, dass die Integralfunktion $F(\xi, \cdot) = \mathbb{E}[f(\omega, \cdot)]$ unterhalbstetig auf $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ bezüglich der schwachen Topologie $\sigma(L_{r'}, L_q)$ ist.

$\mathcal{N}_{r'}(\xi) := \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t})$ bezeichne den Unterraum nichtantizipativer Elemente zum Eingangsprozess ξ . Dann gilt: $\mathcal{N}_{r'}(\xi)$ ist für alle $r' \in [1, \infty]$ abgeschlossen bezüglich der schwachen Topologie $\sigma(L_{r'}, L_q)$. Für $r' \in [1, \infty)$ ist das eine direkte Folgerung aus der Norm-Abgeschlossenheit und der Konvexität von $\mathcal{N}_{r'}(\xi)$. Für $r' = \infty$ sei $(x_\alpha)_{\alpha \in I}$ ein Netz in $\mathcal{N}_\infty(\xi)$ mit einer partiell geordneten Menge (I, \leq) , welches gegen ein Element $x^* \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ konvergiert. Jede Umgebung $U(x^*)$ von x^* ist bezüglich der schwachen Topologie $\sigma(L_\infty, L_1)$ von der Form

$$U(x^*) = \left\{ x \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : \left| \mathbb{E} \left[\sum_{t=1}^T \langle x_t - x_t^*, y_t^i \rangle \right] \right| < \varepsilon_i, i = 1, \dots, n \right\},$$

wobei $n \in \mathbb{N}$, $y^i \in L_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$, $\varepsilon_i > 0$, $i = 1, \dots, n$. Weil das Netz $(x_\alpha)_{\alpha \in I}$ gegen x^* konvergiert, existiert ein $\alpha_0 \in I$, so dass $x_\alpha \in U(x^*)$ falls $\alpha_0 \leq \alpha$. Wenn die Elemente y^i für alle $i = 1, \dots, n$ zu $\times_{t=1}^T L_1(\Omega, \mathcal{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t})$ gehören, so erhält man

$$\begin{aligned} \left| \mathbb{E} \left[\sum_{t=1}^T \langle x_{\alpha,t} - x_t^*, y_t^i \rangle \right] \right| &= \left| \mathbb{E} \left[\sum_{t=1}^T \mathbb{E}[\langle x_{\alpha,t} - x_t^*, y_t^i \rangle | \mathcal{F}_t(\xi)] \right] \right| \\ &= \left| \mathbb{E} \left[\sum_{t=1}^T \langle x_{\alpha,t} - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)], y_t^i \rangle \right] \right| < \varepsilon_i \end{aligned}$$

wegen der Eigenschaft $\mathbb{E}[x_{\alpha,t} | \mathcal{F}_t(\xi)] = x_{\alpha,t}$ für alle $t = 1, \dots, T$ und $\alpha \in I$. Daraus folgt in diesem Fall

$$U(x^*) = U(\mathbb{E}[x_1^* | \mathcal{F}_1(\xi)], \dots, \mathbb{E}[x_T^* | \mathcal{F}_T(\xi)]).$$

Weil $(x_\alpha)_{\alpha \in I}$ gegen x^* konvergiert und weil die schwache Topologie die Hausdorff-Eigenschaft besitzt, folgt weiterhin, dass $x_t^* = \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)]$ für $t = 1, \dots, T$ und damit, dass $x^* \in \mathcal{N}_\infty(\xi)$.

Es verbleibt zu zeigen, dass für ein $\alpha > 0$ die zugehörige α -Levelmenge $l_\alpha(F(\xi, \cdot))$ kompakt bezüglich der schwachen Topologie $\sigma(L_{r'}, L_q)$ ist. Für $r' \in (1, \infty)$ ist der Banach-Raum $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ reflexiv. Darüber hinaus ist jede α -Levelmenge $l_\alpha(F(\xi, \cdot))$ abgeschlossen und konvex. Für ein $\alpha > 0$ ist die zugehörige Levelmenge wegen (A3) beschränkt und folglich kompakt in Bezug auf $\sigma(L_{r'}, L_q)$. Für $r' = 1$ folgt die Kompaktheit der α -Levelmengen bezüglich $\sigma(L_1, L_\infty)$ aus [75, Theorem 3K] wegen der Bedingung (A5). Für $r' = \infty$ existiert eine beschränkte α -Levelmenge wegen (A3) und damit gilt unter Anwendung des Alaoglu'schen Theorems [24, Theorem 3.21], dass sie relativ-kompakt bezüglich $\sigma(L_\infty, L_1)$ ist. Weil aber die Zielfunktion $F(\xi, \cdot)$ unterhalb-stetig ist und weil $\mathcal{N}_\infty(\xi)$ schwach-abgeschlossen in Bezug auf $\sigma(L_\infty, L_1)$ ist, ist die α -Levelmenge auch kompakt bezüglich $\sigma(L_\infty, L_1)$.

Damit ist nach dem Weierstraß'schen Theorem die Menge $S(\xi)$ nichtleer und kompakt bezüglich der schwachen Topologie $\sigma(L_{r'}, L_q)$. Die Konvexität von $S(\xi)$ ist eine unmittelbare Konsequenz aus der Konvexität der Zielfunktion $F(\xi, \cdot)$ des stochastischen Optimierungsproblems (4.4).

Schließlich sei bemerkt, dass die Aussagen (A4) und (A5) im Fall $r' \in (1, \infty)$ für den Beweis, dass $S(\xi)$ nichtleer und kompakt bezüglich der schwachen Topologie $\sigma(L_{r'}, L_q)$ ist, nicht benötigt werden. Diese Eigenschaft folgt dann bereits aufgrund dessen, dass eine lineare stetige Funktion auf einer abgeschlossenen, konvexen sowie beschränkten Teilmenge eines reflexiven Banach-Raums minimiert wird. \square

Um ein Stabilitätsresultat für Lösungsmengen in (4.4) von Proposition 4.3 ableiten zu können, bedarf es einer stärkeren Formulierung des Filtrationsabstands D_f (siehe Abschnitt 4.1.2). Unter der Voraussetzung, dass Lösungen des stochastischen Optimierungsproblems (4.4) für die Eingangsprozesse ξ und $\tilde{\xi}$ existieren, sei

$$\bar{D}_f(\xi, \tilde{\xi}) := \sup \left\{ \sum_{t=2}^T D_t(\xi, \tilde{\xi}) : x \in S(\xi), \tilde{x} \in S(\tilde{\xi}) \right\}$$

mit $D_t(\xi, \tilde{\xi})$ definiert wie in (4.20), hier aber für $t = 2, \dots, T$. Offensichtlich ist das Funktional $\bar{D}_f(\xi, \tilde{\xi})$ eine obere Schranke für das Funktional $D_f(\xi, \tilde{\xi})$. Damit gilt folgende Aussage (eine abgeschwächte Version von [39, Theorem 2.5]):

Theorem 4.4 *In (4.4) seien nur die Kosten und rechten Seiten stochastisch und es seien die Bedingungen (A1)–(A3) für $r \in (1, \infty)$ und $\frac{1}{r} + \frac{1}{r'} = 1$ erfüllt. Wenn $(\xi^{(n)})$ eine Folge in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ ist, die gegen ein ξ bezüglich L_r und \bar{D}_f konvergiert sowie wenn $(x^{(n)})$ eine Folge von Lösungen der zugehörigen Optimierungsprobleme ist (d.h. wenn $x^{(n)} \in S(\xi^{(n)})$), dann existiert eine konvergente Teilfolge $(x^{(n_k)})$ von $(x^{(n)})$, die bezüglich der schwachen Topologie $\sigma(L_{r'}, L_r)$ gegen ein Element aus $S(\xi)$ konvergiert. Ist $S(\xi)$ einelementig, so konvergiert die Folge $(x^{(n)})$ gegen die eindeutige Lösung von (4.4).*

Beweis: Es seien $(\xi^{(n)})$ und $(x^{(n)})$ den Voraussetzungen entsprechend gegeben. Dann existiert ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$, $\|\xi^{(n)} - \xi\|_r \leq \delta$ und $x^{(n)} \in l_\alpha(F(\xi^{(n)}, \cdot))$ gilt, wobei $\alpha > 0$ und $\delta > 0$ gemäß (A3) gewählt seien. Daraus folgt, dass die Folge $(x^{(n)})$ in einer beschränkten Teilmenge des reflexiven Banach-Raums $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ enthalten ist und damit eine konvergente Teilfolge bezüglich der schwachen Topologie $\sigma(L_{r'}, L_r)$ besitzt. Der Grenzwert x^* liegt in $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ und Theorem 4.1 liefert

$$v(\xi^{(n_k)}) = F(\xi^{(n_k)}, x^{(n_k)}) = \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\xi_t^{(n_k)}), x_t^{(n_k)} \rangle \right] \longrightarrow v(\xi) \quad (k \rightarrow \infty).$$

Wegen der Norm-Konvergenz von $(\xi^{(n_k)})$ und der schwachen Konvergenz von $(x^{(n_k)})$ ergibt sich

$$\mathbb{E} \left[\sum_{t=1}^T \langle b_t(\xi_t^{(n_k)}), x_t^{(n_k)} \rangle \right] \longrightarrow \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\xi_t), x_t^* \rangle \right] \quad (k \rightarrow \infty).$$

Folglich genügt es zu zeigen, dass x^* zulässig für (4.4) ist, d.h. es gilt $x^* \in \mathcal{X}(\xi)$ und $x^* \in \mathcal{N}_{r'}(\xi) = \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t})$.

Da die Technologiematrizen als nicht-stochastisch vorausgesetzt wurden, hat die Menge $\mathcal{X}(\xi)$ die Form

$$\mathcal{X}(\xi) = \{x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : x \in X, Ax = h(\xi)\}, \quad (4.23)$$

wobei $X := \times_{t=1}^T X_t$, $h(\xi) := (h_1(\xi_1), \dots, h_T(\xi_T))$ und

$$A := \begin{pmatrix} A_{1,0} & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_{2,1} & A_{2,0} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & A_{T,1} & A_{T,0} \end{pmatrix}.$$

Der Graph von \mathcal{X} , gegeben durch

$$\text{graph } \mathcal{X} = \left\{ (x, \xi) \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \times L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) : x \in \mathcal{X}(\xi) \right\},$$

ist abgeschlossen und konvex. Aufgrund der Norm-Konvergenz von $(\xi^{(n_k)})$ gegen ξ in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ und der schwachen Konvergenz von $(x^{(n_k)})$ gegen x^* konvergiert die Folge $((x^{(n_k)}, \xi^{(n_k)}))$ geordneter Paare aus $\text{graph } \mathcal{X}$ schwach gegen (x^*, ξ) . Wegen der Abgeschlossenheit und der Konvexität von $\text{graph } \mathcal{X}$ liefert die Anwendung des Mazur'schen Theorems [24, Theorem 3.19], dass $\text{graph } \mathcal{X}$ schwach abgeschlossen ist und damit $(x^*, \xi) \in \text{graph } \mathcal{X}$ bzw. $x^* \in \mathcal{X}(\xi)$.

Schließlich ist noch zu zeigen, dass x^* zu $\mathcal{N}_{r'}(\xi)$ gehört, d.h. nichtantizipativ bezüglich ξ ist. Für ein beliebiges $y \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ gilt

$$\begin{aligned}
\left| \sum_{t=1}^T \langle y_t, x_t^* - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right| &\leq \left| \sum_{t=1}^T \langle y_t, x_t^* - x_t^{(n_k)} \rangle \right| \\
&\quad + \left| \sum_{t=1}^T \langle y_t, x_t^{(n_k)} - \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)] \rangle \right| \\
&\quad + \left| \sum_{t=1}^T \langle y_t, \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right| \\
&\leq \left| \sum_{t=1}^T \langle y_t, x_t^* - x_t^{(n_k)} \rangle \right| + \|y\|_r \bar{D}_f(\xi, \xi^{(n_k)}) \\
&\quad + \left| \sum_{t=1}^T \langle y_t, \mathbb{E}[x_t^{(n_k)} - x_t^* | \mathcal{F}_t(\xi)] \rangle \right|.
\end{aligned}$$

Der erste Term auf der rechten Seite der letzten Ungleichung konvergiert (aufgrund der schwachen Konvergenz von $(x^{(n_k)})$ gegen x^*) gegen Null genauso wie der zweite Term (aufgrund der Filtrationskonvergenz von $\bar{D}_f(\xi, \xi^{(n_k)})$ nach Voraussetzung). Auch der verbleibende dritte Term konvergiert gegen Null, weil die Abbildung \mathcal{E} , definiert durch

$$\mathcal{E}(x) := (\mathbb{E}[x_1 | \mathcal{F}_1(\xi)], \dots, \mathbb{E}[x_T | \mathcal{F}_T(\xi)]),$$

linear und beschränkt auf $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ ist und damit die Folge $(\mathcal{E}(x^{(n_k)}))$ schwach gegen $\mathcal{E}(x^*)$ konvergiert. Daher muss also $x_t^* = \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)]$ für alle $t = 1, \dots, T$ gelten und folglich gilt $x^* \in \mathcal{N}_{r'}(\xi)$. \square

4.2 Zweistufige Probleme

Für den Spezialfall zweistufiger stochastischer Optimierungsprobleme behalten natürlich alle in Abschnitt 4.1 hergeleiteten Aussagen ihre Gültigkeit. Dennoch ist es sinnvoll, den zweistufigen Fall auch speziell zu untersuchen. Betrachtet man zum Beispiel das Stabilitätsresultat für Optimalwerte in Theorem 4.1, so ergibt sich, dass im zweistufigen Fall der Filtrationsabstand innerhalb der Stabilitätsungleichung entfällt, d.h. wirklich nur der Verteilungsabstand der Eingangsdatenprozesse von Bedeutung ist. Die Abschätzung (4.10) für die Differenz der Optimalwerte des Zweistufenproblems (d.h. $T = 2$) geht über in die einfache Form:

$$|v(\xi) - v(\tilde{\xi})| \leq L \|\xi - \tilde{\xi}\|_r. \quad (4.24)$$

Die Tatsache, dass im zweistufigen Fall für Stabilitätsresultate in Bezug auf Optimalwerte und Erststufenentscheidungen nur der Abstand zwischen den Eingangsverteilungen von Bedeutung ist, ist bereits seit längerer Zeit bekannt.

Eine weitere Motivation der gesonderten Betrachtung der Zweistufigkeit resultiert aus der einfacheren Struktur des Zweistufenmodells, was die Formulierung speziellerer und damit oft verbesserter Ergebnisse gegenüber dem allgemeinen Fall erlaubt. So gelangt man durch die Wahl angepasster Fortet-Mourier-Metriken als Wahrscheinlichkeitsabstand (siehe Theorem 4.5) zu deutlich schärferen Abschätzungen als beispielsweise in (4.24). Der folgende Abschnitt fasst kurz die neuesten Resultate für zweistufige stochastische Optimierungsprobleme zusammen.

4.2.1 Stabilität zweistufiger Kompensationsmodelle

In zahlreichen Anwendungen in der Praxis werden zweistufige stochastische Optimierungsmodelle mit mehrperiodischer Kompensation verwendet. Typisch für solche Kompensationmodelle ist die Modellierung der Zweitstufenentscheidung als stochastischen (mehrperiodischen) Entscheidungsprozess. Das zweistufige (mehrperiodische) Optimierungsmodell besitzt die Form

$$\min \left\{ \int_{\Xi} f_0(\xi, x) P(d\xi) : x \in X \right\}, \quad (4.25)$$

wobei X eine polyedrische Menge in \mathbb{R}^m ist, Ξ eine abgeschlossene Teilmenge von \mathbb{R}^s ist und P ein Borel'sches Wahrscheinlichkeitsmaß auf Ξ ist. Der Integrand f_0 ist dabei von der Form

$$f_0(\xi, x) = \langle c, x \rangle + g(\xi, x), \quad \text{wobei}$$

$$g(\xi, x) := \inf \left\{ \sum_{j=1}^{\ell} \langle q_j(\xi), y_j \rangle \left| \begin{array}{l} W_1 y_1 = h_1(\xi) - T_1(\xi)x, y_1 \in Y_1, \\ W_j y_j = h_j(\xi) - T_j(\xi)y_{j-1}, \\ y_j \in Y_j, j = 2, \dots, \ell \end{array} \right. \right\}$$

mit $c \in \mathbb{R}^m$, polyedrischen Mengen Y_j aus \mathbb{R}^{m_j} , Kompensations- bzw. Recourse-Kosten $q_j(\xi) \in \mathbb{R}^{m_j}$, rechten Seiten $h_j(\xi) \in \mathbb{R}^{r_j}$, Technologiematrizen $T_j \in \mathbb{R}^{r_j \times m_{j-1}}$ und Recourse-Matrizen $W_j \in \mathbb{R}^{r_j \times m_j}$ für $j = 1, \dots, \ell$ und ein endliches $\ell \in \mathbb{N}$. Die Vektoren $q_j(\cdot)$, $h_j(\cdot)$ und die Matrizen $T_j(\cdot)$ können stochastisch sein, d.h. affin linear von ξ abhängen.

Das Problem der zweiten Stufe besitzt eine separable Blockstruktur und die Recourse-Variable y hat die Form $y = (y_1, \dots, y_\ell)$. Es lässt sich als Standard-Zweistufenmodell mit der Recourse-Entscheidung $y = (y_1, \dots, y_\ell)$ umformulieren, so dass die Recourse-Matrix eine Blockmatrix ist, wobei die Matrizen W_1, \dots, W_ℓ auf der Hauptdiagonalen und die Matrizen $T_1(\xi), \dots, T_\ell(\xi)$ unterhalb der Hauptdiagonalen angeordnet sind.

Das folgende Stabilitätsresultat für Optimalwerte $v(P)$ und ε -optimale Lösungsmengen $S_\varepsilon(P)$ von (4.25) wird in [81] bewiesen.

Theorem 4.5 *Es sei $P \in \mathcal{P}_{\ell+1}(\Xi)$ und die Lösungsmenge $S(P)$ von (4.25) sei nichtleer und beschränkt. Es gelte $h_j(\xi) - T_j(\xi)x \in W_j(Y_j)$ für alle $j = 1, \dots, \ell$ und all Paare $(\xi, x) \in \Xi \times X$ (Relatively-Complete-Recourse). Weiterhin gelte $\ker(W_j) \cap Y_j^\infty = \{0\}$ für alle $j = 1, \dots, \ell - 1$, wobei Y_j^∞ den (polyedrischen) Horizont-Kegel zu Y_j bezeichnet. Dann existieren Konstanten $L > 0$ und $\hat{\varepsilon} > 0$, so dass für alle $\varepsilon \in (0, \hat{\varepsilon})$ die Abschätzungen*

$$|v(P) - v(Q)| \leq L \overset{\circ}{\mu}_{\ell+1}(P, Q) \quad (4.26)$$

$$d_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{L}{\varepsilon} \overset{\circ}{\mu}_{\ell+1}(P, Q) \quad (4.27)$$

gelten, wenn $Q \in \mathcal{P}_{\ell+1}(\Xi)$ und $\overset{\circ}{\mu}_{\ell+1}(P, Q) < \varepsilon$. Der Abstand d_∞ bezeichnet hier den Pompeiu-Hausdorff Abstand auf abgeschlossenen und beschränkten Teilmengen von \mathbb{R}^m .

Der Fall $\ell = 1$ entspricht der Situation von linearen zweistufigen Modellen mit sogenannter fester Kompensation (Fixed-Recourse, siehe [78, Theorem 24]). In Verbindung mit den Resultaten in [70, 78] wird der Parameter r der entsprechenden Fortet-Mourier-Metrik $\overset{\circ}{\mu}_r$ (siehe auch Kapitel 3) je nach gegebener Situation im Problem (4.25) wie folgt gewählt:

$$r = \begin{cases} 1 & , \text{wenn entweder Kosten oder rechte Seiten zufällig sind,} \\ 2 & , \text{wenn Kosten und rechte Seiten zufällig sind,} \\ \ell + 1 & , \text{wenn zusätzlich auch alle Technologie-Matrizen zufällig sind.} \end{cases}$$

Kapitel 5

Szenarioreduktion

Die Modellierung des zwei- und mehrstufigen stochastischen Optimierungsproblems basiert auf der Verwendung einer in der Regel diskreten Wahrscheinlichkeitsverteilung bzw. eines diskreten stochastischen Prozesses. Es ergibt sich jedoch oft das Problem, dass selbst bei einer moderaten Anzahl von verwendeten Szenarien, die Modelle eine behandelbare Komplexität bei weitem übersteigen. Die Möglichkeit einer geeigneten Reduktion von Szenarien kann dann unverzichtbar sein. In letzter Zeit erschienen aufgrund der Bedeutung eine Reihe von relevanten Arbeiten zum Thema Szenarioreduktion innerhalb der stochastischen Optimierung (siehe [18, 19, 36, 40]).

Die Ideen der Szenarioreduktion resultieren aus Approximationsüberlegungen bezüglich unterschiedlicher Verteilungsabstände (vgl. Kapitel 3) und sind bisher insbesondere an zweistufige Modelle geknüpft. In der jüngsten Arbeit [40] werden die Reduktionstechniken in [36] erweitert, so dass sie eine direkte Anwendung der Stabilitätsabschätzungen (4.26) und (4.27) in Theorem 4.5 erlauben. Auch im mehrstufigen Fall besitzt die Szenarioreduktion ihre Bedeutung. Rekursives Anwenden der Reduktionstechniken bildet die Basis für konsistente Szenariobaumkonstruktionen (siehe Kapitel 6).

5.1 Prinzip der optimalen Szenarioreduktion

Es bezeichne P eine diskrete Wahrscheinlichkeitsverteilung mit Szenarien ξ^i und Wahrscheinlichkeiten p_i für $i \in \{1, \dots, N\}$. Mit Q_J sei eine approximative Verteilung bezeichnet, deren Träger eine Teilmenge von Szenarien umfasst:

$$\text{supp}(Q_J) = \{\xi^i : i \in \{1, \dots, N\} \setminus J\} \quad \text{und} \quad J \subset \{1, \dots, N\}.$$

Die zugehörigen Wahrscheinlichkeiten für Q seien mit

$$q_i \quad \text{für} \quad i \notin J$$

bezeichnet. Motiviert durch die Stabilitätsresultate ist es naheliegend, für eine vorgegebenen Anzahl von Elementen, eine Teilmenge J der Indexmenge

$\{1, \dots, N\}$ sowie Wahrscheinlichkeiten q_i für $i \notin J$ so zu bestimmen, so dass der Verteilungsabstand

$$\mu(P, Q_J) \quad (5.1)$$

möglichst klein wird. Das Funktional μ kann hier für einen der Abstände $\hat{\mu}_c$, $\overset{\circ}{\mu}_c$ oder ℓ_r (siehe Kapitel 3) stehen. Eine andere Möglichkeit wäre zu sagen, man maximiere die Anzahl der Elemente von J , so dass (5.1) kleiner ist als eine vorgegebene Tolleranz.

Das Prinzip der optimalen Szenarioreduktion besteht klassischer Weise im Folgenden: Man fixiere ein $n \in \mathbb{N}$ mit $n < N$ und bestimme eine Lösung des Minimumproblems

$$\min \left\{ \mu(P, Q_J) \mid J \subset \{1, \dots, N\}, \#J = N - n, q_i \geq 0, \sum_{i \notin J} q_i = 1 \right\}. \quad (5.2)$$

Ein erster Schritt zur Behandlung dieses Problems ist die Untersuchung der Abhängigkeiten zwischen den Wahrscheinlichkeiten q_i und der Indexmenge J . Insbesondere sind die optimalen Gewichte q_i für $i \notin J$ von Interesse, wenn die Indexmenge J festgehalten wird. Mit anderen Worten, man ist an der Lösung des Problems der besten Approximation

$$\min \left\{ \mu(P, Q_J) \mid q_i \geq 0, \sum_{i \notin J} q_i = 1 \right\} \quad (5.3)$$

interessiert.

Das folgende Resultat bezieht sich auf das letztere Problem (5.3) und wird in [19, Theorem 2] bewiesen. Es liefert eine explizite Darstellung des Infimums für den Monge-Kantorovich-Abstand, d.h. im Fall $\mu = \hat{\mu}_c$.

Theorem 5.1 *Für eine gegebene nichtleere Teilmenge $J \subset \{1, \dots, N\}$ besitzt das Problem (5.3) eine Lösung $Q_J^* = \sum_{i \notin J} q_i^* \delta_{\xi^i}$ und es gilt*

$$\begin{aligned} D_J &:= \hat{\mu}_c(P, Q_J^*) = \min \left\{ \hat{\mu}_c(P, Q_J) \mid q_i \geq 0, \sum_{i \notin J} q_i = 1 \right\} \\ &= \sum_{j \in J} p_j \min_{i \notin J} c(\xi^i, \xi^j), \end{aligned} \quad (5.4)$$

wobei

$$q_i^* = p_i + \sum_{\substack{j \in J \\ i(j)=i}} p_j \quad \text{mit} \quad i(j) \in \arg \min_{i \notin J} c(\xi^i, \xi^j), \quad i \notin J. \quad (5.5)$$

Das heißt, die optimale Neuverteilung besteht darin, die Wahrscheinlichkeit eines gestrichenen Szenarios, zu der Wahrscheinlichkeit eines in Bezug auf c nächstgelegenen Szenarios zu addieren.

Das Resultat lässt sich aufgrund der Darstellung für die L_r -minimalen Metriken (3.5) und der damit verbundenen Identität

$$\ell_r(P, Q) = (\hat{\mu}_c(P, Q))^{\frac{1}{r}} \quad \text{mit} \quad c(\xi^i, \xi^j) := \|\xi^i - \xi^j\|^r$$

ohne Schwierigkeit auf diese übertragen. Für den Optimalwert von Problem (5.3) im Fall $\mu = \ell_r$ ergibt sich dann

$$D_J = \left(\sum_{j \in J} p_j \min_{i \notin J} \|\xi^i - \xi^j\|^r \right)^{\frac{1}{r}}.$$

Die Neuverteilungsvorschrift (5.5) bleibt hier dieselbe. Definiert man ξ als stochastischen Prozess mit den Szenarien ξ^i ($i = 1, \dots, N$) sowie ξ_J als stochastischen Prozess mit den Szenarien ξ^i ($i \notin J$) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so dass

$$\xi_J(\omega) := \begin{cases} \xi^i & , \text{ falls } \xi(\omega) = \xi^i, i \notin J, \\ \xi^{i(j)} & , \text{ falls } \xi(\omega) = \xi^j, j \in J, \end{cases}$$

wobei $i(j)$ eine Abbildung wie in Theorem 5.1 ist, so gilt

$$\|\xi - \xi_J\|_r = D_J = \left(\sum_{j \in J} p_j \min_{i \notin J} \|\xi^i - \xi^j\|^r \right)^{\frac{1}{r}}. \quad (5.6)$$

Es bleibt die Frage nach einer Verallgemeinerung der Ergebnisse von Theorem 5.1 auf die Metriken vom Typ Fortet-Mourier. Wie in Abschnitt 4.2 von Kapitel 4 gezeigt wurde, sind diese speziell für zweistufige stochastische Optimierungsprobleme von großer Bedeutung. Für $\mu = \overset{\circ}{\mu}_c$ gilt folgende Aussage.

Theorem 5.2 *Für eine gegebene nichtleere Teilmenge $J \subset \{1, \dots, N\}$ besitzt das Problem (5.3) eine Lösung $Q_J^* = \sum_{i \notin J} q_i^* \delta_{\xi^i}$ und es gilt*

$$\begin{aligned} D_J &:= \overset{\circ}{\mu}_c(P, Q_J^*) = \min \left\{ \overset{\circ}{\mu}_c(P, Q_J) \mid q_i \geq 0, \sum_{i \notin J} q_i = 1 \right\} \\ &= \sum_{j \in J} p_j \min_{i \notin J} \hat{c}(\xi^i, \xi^j), \end{aligned} \quad (5.7)$$

wobei

$$q_i^* = p_i + \sum_{\substack{j \in J \\ i(j)=i}} p_j \quad \text{mit} \quad i(j) \in \arg \min_{i \notin J} \hat{c}(\xi^i, \xi^j), \quad i \notin J, \quad (5.8)$$

und wobei \hat{c} die reduzierten Kosten von c sind, d.h.

$$\hat{c}(\xi^i, \xi^j) := \min \left\{ \sum_{k=1}^{n-1} c(\xi^{l_k}, \xi^{l_{k+1}}) \mid n \in \mathbb{N}, l_k \in \{1, \dots, N\}, l_1 = i, l_n = j \right\}.$$

Beweis: Aufgrund von Proposition 3.2 gilt die Identität

$$\overset{\circ}{\mu}_c(P, Q_J) = \hat{\mu}_{\hat{c}}(P, Q_J),$$

wobei die reduzierten Kosten \hat{c} eine Metrik auf dem Träger $\{\xi^1, \dots, \xi^N\}$ von P definieren. Die Anwendung von Theorem 5.1 für das Monge-Kantorovich-Funktional impliziert die behauptete Aussage, denn es gilt

$$\min \left\{ \hat{\mu}_{\hat{c}}(P, Q_J) \mid q_i \geq 0, \sum_{i \notin J} q_i = 1 \right\} = \sum_{j \in J} p_j \min_{i \notin J} \hat{c}(\xi^i, \xi^j)$$

zusammen mit der angegebenen Formel zur Neuverteilung der Wahrscheinlichkeitsgewichte. \square

Unter Verwendung der Resultate von Theorem 5.1 und 5.2 kann das Problem (5.2) der optimalen Szenarioreduktion nun in der Form

$$\min \{ D_J \mid J \subset \{1, \dots, N\}, \#J = N - n \}$$

formuliert werden. Das Funktional D_J wird dabei je nach verwendeten Abstand μ durch die Formeln (5.4), (5.6) bzw. (5.7) bestimmt. All diese Probleme führen auf ein Optimierungsproblem von der Form

$$\min \left\{ \sum_{j \in J} p_j \min_{i \notin J} \bar{c}(\xi^i, \xi^j) \mid J \subset \{1, \dots, N\}, \#J = N - n \right\}, \quad (5.9)$$

mit geeigneter Kostenfunktion \bar{c} . Die Aufgabe besteht in der Bestimmung einer kostenoptimalen Indexmenge J mit vorgeschriebener Kardinalität $1 \leq n = \#J < N$. Probleme dieser Art sind aufgrund ihrer kombinatorischen Struktur im Allgemeinen NP-schwer. Das Optimierungsproblem (5.9) stellt insbesondere für $\bar{c} = \hat{c}$ ein sogenanntes k -Median-Problem im metrischen Raum (Ξ, \bar{c}) dar. In [7] und [91, Chapter 25] werden Approximationsmethoden für das k -Median-Problem beschrieben, die das Problem mit polynomialen Aufwand bei einer Genauigkeitsgarantie vom $6\frac{2}{3}$ - bzw. 6-fachen des Optimums lösen.

5.2 Heuristische Algorithmen

Für kombinatorische Probleme ist die Verwendung polynomialer Näherungsalgorithmen oft der einzige Weg zur Bestimmung brauchbarer Lösungen. In [19] und [36] wurden speziell für das Problem der optimalen Szenarioreduktion vielversprechende heuristische Algorithmen, der Vorwärts- und Rückwärtsalgorithmus, zur näherungsweise Lösung von (5.9) entworfen. Bei der Konzeption dieser Algorithmen stand weniger die möglichst exakte Lösung des Problems im Vordergrund als vielmehr die Einfachheit der Algorithmen und die daraus resultierende Schnelligkeit.

5.2.1 Vorwärtsalgorithmus

Die Idee des Vorwärtsalgorithmus zur Szenarioreduktion besteht in der Ausnutzung der einfachen Struktur des Problems im Spezialfalls $n = 1$. In diesem Fall nimmt das Problem (5.9) die Form

$$\min_{u \in \{1, \dots, N\}} \sum_{\substack{j=1 \\ j \neq u}}^N p_j \bar{c}(\xi^u, \xi^j)$$

an. Wird das Minimum in u^* angenommen, erhält man eine optimale Lösung durch $J = \{1, \dots, N\} \setminus \{u^*\}$. Diese Idee der schrittweisen Auswahl von Szenarien kann zu einer Heuristik zur approximativen Lösung von (5.9) fortgesetzt werden. Das Resultat nach k Schritten ist eine Indexmenge $J^{[k]}$ mit einer Anzahl von noch $N - k$ Elementen.

Algorithmus 5.3 (*Vorwärtsalgorithmus Szenarioreduktion*)

Schritt [0]: $J^{[0]} := \{1, \dots, N\}$.

Schritt [k]: $u_k \in \arg \min_{u \in J^{[k-1]}} \sum_{j \in J^{[k-1]} \setminus \{u\}} p_j \min_{i \in J^{[k-1]} \setminus \{u\}} \bar{c}(\xi^i, \xi^j),$

$J^{[k]} := J^{[k-1]} \setminus \{u_k\}$.

Schritt [n+1]: *Optimale Neuverteilung mit $J := J^{[n]}$.*

Der Algorithmus kann in folgender Weise interpretiert werden. Der Szenarioindex u_k in Schritt k wird so gewählt, dass der Wert des Funktionals $D_{J^{[k]}}$ minimiert wird, d.h.

$$D_{J^{[k-1]} \setminus \{u_k\}} = \min_{u \in J^{[k-1]}} D_{J^{[k-1]} \setminus \{u\}}.$$

Die Menge der verbleibenden Szenarien und damit der Träger des Maßes Q_J ist am Ende durch die Indexmenge $\{u_1, \dots, u_n\}$ bestimmt.

5.2.2 Rückwärtsalgorithmus

Die Idee des Rückwärtsalgorithmus zur Szenarioreduktion basiert auf dem zweiten Spezialfall des Problems (5.9). Dieser entsteht für $n = N - 1$ und besitzt die Form

$$\min_{l \in \{1, \dots, N\}} p_l \min_{i \neq l} \bar{c}(\xi^i, \xi^l).$$

Wird das Minimum in l^* angenommen, ergibt sich $J = \{l^*\}$, d.h. dieser Spezialfall führt zum Streichen des Szenarios mit dem Index l^* . Auch dieses Vorgehen lässt sich schrittweise zu einem heuristischen Algorithmus zur approximativen Lösung von (5.9) fortsetzen. Das Ergebnis nach $N - k$ Schritten ist nun eine Indexmenge $J^{[N-k]}$ mit einer Anzahl von $N - k$ Elementen.

Algorithmus 5.4 (*Rückwärtsalgorithmus Szenarioreduktion*)

Schritt [0]: $J^{[0]} := \emptyset$.

Schritt [k]: $l_k \in \arg \min_{l \notin J^{[k-1]}} \sum_{j \in J^{[k-1]} \cup \{l\}} p_j \min_{i \notin J^{[k-1]} \cup \{l\}} \bar{c}(\xi^i, \xi^j),$
 $J^{[k]} := J^{[k-1]} \cup \{l_k\}.$

Schritt [N-n+1]: *Optimale Neuverteilung mit $J := J^{[N-n]}$.*

Die Interpretation des Algorithmus ist dieselbe wie zuvor. Der Szenarioindex l_k in Schritt k wird so gewählt, dass der Wert des Funktionals $D_{J^{[k]}}$ minimiert wird, d.h.

$$D_{J^{[k-1]} \cup \{l_k\}} = \min_{l \notin J^{[k-1]}} D_{J^{[k-1]} \cup \{l\}}.$$

Die Menge der verbleibenden Szenarien und damit der Träger des Maßes Q_J ist hier durch die Indexmenge $\{1, \dots, N\} \setminus \{l_1, \dots, l_{N-n}\}$ bestimmt.

Abbildung 5.1 zeigt ein Beispiel der Szenarioreduktion angewendet auf ein Sampling der zweidimensionalen Standardnormalverteilung mit 10 000 Szenarien. Die Reduktion wurde mit dem Vorwärtsalgorithmus und bezüglich des Monge-Kantorovich-Funktional unter Verwendung des Euklidischen Abstands bestimmt. Reduziert wurde auf 20 Szenarien, die entsprechend ihrer neuen Wahrscheinlichkeit unterschiedlich groß dargestellt sind.

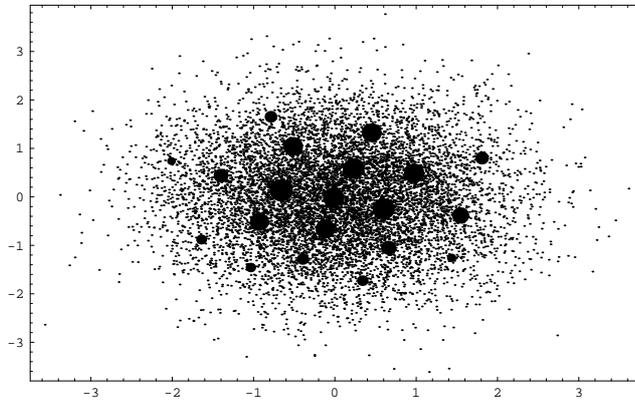


Abbildung 5.1: Ergebnis der Szenarioreduktion (Vorwärtsalgorithmus) ausgehend von einem Sampling mit 10 000 normalverteilten Szenarien. Die durch die Szenarioreduktion ermittelten 20 Szenarien werden entsprechend ihrer neuen Wahrscheinlichkeit unterschiedlich groß dargestellt.

Beide heuristischen Ansätze wurden in [36] im Fall $\mu = \hat{\mu}_c$ ausführlich studiert. Dort wurde gezeigt, dass beide Algorithmen polynomiale Komplexität besitzen. Obwohl keiner der beiden Algorithmen im Allgemeinen zu optimalen

Lösungen führt, so ist die Leistungsfähigkeit dennoch recht beachtlich. Welcher der Algorithmen eingesetzt werden sollte, richtet sich in erster Linie nach der Geschwindigkeit. Algorithmus 5.3 sollte für $n \leq \frac{N}{4}$ verwendet werden, d.h. immer dann, wenn auf nur wenige Szenarien reduziert werden soll. Ansonsten ist Algorithmus 5.4 vorzuziehen. Eine Einschätzung der Heuristiken hinsichtlich Schnelligkeit und Genauigkeit sowie ein Vergleich mit CPLEX, einem der leistungsfähigsten Standardlöser für lineare und gemischt-ganzzahlige Optimierungsprobleme, ist im Kapitel 5 zu finden.

In [43] wird eine Erweiterung der Algorithmen zur Szenarioreduktion bei der Verwendung von Wahrscheinlichkeitsdiskrepanzen entwickelt. Diese sind durch aktuelle Stabilitätsresultate für stochastische Optimierungsprobleme mit Wahrscheinlichkeitsrestriktionen und für Optimierungsprobleme mit Ganzzahligkeitsforderungen motiviert (siehe [15, 44] sowie [70, 84]).

5.3 Bestimmung der reduzierten Kosten

In den vorherigen Abschnitten wurde gezeigt, wie das Problem der optimalen Szenarioreduktion auf ein kombinatorisches Problem des Typs (5.9) zurückgeführt werden kann und wie dieses approximativ mit einer der beiden beschriebenen Algorithmen näherungsweise gelöst werden kann. Offen bleibt die Frage nach der Bestimmung der reduzierten Kosten $\bar{c} := \hat{c}$ im Fall der Verwendung der Fortet-Mourier-Metriken $\overset{\circ}{\mu}_c$.

Für eine endliche Menge von Szenarien $\{\xi^1, \dots, \xi^N\}$ und gegebener Kostenfunktion $c(\cdot, \cdot)$ besitzen die reduzierten Kosten $\hat{c}(\cdot, \cdot)$ die Darstellung

$$\hat{c}(\xi^i, \xi^j) := \min \left\{ \sum_{k=1}^{n-1} c(\xi^{l_k}, \xi^{l_{k+1}}) \mid n \in \mathbb{N}, l_k \in \{1, \dots, N\}, l_1 = i, l_n = j \right\}$$

(vgl. Theorem 5.2). Zwischen je zwei Szenarien ξ^i und ξ^j ist der kostengünstigste Weg bzgl. der durch c definierten Kosten zu ermitteln. Dieses Problem ist in der Literatur auch als *Shortes-Paths-Problem* bekannt (siehe z.B. [10, Kapitel 25 und 26]).

5.3.1 Dijkstra-Algorithmus

Das Shortest-Paths-Problem ist eines der am einfachsten zu behandelnden Netzwerkflussprobleme und lässt sich als Spezialfall eines Minimum-Cost-Flow-Problems formulieren.

Im Folgenden wird ein einfacher Algorithmus zur Bestimmung aller kostenoptimalen Wege eines ausgewählten Szenarios zu allen übrigen beschrieben, der auf dem bekannten Algorithmus von Dijkstra zur Lösung des *Single-Source-Shortes-Paths-Problem* basiert. Dazu bezeichnen

$$V = \{1, \dots, N\} \quad \text{und} \quad E = \{(i, j) \mid i, j \in V\}$$

die Mengen aller Punkte bzw. gerichteter Verbindungen (Kanten). Zusammen definieren sie einen gerichteten Graphen $G = (V, E)$. Die Kosten für eine Verbindung $(i, j) \in E$ sind durch

$$c_{ij} := c(\xi^i, \xi^j), \quad i, j = 1, \dots, N,$$

gegeben. Für ein fest gewähltes Element $s \in V$ besteht die Idee des Algorithmus in der schrittweisen Aufdatierung einer Menge $S \subseteq V$ von Punkten, deren kostenoptimale Wege zu s bereits bestimmt wurden sowie in einer schrittweisen Relaxierung nichtoptimaler Wege.

Algorithmus 5.5 (*Algorithmus zur Bestimmung reduzierter Kosten*)

Schritt [0]: Wähle ein $s \in V$:

$$\text{Setze } S^{[0]} := \emptyset, \quad Q^{[0]} := V \setminus \{s\},$$

$$d_v^{[0]} := \begin{cases} c_{sv} & , \text{ für } v \in Q^{[0]}, \\ 0 & , v = s. \end{cases}$$

Schritt [k]: Wähle $u_k \in \arg \min_{u \in Q^{[k-1]}} d_u^{[k-1]}$.

$$\text{Setze } S^{[k]} := S^{[k-1]} \cup \{u_k\}, \quad Q^{[k]} := Q^{[k-1]} \setminus \{u_k\},$$

$$d_v^{[k]} := \begin{cases} \min\{d_v^{[k-1]}, c_{vu_k} + d_{u_k}^{[k-1]}\} & , v \in Q^{[k]}, \\ d_v^{[k-1]} & , \text{ sonst.} \end{cases}$$

Schritt [N]: Setze $\hat{c}(\xi^s, \xi^v) := d_v^{[N-1]}$ für $v = 1, \dots, N$.

Der Algorithmus nutzt im Wesentlichen die für diese Art von Problemen typische Eigenschaft aus, dass jeder Teilweg eines optimalen Wegs selbst auch wieder optimal ist. Dieses Prinzip ist auch Grundlage für die Anwendung von Techniken der *dynamischen Optimierung* sowie der *Greedy-Algorithmen*. Es zeigt sich, dass der Dijkstra-Algorithmus einen Greedy-Algorithmus (vgl. [58, Kapitel 9]) darstellt. Die Komplexität hängt von der Anzahl der Punkte ab. Es ergibt sich für $|V| = N$ ein Aufwand von $\mathcal{O}(N^2)$ für Algorithmus 5.5.

Abschließend sei bemerkt, dass der hier beschriebene Algorithmus leicht zur Bestimmung aller kostenoptimalen Wege zwischen je zwei Punkten erweitert werden kann, so dass der resultierende Gesamtaufwand von $\mathcal{O}(N^3)$ nicht überschritten wird. Damit hat die Ermittlung der reduzierten Kosten $\bar{c} = \hat{c}$ insgesamt dieselbe Komplexitätsordnung wie die heuristischen Algorithmen 5.3 und 5.4 zur optimalen Szenarioreduktion.

Kapitel 6

Konstruktion multivariater Szenariobäume

Die Konstruktion von Szenariobäumen übernimmt bei der Modellierung mehrstufiger stochastischer Optimierungsprobleme eine zentrale Rolle. Dabei geht es sowohl um die Beschreibung der Verteilungsinformation der im Modell als unsicher angenommenen Eingangsgrößen als auch um die Abbildung der Informationsstruktur. Szenariobäume dienen zusätzlich der numerischen Behandelbarkeit des Optimierungsproblems, da diese in der Regel die Verwendung einer begrenzten Anzahl von Szenarien erfordern. Basierend auf den Stabilitätsresultaten in Kapitel 4 und unter Verwendung der Techniken zur optimalen Szenarioreduktion in Kapitel 5 stehen zwei rekursive Algorithmen, der Vorwärts- und Rückwärtsalgorithmus zur schrittweisen Konstruktion von Szenariobäumen, im Vordergrund dieses Kapitels. Während sich der erste Teil zunächst der detaillierte Beschreibung der Algorithmen widmet, werden im zweiten Teil Konvergenzeigenschaften diskutiert.

6.1 Die Konstruktionsalgorithmen

Bevor in den Abschnitten 6.1.3 und 6.1.4 mit der mathematischen Formulierung der beiden Konstruktionsalgorithmen zur Szenariobaumgenerierung begonnen wird, sollen in den folgenden zwei Abschnitten das Grundkonzept erläutert sowie generelle Bezeichnungen festgelegt werden.

6.1.1 Konzeption

Es sei ξ der originale stochastische Prozess definiert auf irgendeinem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Zeitparameter $\{1, \dots, T\}$, wobei die Realisierungen zu verschiedenen Zeitstufen Werte im \mathbb{R}^d annehmen. Ziel ist die Generierung eines approximativen diskreten Prozesses ξ_{tr} auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so dass folgende Eigenschaften gelten:

- (i) der Träger von ξ_{tr} ist endlich,
- (ii) die Szenarien besitzen (im Allgemeinen) eine Baumstruktur,
- (iii) sowohl der Verteilungsabstand als auch der Informationsabstand zwischen ξ_{tr} und ξ sind klein.

Eigenschaft (iii) sichert dann aufgrund des Stabilitätsresultats in Theorem 4.1, dass die Optimalwerte der stochastischen Optimierungsprobleme (2.3) mit Eingangsprozessen ξ bzw. ξ_{tr} nahe beieinander liegen, also

$$|v(\xi) - v(\xi_{\text{tr}})|$$

klein ist. In Abschnitt 6.2 wird später gezeigt, dass der Prozess ξ in (iii) auch durch eine gute Näherung $\hat{\xi}$ ersetzt werden kann. Mit einer guten Näherung ist hier gemeint, dass sich die Optimalwerte der jeweiligen Probleme nur wenig unterscheiden.

Die Konzeption zur Generierung des Szenariobaums ξ_{tr} besteht im Wesentlichen aus zwei Schritten:

- (1) Approximation der Verteilung durch (individuelle) Szenarien,
- (2) Rekonstruktion der Informationsstruktur.

Im ersten Schritt soll zunächst die Verteilung des originalen Eingangsprozess ξ durch eine gute diskreten Approximation mit einer endlichen Anzahl von Szenarien wiedergegeben werden. Eine diskrete Näherung in dieser Form wird auch Szenariofächer (oder Szenariofan) genannt und mit ξ_f bezeichnet. Im Allgemei-

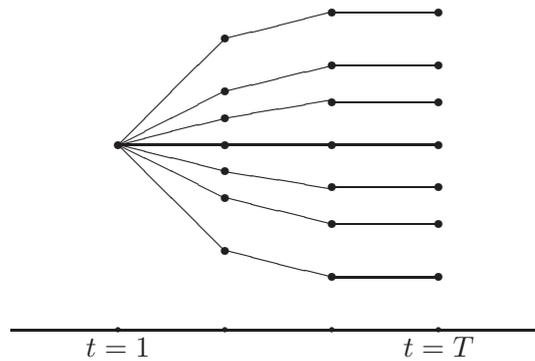


Abbildung 6.1: Szenariofächers individueller Szenarien mit $T = 4$ und $N = 7$

nen besitzt ein solche Approximation keinerlei Baumstruktur und besteht so in der Regel aus individuellen Szenarien wie in Abbildung 6.1. Der Szenariofächer ξ_f sei im Folgenden gegeben durch die individuellen Szenarien

$$\xi^i = (\xi_1^i, \dots, \xi_T^i) \quad \text{mit Wahrscheinlichkeiten } p_i \quad \text{für } i = 1, \dots, N.$$

Dabei gelte $\xi_1^1 = \dots = \xi_1^N =: \xi_1^*$. Weiterhin sei vorausgesetzt, dass ξ_f eine gute Approximation von ξ bzgl. der Verteilung ist, d.h. in dem Sinne eine gute Näherung ist, dass für den L_r -Abstand auf einem Wahrscheinlichkeitsraum

$$\|\xi - \xi_f\|_r \leq \varepsilon_{\text{appr}} \quad (6.1)$$

gilt, wobei die Approximationsgenauigkeit $\varepsilon_{\text{appr}}$ klein ist.

In den meisten praktisch relevanten Anwendungen sind derartige Approximationen mit individuellen Szenarien ohne größere Schwierigkeiten realisierbar. Eine weit verbreitete Möglichkeit besteht in der Auswertung verfügbarer historischer Daten. Insbesondere die Anpassung statistischer Zeitreihenmodelle und der darauf basierenden Simulation von Szenarien führt in vielen Fällen zu guten Anfangsnäherungen in Form individueller Szenarien.

Der zweite Schritt befasst sich mit dem Problem der Informationsstruktur. Individuelle Szenarien weisen in den meisten Fällen formal keine Informationsstruktur auf. Das gilt insbesondere dann, wenn die Szenarien durch eine endliche Simulation eines stetigen Prozesses generiert werden. Das liegt darin begründet, dass die von ξ_f erzeugten σ -Algebren $\mathcal{F}_t(\xi_f)$ in der Regel unabhängig von $t \in \{2, \dots, T\}$ sind. Der zweite Schritt zur Generierung von Szenariobäumen (die eigentliche Baumkonstruktion) besteht daher darin, die Informationsstruktur des originalen Prozesses, durch geeignetes auf Diskretisierungen basierendem Clustering, zu rekonstruieren. Dabei wird wie folgt vorgegangen: Ausgehend von ξ_f wird ein Prozess ξ_{tr} durch rekursives Zusammenlegen ähnlicher Szenarien auf Zeitintervallen $\{1, \dots, t\}$ mit absteigendem bzw. ansteigendem Zeitparameter t bestimmt (siehe Abschnitt 6.1.3 bzw. 6.1.4), wobei die Techniken der Szenarioreduktion (vgl. Kapitel 5) zur Anwendung kommen.

Die spezielle Form des Clustering bei der Generierung der Szenariobaumprozesse ξ_{tr} erlaubt zunächst eine Abschätzung der Form

$$\|\xi_f - \xi_{\text{tr}}\|_r \leq \varepsilon_r,$$

womit sich mit (6.1) insgesamt die L_r -Abschätzung

$$\|\xi - \xi_{\text{tr}}\| \leq \varepsilon_{\text{appr}} + \varepsilon_r$$

ergibt. Der Gesamtfehler hinsichtlich des L_r -Abstands zwischen dem originalen Eingangsprozess ξ und dem Szenariobaumprozess ξ_{tr} hängt also von der Güte der Ausgangsnäherung durch die individuellen Szenarien sowie von der Fehlergenauigkeit ε_r ab. Letztere kann direkt innerhalb der Algorithmen gesteuert werden. Die Genauigkeit ε_r beeinflusst aber auch entscheidend die Baumstruktur von ξ_{tr} . Insbesondere die Wahl der Genauigkeiten innerhalb der verschiedenen Zeitstufen (siehe Algorithmus 6.2 und 6.5) entscheidet maßgeblich über die jeweilige Verzweigungsstruktur. In Abschnitt 6.2 wird gezeigt, dass im Zusammenspiel mit der Anzahl der individuellen Szenarien neben der Verteilungsinformation auch die Rekonstruktion der Informationsstruktur erfolgen kann, womit ebenso eine Approximation bzgl. des Filtrationsabstands möglich ist.

6.1.2 Generelle Bezeichnungen

In der Beschreibung der Algorithmen der folgenden zwei Abschnitte werden L_r -Seminormen $\|\cdot\|_{r,t}$ auf $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ mit $s = Td$ verwendet, die durch

$$\|\xi\|_{r,t} := \left(\mathbb{E}[\|\xi\|_t^r] \right)^{\frac{1}{r}} = \left(\sum_{i=1}^N p_i \|\xi^i\|_t^r \right)^{\frac{1}{r}} \quad (6.2)$$

gegeben sind und die jeweils im Schritt t der Konstruktionsalgorithmen benötigt werden. Hierbei bezeichnet $\|\cdot\|_t$ die Seminorm auf \mathbb{R}^d definiert durch

$$\|\xi\|_t := \|(\xi_1, \dots, \xi_t, 0, \dots, 0)\|$$

für ein $\xi = (\xi_1, \dots, \xi_T) \in \mathbb{R}^{Td}$.

6.1.3 Der Rückwärtsalgorithmus

Aufgrund der etwas einfacheren Darstellung wird mit der Beschreibung des Rückwärtsalgorithmus begonnen. Die Rückwärtsstrategie basiert auf einer rekursiven Szenarioreduktion angewendet auf Zeitabschnitte $\{1, \dots, t\}$ mit absteigendem Zeitparameter t . Ausgehend von einem Szenariofächer ξ_t als Startprozess, d.h. es sei

$$\bar{\xi}^{T+1} := \xi_t,$$

ist für $t = T \dots, 1$ eine Folge von stochastischen Prozessen

$$\bar{\xi}^t \text{ mit Szenarien } \{\bar{\xi}^{t,i} := \xi^i\}_{i \in I_t} \text{ und Wahrscheinlichkeiten } \{\pi_t^i\}_{i \in I_t}$$

das Ergebnis, wobei die Indexmengen I_t jeweils Teilmengen von $I := \{1, \dots, N\}$ bilden. Die Methode arbeitet dabei so, dass rückwärts (mit $I_{T+1} := I$ beginnend) eine Kette von Indexmengen mit der Eigenschaft

$$I_1 = \{i_*\} \subseteq I_2 \subseteq \dots \subseteq I_{t-1} \subseteq I_t \subseteq \dots \subseteq I_T \subseteq I_{T+1} = I$$

entsteht. Die Differenz der Indexmengen I_{t+1} und I_t wird jeweils im Schritt t mit Hilfe eines Szenarioreduktionsschritts bzgl. der Seminorm $\|\cdot\|_{r,t}$ bestimmt. Mit den gleichen Bezeichnungen wie in Kapitel 5 sei

$$I_t := I_{t+1} \setminus J_t \quad \text{für jedes } t = 1, \dots, T, \quad (6.3)$$

das heißt, J_t bezeichne die Indexmenge der im Schritt t gestrichenen Szenarien. Dementsprechend werden die Wahrscheinlichkeiten π_t^i rekursiv gemäß der Neuverteilungsvorschrift (5.5) von Theorem 5.1 berechnet. Es sei $\pi_{T+1}^i := p_i$ ($i \in I_{T+1}$) und

$$\pi_t^i = \pi_{t+1}^i + \sum_{\substack{j \in J_t \\ i_t(j)=i}} \pi_{t+1}^j \quad (i \in I_t) \quad \text{mit} \quad i_t(j) \in \arg \min_{i \in I_t} \|\xi^i - \xi^j\|_t^r. \quad (6.4)$$

Theorem 5.1 sowie die Folgerung (5.6) liefern für die so konstruierten stochastischen Prozesse die Eigenschaft

$$\|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}^r = \sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} \|\xi^i - \xi^j\|_t^r. \quad (6.5)$$

Die Vorschrift $i_t(j)$ definiert die Zuordnungsabbildung entsprechend der optimalen Neuverteilung von der Menge der zum Zeitpunkt t eliminierten Szenarien in die Menge der verbleibenden Szenarien, d.h. man erhält $i_t(j) : J_t \rightarrow I_t$. Setzt man

$$J_{t,i} := \{j \in J_t \mid i_t(j) = i\} \quad (i \in I_t),$$

so ergibt sich mit $1 + |J_{t,i}|$ der Verzweigungsgrad eines jeden Szenarios $i \in I_t$ zum Zeitpunkt t . Für die abschließende Konstruktion des Szenariobaums, d.h. eines Prozesses ξ_{tr} mit Baumstruktur, werden die partiellen Vorschriften $i_t(\cdot)$ rekursiv zu einer Folge von gesamtheitlichen Stufenzuweisungen $\alpha_t : I \rightarrow I_t$ erweitert. Dazu sei α_{T+1} die Identität und

$$\alpha_t(j) := \begin{cases} i_t(\alpha_{t+1}(j)) & , \alpha_{t+1}(j) \in J_t, \\ \alpha_{t+1}(j) & , \text{sonst,} \end{cases} \quad (6.6)$$

für $j \in I$ und $t = 1, \dots, T$. Der Szenariobaum ξ_{tr} wird dann durch $|I_T|$ Szenarien ξ_{tr}^j für $j \in I_T$ gebildet, wobei die Szenarienwahrscheinlichkeiten durch π_T^j gegeben sind. Jede Komponente $\xi_{\text{tr},t}^j$ beschreibt einen Knoten mit Knotenwahrscheinlichkeit π_t^j im Szenariobaum und gehört jeweils zur Menge $\{\xi_t^i\}_{i \in I_t}$. Der entsprechende Index $i \in I_t$ ist durch $i = \alpha_t(j)$ bestimmt. Insgesamt ergeben sich so mit

$$\bar{I}_{t,i} := \{j \in I : \alpha_t(j) = i\} \quad (i \in I_t)$$

für alle $i \in I_t$ jeweils verschiedene Cluster ähnlicher Szenarien für den Zeitraum $\{1, \dots, t\}$, die zusammen für jedes t eine Partition der Szenariomenge $\{1, \dots, N\}$ bilden. Eine Abschätzung für den L_r -Abstand zwischen ξ_{tr} und ξ_f liefert das folgende Theorem.

Theorem 6.1 *Es sei ξ_f ein stochastischer Prozess mit fixiertem Anfangsknoten ξ_1^* , Szenarien ξ^i und Wahrscheinlichkeiten p_i , $i = 1, \dots, N$. Weiter sei ξ_{tr} ein stochastischer Prozess mit Szenarien*

$$\xi_{\text{tr}}^i = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^i \right) \text{ und Wahrscheinlichkeiten } \pi_T^i, i \in I_T.$$

Dann gilt die Abschätzung

$$\|\xi_f - \xi_{\text{tr}}\|_r \leq \sum_{t=2}^T \left(\sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} \|\xi^i - \xi^j\|_t^r \right)^{\frac{1}{r}}. \quad (6.7)$$

Beweis: Es sei $(\hat{\xi}^\tau)_\tau$ eine Folge stochastischer Prozesse, definiert über Szenarien $\hat{\xi}^{\tau,i}$ mit Wahrscheinlichkeiten π_T^i für $i \in I_T$, wobei

$$\hat{\xi}_t^{\tau,i} := \begin{cases} \xi_t^{\alpha_t(i)} & , t \geq \tau, \\ \xi_t^{\alpha_\tau(i)} & , t < \tau, \end{cases} \quad (t = 1, \dots, T)$$

für $\tau = 1, \dots, T$. Entsprechend der Konstruktion gilt $\hat{\xi}^T = \bar{\xi}^T$ und $\hat{\xi}^1 = \xi_{\text{tr}}$. Die Prozesse $\hat{\xi}^\tau$ sind in der Abbildung 6.2 veranschaulicht (wobei die Bildsequenz der Folge von Prozessen $\xi_t, \hat{\xi}^T, \dots, \hat{\xi}^2, \xi_{\text{tr}}$ entspricht).

Als nächstes wird die Beziehung

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}. \quad (6.8)$$

für alle $t = 1, \dots, T-1$ gezeigt. Es gilt offenbar

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \sum_{i \in I_T} \pi_T^i \|\hat{\xi}^{t+1,i} - \hat{\xi}^{t,i}\|_r \quad (6.9)$$

und da alle Komponenten von $\hat{\xi}^{t+1,i}$ und $\hat{\xi}^{t,i}$ ab $T-t$ identisch sind, kann in (6.9) die Norm $\|\cdot\|$ auch durch die Seminorm $\|\cdot\|_t$ ersetzt werden. Da darüber hinaus die ersten t Komponenten von $\hat{\xi}^{t+1,i}$ und $\hat{\xi}^{t,i}$ gerade $\xi_\tau^{\alpha_{t+1}(i)}$ bzw. $\xi_\tau^{\alpha_t(i)}$ für $\tau = 1, \dots, t$ sind, gilt weiter

$$\sum_{i \in I_T} \pi_T^i \|\hat{\xi}^{t+1,i} - \hat{\xi}^{t,i}\|_r = \sum_{i \in I_T} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r.$$

Wegen der Eigenschaft $\alpha_t(j) = \alpha_{t+1}(j)$ im Fall $\alpha_{t+1}(j) \notin J_t$ erhält man

$$\sum_{i \in I_T} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r = \sum_{\substack{i \in I_T \\ \alpha_{t+1}(i) \in J_t}} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r.$$

Unter Verwendung der Definition für α_t (6.6) und mit der Eigenschaft (6.5) kann die letzte Summe wie folgt umgeformt werden:

$$\begin{aligned} \sum_{\substack{i \in I_T \\ \alpha_{t+1}(i) \in J_t}} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r &= \sum_{j \in J_t} \sum_{\substack{k \in I_T \\ \alpha_{t+1}(k)=j}} \pi_T^k \|\xi^{\alpha_{t+1}(k)} - \xi^{\alpha_t(k)}\|_t^r \\ &= \sum_{j \in J_t} \left(\sum_{\substack{k \in I_T \\ \alpha_{t+1}(k)=j}} \pi_T^k \right) \|\xi^j - \xi^{i_t(j)}\|_t^r \\ &= \sum_{j \in J_t} \pi_{t+1}^j \|\xi^j - \xi^{i_t(j)}\|_t^r \\ &= \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}^r, \end{aligned}$$

womit der Beweis der Formel (6.8) für $t = 1, \dots, T-1$ komplett ist.

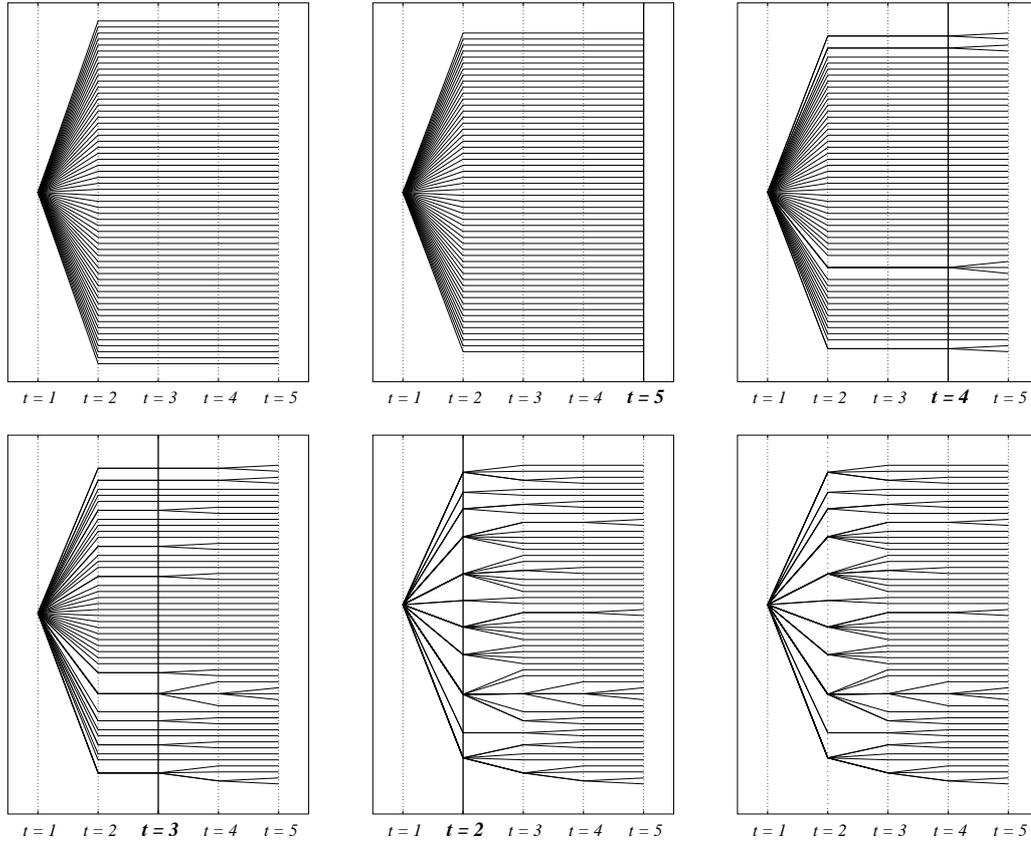


Abbildung 6.2: Illustration der Rückwärtskonstruktion von Szenariobäumen an einem Beispiel mit $T = 5$ Zeitperioden, ausgehend von einem Szenariofächer individueller Szenarien.

Damit kann jetzt die Abschätzung (6.7) durch wiederholtes Anwenden der Dreiecksungleichung für die L_r -Norm $\|\cdot\|_r$ unter Ausnutzung von (6.8) sowie der Identitäten $\xi_f = \bar{\xi}^{T+1}$, $\hat{\xi}^T = \bar{\xi}^T$ und $\hat{\xi}^1 = \xi_{tr}$ gezeigt werden:

$$\begin{aligned}
 \|\xi_f - \xi_{tr}\|_r &\leq \|\xi_f - \hat{\xi}^T\|_r + \|\hat{\xi}^T - \xi_{tr}\|_r \\
 &\leq \|\bar{\xi}^{T+1} - \bar{\xi}^T\|_r + \sum_{k=1}^{T-1} \|\hat{\xi}^{T-k+1} - \hat{\xi}^{T-k}\|_r \\
 &= \sum_{k=0}^{T-1} \|\bar{\xi}^{T-k+1} - \bar{\xi}^{T-k}\|_{r, T-k} \\
 &= \sum_{t=1}^T \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r, t},
 \end{aligned}$$

wobei der Summand für $t = 1$ in der letzten Summe verschwindet. Zusammen mit der Eigenschaft (6.5) von $\|\cdot\|_{r, t}$, ist die Behauptung damit bewiesen. \square

Das bewiesene Resultat erlaubt die Kontrolle über die Qualität der Szenariobäume innerhalb des Konstruktionsprozesses. Ist zum Beispiel die Baumstruktur durch die Vorgabe einer absteigenden Folge von Zahlen $N_t \leq N$ ($t = T, \dots, 1$) stufenweise festgelegt, kann mit der beschriebenen Methode eine nahezu optimale Transformation auf die vorgegebene Struktur erfolgen und gleichzeitig ermöglicht Theorem 6.1 eine Einschätzung der Güte des Szenariobaums. Zusätzlich erlaubt die Abschätzung (6.7) quantitative Aussagen bzgl. des relativen Fehler zu jedem Zeitpunkt t und ermöglicht darüber hinaus eine gezielte Einflussnahme auf die Struktur der Bäume.

Ist die Struktur des Szenariobaums frei, erlaubt der folgende flexible Algorithmus die Konstruktion von Szenariobäumen, die einer gewissen Genauigkeit in Bezug auf den L_r -Abstand genügen.

Algorithmus 6.2 (*Rückwärtsalgorithmus Baumkonstruktion*)

Es seien N Szenarien ξ^i mit Wahrscheinlichkeiten p_i , $i = 1, \dots, N$, und festem Anfangsknoten $\xi_1^* \in \mathbb{R}^d$ gegeben. Weiterhin seien $r \geq 1$ und Stufengenauigkeiten ε_t , $t = 2, \dots, T$, so gewählt, dass $\sum_{t=2}^T \varepsilon_t \leq \varepsilon$.

Schritt [0]: Setze $\bar{\xi}^{T+1} := \xi_f$ und $I_{T+1} = \{1, \dots, N\}$.
Bestimme eine Indexmenge $I_T \subseteq I_{T+1}$ und einen stochastischen Prozess $\bar{\xi}^T$ mit $|I_T|$ Szenarien, so dass $\|\bar{\xi}^{T+1} - \bar{\xi}^T\|_r \leq \varepsilon_T$.

Schritt [t]: Bestimme eine Indexmenge $I_{T-t} \subseteq I_{T-t+1}$ und einen stochastischen Prozess $\bar{\xi}^{T-t}$ mit $|I_{T-t}|$ Szenarien, so dass $\|\bar{\xi}^{T-t+1} - \bar{\xi}^{T-t}\|_{r, T-t} \leq \varepsilon_{T-t}$.

Schritt [T-1]: Konstruiere einen stochastischen Prozess ξ_{tr} mit $|I_T|$ Szenarien ξ_{tr}^j , $j \in I_T$, so dass $\xi_{tr,t}^j := \xi_t^{\alpha_t(j)}$, $t = 1, \dots, T$, wobei $\alpha_t(\cdot)$ wie in (6.6) definiert.

Eine Veranschaulichung des Algorithmus ist in Abbildung 6.2 zu sehen. Das erste Bild zeigt den Ausgangsprozess ξ_f . Das zweite Bild zeigt die Situation nach Schritt 0, welcher einer gewöhnlichen Szenarioreduktion entspricht. Die folgenden drei Bilder zeigen die Resultate der Schritte 1–3. Das letzte Bild zeigt den in Schritt 4 bestimmten Szenariobaum ξ_{tr} .

Proposition 6.3 *Es sei ξ_f ein stochastischer Prozess mit festem Anfangsknoten ξ_1^* , Szenarien ξ^i und Wahrscheinlichkeiten p_i , $i = 1, \dots, N$. Mit ξ_{tr} sei der mit Algorithmus 6.2 konstruierte Prozess bezeichnet. Dann gilt*

$$\|\xi_f - \xi_{tr}\|_r \leq \sum_{t=2}^T \varepsilon_t \leq \varepsilon.$$

Beweis: Die Aussage ist eine direkte Konsequenz der Abschätzung (6.7) in Theorem 6.1, welche die Form

$$\|\xi_f - \xi_{tr}\|_r \leq \sum_{t=2}^T \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}$$

annimmt. □

Wird Algorithmus 6.2 in der Praxis verwendet, werden zunächst ein geeignetes $r \geq 1$ und Genauigkeiten ε_t für $t = 2, \dots, T$ gewählt. Es ist zweckmäßig, r in Abstimmung mit den Eigenschaften des Originalprozess ξ und der geforderten Approximation der Lösungsszenarien, ausgedrückt in der Norm $\|\cdot\|_{r'}$, zu wählen (siehe Abschnitt 4.1.1). Die Wahl der Genauigkeiten ε_t ist dagegen weitgehend offen. Allerdings wirken sich unterschiedliche Genauigkeiten zum Zeitpunkt t in folgender Weise auf die Struktur des konstruierten Baums aus: Die Anzahl der Verzweigungen zum Zeitpunkt t wächst, wenn auch der Wert ε_t wächst. Der Wert $\varepsilon_t = 0$ führt dagegen zu keiner Verzweigung der Szenarien zum Zeitpunkt t . Einige numerische Erfahrungen mit der Steuerung der Genauigkeiten sind in Kapitel 7 zu finden.

6.1.4 Der Vorwärtsalgorithmus

Die Strategie des Vorwärtsalgorithmus zur Konstruktion von Szenariobäumen basiert, genau wie die Rückwärtsvariante, auf einer rekursiven Szenarioreduktion, wird im Gegensatz zu dieser aber auf wachsende Zeiträume $\{1, \dots, t\}$, d.h. auf Zeitintervalle mit steigendem Zeitparameter t angewendet. Die Vorwärtsmethode generiert ausgehend vom Szenariofächer individueller Szenarien ξ_f für $t = 1, \dots, T$ rekursiv stochastische Prozesse

$$\hat{\xi}^t \quad \text{mit Szenarien} \quad \{\hat{\xi}^{t,i}\}_{i \in I} \quad \text{und Wahrscheinlichkeiten} \quad \{p_i\}_{i \in I}$$

sowie Partitionen von I der Form

$$\mathcal{C}_t := \{C_t^1, \dots, C_t^{K_t}\} \quad (K_t \in \mathbb{N}),$$

wobei $I := \{1, \dots, N\}$. Die Partitionen erfüllen für jedes t die Eigenschaften

$$C_t^k \cap C_t^{k'} = \emptyset \quad (k \neq k') \quad \text{und} \quad \bigcup_{k=1}^{K_t} C_t^k = I. \quad (6.10)$$

Die Elemente einer Partition \mathcal{C}_t werden auch Szenariocluster genannt. Sie vereinen Szenarien ähnlichen Verlaufs bis zum Zeitpunkt t . Zur Bestimmung von $\xi_{tr} := \hat{\xi}^T$ erfolgt die Initialisierung der Prozedur mit

$$\hat{\xi}^1 := \xi_f \quad \text{und} \quad \mathcal{C}_1 := \{I\}.$$

Im Schritt t (mit $t > 1$) wird jedes Cluster C_{t-1}^k , d.h. jede Szenarieteilmenge der Form $\{\hat{\xi}^{t-1,i}\}_{i \in C_{t-1}^k}$ separat einer Szenarioreduktion unterzogen, wobei wieder die Seminorm $\|\cdot\|_{r,t}$ verwendet wird. Das Resultat sind disjunkte Indexmengen I_t^k verbleibender Szenarien bzw. J_t^k gestrichener Szenarien, wobei

$$I_t^k \cup J_t^k = C_{t-1}^k \quad (k = 1, \dots, K_{t-1})$$

gilt. Eine partielle Zuordnung von Elementen aus J_t^k zu Elementen in I_t^k erfolgt hier über eine Vorschrift $i_t^k(j)$, definiert durch

$$i_t^k(j) \in \arg \min_{i \in I_t^k} \|\hat{\xi}^{t-1,i} - \hat{\xi}^{t-1,j}\|_t^r \quad (j \in J_t^k) \quad (6.11)$$

gemäß der optimalen Neuverteilungsregel (5.5). Um die Szenarien des Prozesses $\hat{\xi}^t$ zu definieren, werden die partiellen Vorschriften $i_t^k(j)$, ähnlich wie beim Rückwärtsansatz, zu gesamtheitlichen Stufenzuweisungen $\alpha_t : I \rightarrow I$ erweitert. Es sei $\alpha_1(j) \equiv i^*$ ($i^* \in I$ ist hier frei wählbar aufgrund des gemeinsamen Anfangsknotens aller Szenarien) und

$$\alpha_t(j) := \begin{cases} i_t^k(j) & , j \in J_t^k \text{ für ein } k \in \{1, \dots, K_{t-1}\}, \\ j & , \text{sonst,} \end{cases} \quad (6.12)$$

für $t = 2, \dots, T$. Die stochastischen Prozesse $\hat{\xi}^t$ lassen sich somit über die Szenarien $\hat{\xi}^{t,i}$ mit den Komponenten

$$\hat{\xi}_\tau^{t,i} = \begin{cases} \xi_\tau^{\alpha_\tau(i)} & , \tau \leq t, \\ \xi_\tau^i & , \text{sonst,} \end{cases} \quad (\tau = 1, \dots, T) \quad (6.13)$$

versehen mit den Wahrscheinlichkeiten p_i ($i \in I$) definieren. Eine Illustration ist in Abbildung 6.3 zu sehen, wobei die Einzelbilder die Prozesse $\hat{\xi}^1, \dots, \hat{\xi}^5$ sowie $\xi_{\text{tr}} = \hat{\xi}^5$ zeigen. Die Partition \mathcal{C}_t zum Zeitpunkt t wird durch die inverse Abbildung zu α_t gebildet. Es sei

$$\mathcal{C}_t := \{\alpha_t^{-1}(i) \mid i \in I_t^k, k = 1, \dots, K_{t-1}\}, \quad (6.14)$$

das heißt, jedes der Elemente der Indexmengen I_t^k definiert ein neues Cluster, welche zusammen die neue Partition \mathcal{C}_t bilden. Die Partition \mathcal{C}_t stellt eine Verfeinerung der vorherigen Partition \mathcal{C}_{t-1} dar. Für die Mengen $I_t, \bar{I}_{t,i}$ sowie Wahrscheinlichkeiten π_t^i in der Notation des Rückwärtsalgorithmus in Abschnitt 6.1.3 ergeben sich nun folgende Darstellungen. Für die Menge der Realisierungen zum Zeitpunkt t ergibt sich

$$I_t = \bigcup_{k=1}^{K_{t-1}} I_t^k$$

und mit

$$J_{t,i}^k := \{j \in J_t^k : i = i_t^k(j)\} \quad (i \in I_t^k)$$

erhält man

$$\bar{I}_{t,i} := J_{t,i}^k \cup \{i\} \in C_t \quad (i \in I_t^k)$$

sowie

$$\pi_t^i = \sum_{j \in J_{t,i}^k \cup \{i\}} p_j \quad (i \in I_t^k) \quad (6.15)$$

für $k = 1, \dots, K_{t-1}$.

Die Anzahl der Szenarien und die Wahrscheinlichkeiten sowie die Struktur des stochastischen Szenariobaumprozesses $\xi_{\text{tr}} = \hat{\xi}^T$ ergeben sich aus den Parametern der Partition \mathcal{C}_T bzw. aus den Eigenschaften der Abbildungen α_t . Man erhält für jedes $k = 1, \dots, K_T$ ein wohldefiniertes Szenario

$$\xi_{\text{tr}}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^i \right) \text{ mit Wahrscheinlichkeit } \pi_T^i,$$

wobei i derjenige Index ist, der das Cluster C_T^k erzeugt. Für diesen Index gilt $i \in C_T^k$ sowie $\alpha_T(j) \equiv i$ für alle Indizes $j \in C_T^k$ und damit $\alpha_T(i) = i$. Als Erzeuger von C_T^k liegt i auch in einer der Mengen I_T^l für ein Index $l \in \{1, \dots, K_{T-1}\}$, wodurch die Wahrscheinlichkeit π_T^i in (6.15) wohldefiniert ist. Der Verzweigungsgrad des Szenarios k zum Zeitpunkt t kann durch $1 + |J_{t,i}^k|$ mit der zugehörige Indexmenge $J_{t,i}^k$ berechnet werden.

Es lässt sich ein zu Theorem 6.1 analoges Resultat beweisen. Für einen mit der Vorwärtsmethode konstruierten stochastischen Prozess ξ_{tr} gilt folgende Aussage in Bezug auf den L_r -Abstand.

Theorem 6.4 *Es sei ξ_f ein stochastischer Prozess mit fixiertem Anfangsknoten ξ_1^* , Szenarien ξ^i und Wahrscheinlichkeiten p_i , $i = 1, \dots, N$. Weiter sei ξ_{tr} ein stochastischer Prozess ξ_{tr} mit Szenarien*

$$\xi_{\text{tr}}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)} \right) \text{ und Wahrscheinlichkeiten } \pi_T^{\alpha_T(i)},$$

wobei $k = 1, \dots, K_T$ und $i \in C_T^k$. Dann gilt die Gleichung

$$\|\xi_f - \xi_{\text{tr}}\|_r = \left(\sum_{t=2}^T \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} \|\xi_t^i - \xi_t^j\|^r \right)^{\frac{1}{r}}. \quad (6.16)$$

Beweis: Für den L_r -Abstand zwischen ξ_f und ξ_{tr} gilt nach Konstruktion

$$\begin{aligned} \|\xi_f - \xi_{\text{tr}}\|_r^r &= \sum_{t=1}^T \sum_{i=1}^N p_i \|\xi_t^i - \xi_t^{\alpha_t(i)}\|^r \\ &= \sum_{t=2}^T \sum_{k=1}^{K_{t-1}} \sum_{j \in C_{t-1}^k} p_j \|\xi_t^j - \xi_t^{\alpha_t(j)}\|^r. \end{aligned} \quad (6.17)$$

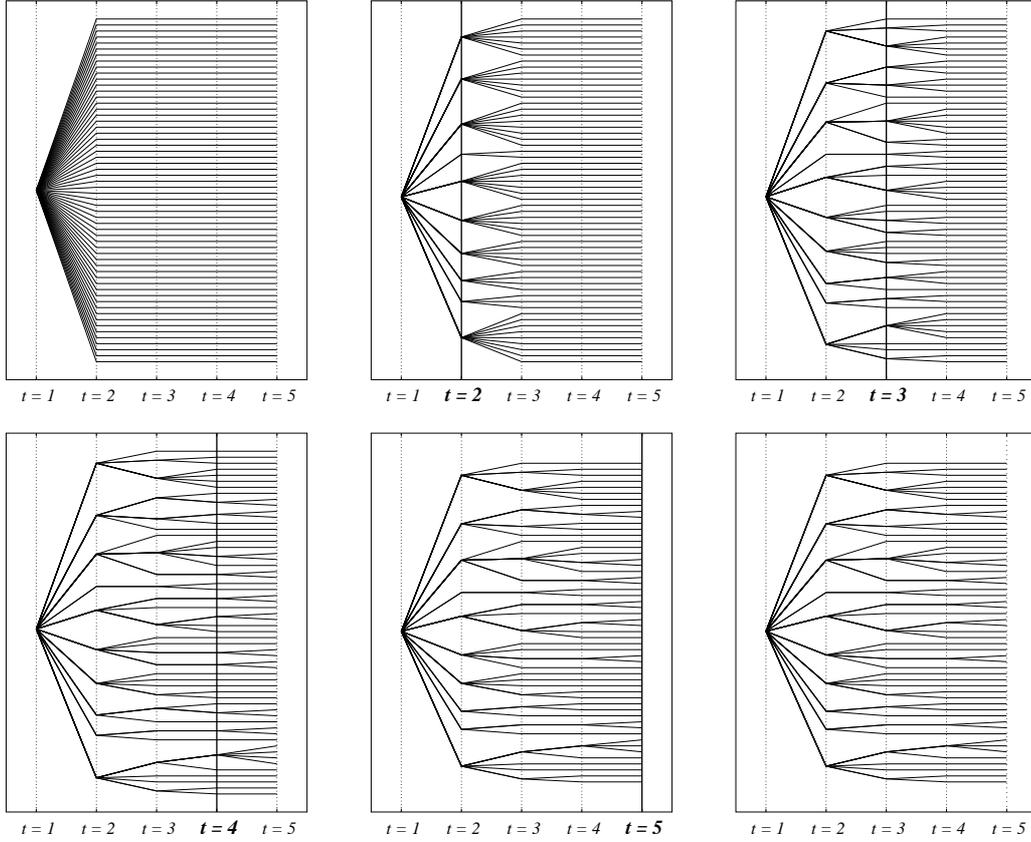


Abbildung 6.3: Illustration der Vorwärtskonstruktion von Szenariobäumen an einem Beispiel mit $T = 5$ Zeitperioden, ausgehend von einem Szenariofächer individueller Szenarien.

Mit der Definition (6.12) für α_t lässt sich die Identität fortsetzen mit

$$\|\xi_f - \xi_{tr}\|_r^r = \sum_{t=2}^T \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \|\xi_t^j - \xi_t^{i_t^k(j)}\|_r^r.$$

Aufgrund der Konstruktion der Partitionen \mathcal{C}_t in (6.14) gilt für alle t die Eigenschaft $\alpha_t(i) = \alpha_t(j)$ für Indizes i und j , die einem selben Cluster in \mathcal{C}_t angehören. Da \mathcal{C}_t stets eine Verfeinerung von \mathcal{C}_{t-1} ist, folgt daraus

$$\alpha_{t-\tau}(i) = \alpha_{t-\tau}(j) \quad \text{für alle } i \in I_t^k, j \in J_t^k, \tau = 1, \dots, t-1$$

mit $k \in \{1, \dots, K_{t-1}\}$. Damit gilt wegen der Darstellung der Szenarien $\hat{\xi}^{t-1, i}$ des stochastischen Prozess $\hat{\xi}^t$ in (6.13) bzgl. der Seminorm $\|\cdot\|_t$

$$\|\hat{\xi}^{t-1, i} - \hat{\xi}^{t-1, j}\|_t^r = \|\xi_t^i - \xi_t^j\|_t^r \quad \text{für alle } i \in I_t^k, j \in J_t^k.$$

Das Verwenden der letzten Eigenschaft in (6.11) liefert schließlich

$$\|\xi_f - \xi_{\text{tr}}\|_r^r = \sum_{t=2}^T \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} \|\xi_t^i - \xi_t^j\|_r^r,$$

womit durch Ziehen der r -ten Wurzel die Aussage bewiesen ist. \square

Die Aussage über den L_r -Fehler in Theorem 6.4 ist etwas schärfer als die Fehlerabschätzung in Theorem 6.1. Das liegt in der Tatsache begründet, dass im Fall des Vorwärtsalgorithmus der Beweis ohne die Verwendung der Dreiecksungleichung für die L_r -Norm $\|\cdot\|_r$ geführt werden kann. Dennoch sind die Resultate beider Theoreme sehr ähnlich. Auch die Vorwärtskonstruktion erlaubt den Fehler des t -ten Konstruktionsschritts zu quantifizieren. Wie im vorherigen Abschnitt kann damit ein flexibler Algorithmus zur Bestimmung verschiedener Szenariobäume, die einer vorgegebenen Genauigkeit genügen, aufgestellt werden.

Algorithmus 6.5 (*Vorwärtsalgorithmus Baumkonstruktion*)

Es seien N Szenarien ξ^i mit Wahrscheinlichkeiten p_i , $i = 1, \dots, N$, und festem Anfangsknoten $\xi_1^* \in \mathbb{R}^d$ gegeben. Weiterhin seien $r \geq 1$ und Stufengenauigkeiten ε_t , $t = 2, \dots, T$, so gewählt, dass $\sum_{t=2}^T \varepsilon_t \leq \varepsilon$.

Schritt [1]: Setze $\hat{\xi}^1 := \xi_f$ und $\mathcal{C}_1 = \{\{1, \dots, N\}\}$.

Schritt [t]: Es sei $\mathcal{C}_{t-1} = \{C_{t-1}^1, \dots, C_{t-1}^{K_{t-1}}\}$:
Bestimme disjunkte Indextmengen I_t^k und J_t^k , so dass $I_t^k \cup J_t^k = C_{t-1}^k$, Abbildungen $\alpha_t(\cdot)$ gemäß (6.12) sowie einen stochastischen Prozess $\hat{\xi}^t$ mit N Szenarien $\hat{\xi}^{t,i}$ und Wahrscheinlichkeiten p_i gemäß (6.13), so dass $\|\hat{\xi}^{t-1} - \hat{\xi}^t\|_{r,t} \leq \varepsilon_t$.
Setze $\mathcal{C}_t := \{\alpha_t^{-1}(i) \mid i \in I_t^k, k = 1, \dots, K_{t-1}\}$.

Schritt [T+1]: Es sei $\mathcal{C}_T = \{C_T^1, \dots, C_T^{K_T}\}$:
Konstruiere einen stochastischen Prozess ξ_{tr} mit K_T Szenarien ξ_{tr}^k , so dass $\xi_{\text{tr},t}^k := \xi_t^{\alpha_t(i)}$, $i \in C_T^k$, $k = 1, \dots, K_T$ und $t = 1, \dots, T$.

Zur Veranschaulichung des Algorithmus 6.5 zeigt das erste Bild in Abbildung 6.3 den Eingangsprozess ξ_f . Die folgenden Bilder entsprechen jeweils der Situation nach den Schritten 2–5, d.h. sie zeigen die nacheinander konstruierten Prozesse $\hat{\xi}^t$ für $t = 2, \dots, 5$. Das letzte Bild ist dem Schritt 6 zugeordnet und zeigt den mit der Vorwärtsmethode generierten Szenariobaum ξ_{tr} .

Proposition 6.6 *Es sei ξ_f ein stochastischer Prozess mit festem Anfangsknoten ξ_1^* , Szenarien ξ^i und Wahrscheinlichkeiten p_i , $i = 1, \dots, N$. Mit ξ_{tr} sei der durch Algorithmus 6.5 konstruierte Prozess bezeichnet. Dann gilt*

$$\|\xi_f - \xi_{tr}\|_r \leq \left(\sum_{t=2}^T \varepsilon_t^r \right)^{\frac{1}{r}} \leq \sum_{t=2}^T \varepsilon_t \leq \varepsilon. \quad (6.18)$$

Beweis: Aufgrund der Darstellung (6.13) für den stochastischen Prozess $\hat{\xi}^t$ gilt

$$\begin{aligned} \|\hat{\xi}^{t-1} - \hat{\xi}^t\|_{r,t}^r &= \sum_{i=1}^N p_i \|\hat{\xi}^{t-1,i} - \hat{\xi}^{t,i}\|_t^r \\ &= \sum_{k=1}^{K_{t-1}} \sum_{j \in C_{t-1}^k} p_j \|\xi_t^j - \xi_t^{\alpha_t(j)}\|^r. \end{aligned}$$

Damit ist die Abschätzung (6.18) eine direkte Konsequenz der Gleichung (6.17) im Beweis von Theorem 6.4, welche die Form

$$\|\xi_f - \xi_{tr}\|_r^r = \sum_{t=2}^T \|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t}^r \leq \sum_{t=2}^T \varepsilon_t^r$$

annimmt. □

Wird der Algorithmus 6.5 in der Praxis verwendet, so sollte die Wahl des Parameters $r \geq 1$ nach den gleichen Kriterien wie in Algorithmus 6.2 erfolgen, d.h. in Abhängigkeit des verwendeten stochastischen Modells sowie der Eigenschaften des Originalprozess ξ (hier sei noch einmal auf die Diskussion in Abschnitt 4.1 von Kapitel 4 verwiesen). Die Auswahl der Stufengenauigkeiten ε_t unterscheidet sich dagegen etwas von der Rückwärtsvariante der Baumkonstruktion. Hier ist es zweckmäßig so vorzugehen, dass die Werte ε_t für $t = 2, \dots, T$ eine nichtfallende Folge von Toleranzen bilden. Ein sehr kleiner Wert für ε_t würde bewirken, dass der relative Grad der Verzweigung zum Zeitpunkt $t - 1$ sehr groß wird. Einige Erfahrungen mit der Steuerung der Genauigkeiten sind im abschließenden Kapitel 7 für verschiedene Anwendungen zu finden.

6.2 Konvergenz

Nach der Beschreibung der Konstruktionsalgorithmen geht es in diesem Abschnitt nun um die generelle Frage nach der Konvergenz. Das Stabilitätsresultat in Theorem 4.1 und das Beispiel 4.1.4 zeigen, dass L_r -Nähe zwischen Eingangsprozessen allein nicht hinreichend für die Nähe der Optimalwerte ist.

Im Folgenden soll daher das Verhalten der Algorithmen hinsichtlich der Informationsstruktur genauer untersucht werden. Insbesondere die Konvergenz des Filtrationsabstands steht im Mittelpunkt dieses Abschnitts.

Um die Darstellung des Filtrationsabstands einfach zu halten, wird die Existenz von Lösungen des stochastischen Optimierungsproblems im Weiteren vorausgesetzt. Im Fall nichtleerer Lösungsmengen $S(\xi)$ bzw. $S(\tilde{\xi})$ ergibt sich für den relevanten Filtrationsabstand zwischen zwei Prozessen ξ und $\tilde{\xi}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ nach (4.19)

$$D_f(\xi, \tilde{\xi}) = \inf_{\substack{x \in S(\xi) \\ \tilde{x} \in S(\tilde{\xi})}} \sum_{t=2}^{T-1} \max \left\{ \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'} \right\}. \quad (6.19)$$

Wie in Kapitel 4 gezeigt wurde, ist für die Existenz von Lösungen beispielsweise die Gültigkeit der Voraussetzungen von Proposition 4.3 (vgl. Abschnitt 4.1.5) hinreichend.

Ein grundlegendes Resultat betrifft die Konvergenzeigenschaft des Filtrationsabstands (6.19) für Folgen adaptierter Prozesse, die in L_r konvergieren:

Proposition 6.7 *Es seien (A1), (A2) und (A3) für $1 \leq r' < +\infty$ erfüllt und $S(\xi)$ sei nichtleer. Des Weiteren sei $(\hat{\xi}^{(n)})$ eine Folge von zu ξ adaptierten Prozessen, die gegen ξ in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ konvergiert, so dass die Folge der σ -Algebren $\mathcal{F}_t(\hat{\xi}^{(n)})$ nichtfallend bzgl. $n \in \mathbb{N}$ ist. Darüber hinaus sei $S(\hat{\xi}^{(n)})$ für hinreichend große $n \in \mathbb{N}$ ebenfalls nichtleer. Dann gilt*

$$\lim_{n \rightarrow \infty} D_f(\xi, \hat{\xi}^{(n)}) = \lim_{n \rightarrow \infty} \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\hat{\xi}^{(n)})]\|_{r'} = 0.$$

Beweis: Zu jedem $t = 1, \dots, T$ bezeichne $\hat{\mathcal{F}}_t$ die kleinste σ -Algebra, die die Mengensysteme $\mathcal{F}_t(\hat{\xi}^{(n)})$ für alle $n \in \mathbb{N}$ enthält, d.h. es sei

$$\hat{\mathcal{F}}_t := \sigma \left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_t(\hat{\xi}^{(n)}) \right).$$

Da die Folge $(\hat{\xi}^{(n)})$ zu ξ adaptiert ist, gilt

$$\mathcal{F}_t(\hat{\xi}^{(n)}) \subseteq \mathcal{F}_t(\xi) \quad \text{für alle } n \in \mathbb{N}, t = 1, \dots, T,$$

und folglich

$$\hat{\mathcal{F}}_t \subseteq \mathcal{F}_t(\xi) \quad \text{für alle } t = 1, \dots, T.$$

Insbesondere gilt $\hat{x}_t^{(n)} = \mathbb{E}[\hat{x}_t^{(n)} | \mathcal{F}_t(\xi)]$ für alle $t = 1, \dots, T$ und jede Lösung $\hat{x}^{(n)} \in S(\hat{\xi}^{(n)})$, womit sich der Filtrationsabstand vereinfacht zu

$$D_f(\xi, \hat{\xi}^{(n)}) = \inf_{x \in S(\xi)} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\hat{\xi}^{(n)})]\|_{r'}.$$

Aufgrund der Annahme, dass die σ -Algebren $\mathcal{F}_t(\hat{\xi}^{(n)})$ nichtfallend bzgl. $n \in \mathbb{N}$ sind, sichert die Konvergenz von $(\hat{\xi}^{(n)})$ gegen ξ in L_r , dass

$$\xi_t \in L_r(\Omega, \hat{\mathcal{F}}_t, \mathbb{P}) \quad \text{für jedes } t = 1, \dots, T.$$

Für ein beliebig gewähltes $x \in S(\xi)$ liefern klassische Konvergenzsätze für bedingte Erwartungen (siehe zum Beispiel [25])

$$\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\hat{\xi}^{(n)})]\|_{r'} \longrightarrow 0 \quad (n \rightarrow \infty)$$

für alle $t = 1, \dots, T$, woraus die die Behauptung folgt. \square

Die Bedeutung von Proposition 6.7 wird im folgenden Abschnitt deutlich. Dort wird gezeigt, wie beliebige Ausgangsprozesse, unter Verwendung von Diskretisierungstechniken, durch endliche und zugleich adaptierte Prozesse approximiert werden können.

6.2.1 Konvergenz von Diskretisierungen

Es bezeichne $\xi = (\xi_1, \dots, \xi_T)$ den originalen Eingangsprozess mit Komponenten $\xi_t \in \Xi_t$, wobei $\Xi_t \subseteq \mathbb{R}^d$ abgeschlossen. Der Prozess ξ sei auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben. Ohne Beschränkung der Allgemeinheit kann dieser Wahrscheinlichkeitsraum mit dem Zustandsraum

$$(\Xi, \mathcal{B}(\Xi), P_\xi)$$

identifiziert werden, wobei $\Xi = \times_{t=1}^T \Xi_t$ eine abgeschlossene Teilmenge in \mathbb{R}^{Td} ist, $\mathcal{B}(\Xi)$ die Menge aller Borel-Mengen in Ξ bezeichnet und P_ξ die Verteilung von ξ ist. Folglich gilt

$$\mathbb{P}(\{\omega \in \Omega : \xi(\omega) \in B\}) = P_\xi(B) = \int_B P_\xi(d\xi), \quad \text{für alle } B \in \mathcal{B}(\Xi).$$

Die durch Realisierungen $\xi^t = (\xi_1, \dots, \xi_t)$ zur Zeitstufe t erzeugten σ -Algebren $\mathcal{F}_t(\xi) = \sigma(\xi^t)$ besitzen damit die Form

$$\mathcal{F}_t(\xi) = \left\{ \xi^{-1}(B_1 \times \dots \times B_t \times \Xi_{t+1} \times \dots \times \Xi_T) : B_t \in \mathcal{B}(\Xi_t) \right\}, \quad (6.20)$$

wobei $\mathcal{B}(\Xi_t)$ die Borel-Mengen in Ξ_t bezeichnen.

Zerlegung des Zustandsraums

Ziel ist es, den stochastischen Prozess ξ durch eine Folge geeigneter diskreter Prozesse zu approximieren. Die Grundlage dafür ist eine endliche Zerlegung des Zustandsraums Ξ . Dazu werden zunächst Folgen endlicher Zerlegungen

$$\mathcal{D}_t^{(n)} \subset \mathcal{B}(\Xi_t), \quad n \in \mathbb{N},$$

in Ξ_t betrachtet, für die folgende Zulässigkeitsbedingungen erfüllt seien:

- (E1) Die Elemente von $\mathcal{D}_t^{(n)}$ sind für alle $n \in \mathbb{N}$ jeweils disjunkt,
- (E2) $\mathcal{D}_t^{(n)}$ ist endlich und $\cup_{D_t \in \mathcal{D}_t^{(n)}} D_t = \Xi_t$ für alle $n \in \mathbb{N}$,
- (E3) Für $\delta_{t,n} := \sup_{D_t \in \mathcal{D}_t^{(n)}} \left\{ \|\xi_t - \tilde{\xi}_t\| : \xi_t, \tilde{\xi}_t \in D_t; \|\xi_t\|, \|\tilde{\xi}_t\| \leq n \right\}$ gilt
 $\lim_{n \rightarrow \infty} \delta_{t,n} = 0$ für alle $n \in \mathbb{N}$.

Die Eigenschaften (E1) und (E2) sichern, dass durch jede Zerlegung der gesamte Träger von ξ , d.h. der Zustandsraum Ξ_t , disjunkt überdeckt wird. Insbesondere lässt sich jedes Element in Ξ_t eindeutig einer Menge in $\mathcal{D}_t^{(n)}$ zuordnen. Eigenschaft (E3) besagt darüber hinaus, dass die Zerlegungen gleichmäßig für $n \rightarrow \infty$, innerhalb einer mit n wachsenden Kugel, beliebig fein werden. Betrachtet man nun das kartesische Produkt der Zerlegungen $\mathcal{D}_t^{(n)}$ für $t = 1, \dots, T$, so gelangt man auf natürliche Weise zu einer Folge von Zerlegungen

$$\mathcal{D}^{(n)} := \left\{ D_1 \times \dots \times D_T : D_t \in \mathcal{D}_t^{(n)} \right\}, \quad n \in \mathbb{N}, \quad (6.21)$$

im gesamten Zustandsraum Ξ .

Diskretisierung des stochastischen Prozesses

Mit Hilfe der Zerlegungen $\mathcal{D}^{(n)}$ von Ξ soll im nächsten Schritt eine Folge von approximativen stochastischen Prozessen $\hat{\xi}^{(n)} = (\hat{\xi}_1^{(n)}, \dots, \hat{\xi}_T^{(n)})$ definiert werden. Dazu bezeichne

$$\mathbb{1}_B(x) := \begin{cases} 1, & \text{falls } x \in B \\ 0, & \text{sonst} \end{cases}, \quad B \in \mathcal{B}(\Xi),$$

die charakteristische Funktion auf Ξ zu B . Weiterhin werde für jedes n und für jede Menge $D_1 \times \dots \times D_T \in \mathcal{D}^{(n)}$ ein für jedes t nichtantizipatives, d.h. nur von den Mengen D_1, \dots, D_t abhängendes Element

$$\hat{\xi}_t^{D_1, \dots, D_t, n} \in D_t \quad \text{mit} \quad \|\hat{\xi}_t^{D_1, \dots, D_t, n}\| \leq C \cdot n \quad (6.22)$$

ausgewählt. Die Normierungsbedingung mit einer beliebigen aber festen Konstanten C sorgt lediglich dafür, dass die Norm der gewählten Elemente maximal linear mit dem Index n wächst und stellt keine wesentliche Einschränkung dar. Für jedes $D_1 \times \dots \times D_T \in \mathcal{D}^{(n)}$ ergibt sich so ein wohl bestimmtes Szenario

$$\hat{\xi}_{D_1 \times \dots \times D_T}^{(n)} := \left(\hat{\xi}_1^{D_1, n}, \dots, \hat{\xi}_T^{D_1, \dots, D_T, n} \right).$$

Mit dieser Konstruktion erhält man in Abhängigkeit der Zerlegungen $\mathcal{D}^{(n)}$ zu ξ eine Folge approximativer Prozesse $(\hat{\xi}^{(n)})_{n \in \mathbb{N}}$ mit

$$\hat{\xi}^{(n)}(\xi) := \sum_{D_1 \times \dots \times D_T \in \mathcal{D}^{(n)}} \hat{\xi}_{D_1 \times \dots \times D_T}^{(n)} \cdot \mathbb{1}_{D_1 \times \dots \times D_T}(\xi). \quad (6.23)$$

Die stochastischen Prozesse $\hat{\xi}^{(n)}$ definieren einfache Treppenfunktionen im Zustandsraum $(\Xi, \mathcal{B}(\Xi), P_\xi)$ und es gilt folgende Aussage.

Proposition 6.8 *Es seien (E1) – (E3) erfüllt und $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$. Dann gilt für die durch (6.23) definierten Prozesse $\hat{\xi}^{(n)}$:*

- (i) $\mathcal{F}_t(\hat{\xi}^{(n)}) \subseteq \mathcal{F}_t(\xi)$ für alle $n \in \mathbb{N}$,
- (ii) $\lim_{n \rightarrow \infty} \|\xi - \hat{\xi}^{(n)}\|_r = 0$.

Beweis: Zunächst wird die Adaptiertheit der Folge $(\hat{\xi}^{(n)})_{n \in \mathbb{N}}$ in Bezug auf ξ gezeigt. Es gilt nach (6.20)

$$\mathcal{F}_t(\xi) = \left\{ \xi^{-1} \left(B_1 \times \cdots \times B_t \times \Xi_{t+1} \times \cdots \times \Xi_T \right) : B_t \in \mathcal{B}(\Xi_t) \right\}.$$

Nach Konstruktion von $\hat{\xi}^{(n)}$ definiert

$$\left\{ \xi^{-1} \left(D_1 \times \cdots \times D_t \times \Xi_{t+1} \times \cdots \times \Xi_T \right) : D_t \in \mathcal{D}_t^{(n)} \right\}$$

eine Erzeugermenge für die σ -Algebra $\mathcal{F}_t(\hat{\xi}^{(n)})$. Folglich gilt $\mathcal{F}_t(\hat{\xi}^{(n)}) \subseteq \mathcal{F}_t(\xi)$ für alle $n \in \mathbb{N}$ und damit Aussage (i). Um die L_r -Konvergenz zeigen zu können, bezeichne

$$B_{\gamma_n}(0) := \left\{ x \in \mathbb{R}^{Td} : \|x_t\| \leq \gamma_n, t = 1, \dots, T \right\}$$

die abgeschlossene Kugel in Ξ mit Radius

$$\gamma_n := n - \max_{t=1, \dots, T} \delta_{t,n}.$$

Dann gilt mit Eigenschaften (E1) und (E2)

$$\begin{aligned} \|\xi - \hat{\xi}^{(n)}\|_r^r &= \int_{\Omega} \|\xi - \hat{\xi}^{(n)}\|_r^r \mathbb{P}(d\omega) = \int_{\Xi} \|\xi - \hat{\xi}^{(n)}\|_r^r P_\xi(d\xi) \\ &= \sum_{D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}} \int_{D_1 \times \cdots \times D_T} \sum_{t=1}^T \|\xi_t - \hat{\xi}_t^{D_1, \dots, D_t, n}\|_r^r P_\xi(d\xi). \end{aligned}$$

Aufteilung der Integrationsbereiche in Anteile der Kugel $B_{\gamma_n}(0)$ und ihrem Komplement ergibt unter Ausnutzung der Eigenschaft (E3) sowie der Wachstumsbedingung (6.22) die Abschätzung

$$\begin{aligned} \|\xi - \hat{\xi}^{(n)}\|_r^r &\leq \sum_{t=1}^T \delta_{t,n}^r + \int_{\Xi \setminus B_{\gamma_n}(0)} \sum_{t=1}^T (\|\xi_t\| + C \cdot n)^r P_\xi(d\xi) \\ &\leq \hat{C} \left(\max_{t=1, \dots, T} \delta_{t,n}^r + \int_{\Xi \setminus B_{\gamma_n}(0)} \|\xi\|_r^r P_\xi(d\xi) \right), \end{aligned}$$

wobei $\hat{C} > 0$ eine von n unabhängige Konstante ist. Da beide Summanden der letzten Ungleichung gegen 0 konvergieren, folgt auch Behauptung (ii). \square

Die Aussagen der Propositionen 6.7 und 6.8 liefern folgendes Resultat. Wenn die Diskretisierung des Zustandsraums so gewählt wird, dass jeweils $\mathcal{D}_t^{(n+1)}$ eine Verfeinerung von $\mathcal{D}_t^{(n)}$ darstellt, so erhält man

$$v(\hat{\xi}^{(n)}) \rightarrow v(\xi) \quad (n \rightarrow \infty). \quad (6.24)$$

Der Optimalwert des mehrstufigen stochastischen Optimierungsproblems mit Eingangsprozess ξ ist also Grenzwert der Folge von Optimalwerten bzgl. der diskreten Approximationen $\hat{\xi}^{(n)}$, da die Diskretisierungen sowohl bzgl. L_r als auch D_t konvergieren.

Das Konvergenzresultat für Diskretisierungen ist für diese Arbeit von nicht geringer theoretischer Bedeutung. Es bildet die Grundlage für weiterer Betrachtungen in Abschnitt 6.2.2. Würde man Diskretisierungen direkt zur Approximation stochastischer Optimierungsprobleme nutzen wollen, steht man vor dem Problem, alle erforderlichen Wahrscheinlichkeiten

$$P_\xi(D_1 \times \cdots \times D_T) \quad \text{für alle } D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$$

bestimmen zu müssen, was sich im Allgemeinen als schwierig erweist. Dennoch führt in besonderen Fällen, wenn der stochastische Prozess ξ durch unabhängige \mathbb{R}^{d_t} -wertige Zufallsvariablen z_t mit Wahrscheinlichkeitsverteilungen P_t ($t = 2, \dots, T$) erzeugt wird, die Kenntnis von $P_t(D_t)$ für alle $D_t \in \mathcal{D}_t^{(n)}$ und Partitionen $\mathcal{D}_t^{(n)}$ des Trägers von z_t , direkt zur Wahrscheinlichkeitsverteilung von $\hat{\xi}^{(n)}$ (siehe [65] für mehr Details).

6.2.2 Konvergenz von Schätzungen

Eine praktikable Möglichkeit zur Bestimmung von diskreten Approximationen, wie sie bei der Diskretisierung auftreten, bieten Schätzungen über empirische Verteilungen. Als Motivation dient folgende Aussage:

Proposition 6.9 *Es sei P eine Wahrscheinlichkeitsverteilung auf einer endlichen Trägermenge $\Xi = \{\xi^1, \dots, \xi^N\} \subseteq \mathbb{R}^{T^d}$ mit positiven Wahrscheinlichkeiten $p_i := P(\{\xi^i\})$. Dann existieren für jede Folge von Verteilungen $(P^{(n)})_{n \in \mathbb{N}}$ in $\mathcal{P}(\Xi)$ mit Wahrscheinlichkeiten $p_i^{(n)} := P^{(n)}(\{\xi^i\})$ unter der Bedingung*

$$P^{(n)} \rightarrow P \quad \text{bzw.} \quad p_i^{(n)} \rightarrow p_i \quad (n \rightarrow \infty), \quad i = 1, \dots, N,$$

Zufallsvariablen ξ und $(\xi^{(n)})_{n \in \mathbb{N}}$ auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so dass für $1 \leq r < \infty$ und $1 \leq r' < \infty$

- (i) $\|\xi - \xi^{(n)}\|_r \rightarrow 0 \quad (n \rightarrow \infty)$ und
- (ii) $\sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})]\|_{r'} \rightarrow 0 \quad (n \rightarrow \infty),$

wobei $\mathcal{L}(\xi) = P$ sowie $\mathcal{L}(\xi^{(n)}) = P^{(n)}$ für alle $n \in \mathbb{N}$, d.h. die Verteilungen von ξ und $\xi^{(n)}$ sind gerade P bzw. $P^{(n)}$. Die Menge \mathcal{B}_∞ ist dabei durch

$$\mathcal{B}_\infty := \{x = (x_1, \dots, x_T) \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : \|x_t(\omega)\| \leq 1 \text{ f.ü.}\} \quad (6.25)$$

gegeben und bezeichnet die Menge aller fast überall beschränkter Funktionen.

Beweis: Die Konvergenz der Folge von Verteilungen $(P^{(n)})_{n \in \mathbb{N}}$ gegen die Verteilung P impliziert die Existenz von Zufallsvariablen ξ und $\xi^{(n)}$ mit Verteilungen $\mathcal{L}(\xi) = P$ bzw. $\mathcal{L}(\xi^{(n)}) = P^{(n)}$ auf einem (gemeinsamen) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so dass

$$\xi^{(n)} \rightarrow \xi \quad \mathbb{P}\text{-fast sicher}$$

(vgl. [17, Theorem 11.7.1]). Da der Träger Ξ endlich ist, folgt aus der \mathbb{P} -fast sicheren Konvergenz die Konvergenz im r -ten Mittel, d.h.

$$\|\xi - \xi^{(n)}\|_r \rightarrow 0 \quad (n \rightarrow \infty).$$

Es bleibt zu zeigen, dass die Differenzen der bedingten Erwartungen für beschränkte Funktionen gleichmäßig gegen 0 konvergiert. Dazu werden die Partitionen $\{E_{tk}\}_{k \in I_t}$ bzw. $\{E_{tk}^{(n)}\}_{k \in I_t}$ von Ω betrachtet, welche die σ -Algebren $\mathcal{F}_t(\xi)$ und $\mathcal{F}_t(\xi^{(n)})$ erzeugen. Es sei

$$\begin{aligned} E_{tk} &:= \{\omega \in \Omega : (\xi_1, \dots, \xi_t)(\omega) = (\xi_1^k, \dots, \xi_t^k)\}, \quad k \in I_t, \\ E_{tk}^{(n)} &:= \{\omega \in \Omega : (\xi_1^{(n)}, \dots, \xi_t^{(n)})(\omega) = (\xi_1^k, \dots, \xi_t^k)\}, \quad k \in I_t, \end{aligned}$$

wobei $I_t \subseteq \{1, \dots, N\}$ die Indexmengen unterscheidbarer Szenarien zu den Zeitpunkten $t = 2, \dots, T-1$ bezeichnen. Setzt man $p_{tk} := \mathbb{P}(E_{tk})$ und $p_{tk}^{(n)} := \mathbb{P}(E_{tk}^{(n)})$ für $k \in I_t$ sowie darüber hinaus

$$\begin{aligned} \bar{E}_{tk}^{(n)} &:= E_{tk} \cap E_{tk}^{(n)} \quad \text{und} \\ \bar{\Omega}_t^{(n)} &:= \Omega \setminus \bigcup_{k \in I_t} \bar{E}_{tk}^{(n)}, \end{aligned}$$

so gilt für ein beliebiges $x \in \mathcal{B}_\infty$:

$$\begin{aligned} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})]\|_{r'}^{r'} &= \int_{\Omega} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})]\|^{r'} \mathbb{P}(d\omega) \\ &\leq \sum_{k \in I_t} \int_{\bar{E}_{tk}^{(n)}} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})]\|^{r'} \mathbb{P}(d\omega) \\ &\quad + \int_{\bar{\Omega}_t^{(n)}} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{(n)})]\|^{r'} \mathbb{P}(d\omega) \\ &\leq \sum_{k \in I_t} \mathbb{P}(\bar{E}_{tk}^{(n)}) \left\| \frac{\int_{E_{tk}} x_t \mathbb{P}(d\omega)}{p_{tk}} - \frac{\int_{E_{tk}^{(n)}} x_t \mathbb{P}(d\omega)}{p_{tk}^{(n)}} \right\|^{r'} + 2\mathbb{P}(\bar{\Omega}_t^{(n)}) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{k \in I_t} \mathbb{P}(\bar{E}_{tk}^{(n)}) \left(\left| \frac{p_{tk}^{(n)} - p_{tk}}{p_{tk} p_{tk}^{(n)}} \right| + \left| \frac{p_{tk}^{(n)} - p_{tk}}{p_{tk} p_{tk}^{(n)}} \right| \right)^{r'} \\ &\quad + 2 \sum_{k \in I_t} |p_{tk}^{(n)} - p_{tk}|, \end{aligned}$$

wobei im Wesentlichen die fast überall Beschränktheit $\|x_t\| \leq 1$ bzgl. der Vektornorm ausgenutzt wurde. Da $p_{tk}^{(n)}$ für alle $k \in I_t$ gegen p_{tk} konvergiert, folgt aus der Abschätzung die Konvergenz gegen Null für alle $t = 2, \dots, T-1$. Da die letzte Ungleichung unabhängig vom gewählten x gilt, folgt auch die gleichmäßige Konvergenz für alle $x \in \mathcal{B}_\infty$ und damit Behauptung (ii). \square

Unter Verwendung einer von den Lösungen der Optimierungsprobleme unabhängigen Formulierung des Filtrationsabstands kann Proposition 6.9 für Konvergenzaussagen bzgl. Approximationen stochastischer Optimierungsprobleme genutzt werden. Betrachtet man den Filtrationsabstand

$$D_f^\infty(\xi, \tilde{\xi}) := \sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \quad (6.26)$$

wobei \mathcal{B}_∞ definiert wie in Proposition 6.9, so gelangt man unter der Voraussetzung der wesentlichen Beschränktheit von Lösungen (bzw. der Levelmengen) zu einer Abschätzung

$$D_f(\xi, \tilde{\xi}) \leq C D_f^\infty(\xi, \tilde{\xi}) \quad (6.27)$$

für alle $\xi, \tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{Td})$ und einer festen Konstanten $C > 0$ (vgl. [38, Bedingung (A3)]). Die Bedeutung von Proposition 6.9 liegt auf der Hand. Gelingt es für eine hinreichend feine Diskretisierung $\mathcal{D}^{(n_0)}$, eine gute Näherung ξ_{tr} für ein $\hat{\xi}^{(n_0)}$ im Sinne von Proposition 6.9 empirisch zu schätzen, sichern die Aussagen (i) und (ii) bzgl. des existierenden gemeinsamen Wahrscheinlichkeitsraums, in Verbindung mit der Abschätzung (6.27) und durch Anwendung des Stabilitätstheorems 4.1, die Nähe der Optimalwerte $v(\xi_{\text{tr}})$ und $v(\hat{\xi}^{(n_0)})$.

Empirische Verteilungen

Es sei ξ ein Ξ -wertiger stochastischer Prozess definiert auf irgendeinem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Verteilung P . Des Weiteren sei $(\xi^k)_{k \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter und ebenfalls Ξ -wertiger Zufallsvariablen definiert auf einem Wahrscheinlichkeitsraum $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$, so dass $\mathcal{L}(\xi^{(1)}) = P$ gilt. Betrachtet man die Folge empirische Verteilungen $(P^{(k)}(\omega^*))$, definiert durch

$$P^{(k)}(\omega^*)(B) := \frac{1}{k} \sum_{j=1}^k \delta_{\xi^j(\omega^*)}(B), \quad n \in \mathbb{N}, \omega^* \in \Omega^*, B \in \mathcal{B}(\Xi), \quad (6.28)$$

wobei δ_x das Dirac-Maß an der Stelle $x \in \Xi$ bezeichnet, so konvergiert diese fast sicher gegen die Verteilung P von ξ (vgl. zum Beispiel [17, Kapitel 11.4]). Eine wichtige Konsequenz der Verteilungskonvergenz ist folgende Aussage, die sich direkt von [17, Theorem 11.1.1] ableiten lässt.

Proposition 6.10 *Es sei $(P^{(k)})$ eine Folge von empirischen Verteilungen nach Konstruktion (6.28). Dann gilt für fast alle $\omega^* \in \Omega^*$*

$$\lim_{k \rightarrow \infty} P^{(k)}(\omega^*)(B) = P(B), \quad \text{für alle } B \in \mathcal{B}(\Xi) \text{ mit } P(\partial B) = 0,$$

wobei ∂B den Rand der Borel-Menge B im normierten Raum R^s bezeichnet.

In Verbindung mit konvergenten Diskretisierungen und dem Konvergenzkriterium von Proposition 6.9 stellt dieses Resultat das noch fehlende Bindeglied zur Rechtfertigung von solchen Algorithmen zur Konstruktion von approximativen Szenariobaumprozessen dar, die (wie die beiden Algorithmen 6.2 und 6.5) auf empirischen oder statistischen Schätzungen aufbauen.

6.2.3 Anwendung auf die Szenariobaumkonstruktion

Auf Grundlage der bisherigen Betrachtungen wird nun ein genereller algorithmischer Ansatz zur Generierung approximativer Szenariobaumprozesse abgeleitet, der unter Ausnutzung der vorangegangenen Resultate, in Bezug auf Optimalwerte, Folgen konvergenter Prozesse in Form von Szenariobäumen generiert. Die Konstruktionsalgorithmen in Kapitel 6.1, d.h. sowohl der Vorwärts- als auch der Rückwärtsalgorithmus zur Baumkonstruktion, können dann als spezielle (heuristische) Umsätzungen des generellen Ansatzes interpretiert werden. Folgender Prototyp eines Approximationsalgorithmus wird betrachtet:

Algorithmus 6.11 *Es sei ξ der originale Ξ -wertige stochastische Eingangsprozess des linearen stochastischen Optimierungsproblems (4.1), definiert auf irgendeinem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Verteilung $P_\xi \in \mathcal{P}(\Xi)$:*

- Schritt [1]:** *Bestimme eine Folge von Diskretisierungen $\mathcal{D}^{(n)}$ in Ξ , so dass Eigenschaften (E1)–(E3) erfüllt sind (vgl. Abschnitt 6.2.1) und fixiere ein $n \in \mathbb{N}$.*
- Schritt [2]:** *Bestimme ein Sampling von k ($k \in \mathbb{N}$) Szenarien $\{\xi^1, \dots, \xi^k\} \subseteq \Xi$, so dass ξ^j ($1 \leq j \leq k$) Realisierungen unabhängig, identisch P_ξ -verteilter Zufallsgrößen sind.*
- Schritt [3]:** *Bestimme die Wahrscheinlichkeiten $P^{(k)}(D_1 \times \dots \times D_T)$ für alle $D_1 \times \dots \times D_T \in \mathcal{D}^{(n)}$ der empirischen Verteilung $P^{(k)}$ gemäß Formel (6.28).*

Schritt [4]: Wähle für jedes $D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}$ nichtantizipative Szenarien, deren t -ten Komponenten zu D_t gehören und durch $C \cdot n$ ($C > 0$) beschränkt sind (analog zur Konstruktion (6.22)).

Schritt [5]: Konstruiere abschließend einen stochastischen Prozess $\xi_{\text{tr}} := \xi_{\text{tr}}^{(n,k)}$ mit den in Schritt 4 ausgewählten Szenarien, versehen mit den in Schritt 3 bestimmten zugehörigen empirischen Wahrscheinlichkeiten.

Betrachtet man eine Folge von Szenariobaumprozessen $(\xi_{\text{tr}}^{(n,k)})_{n,k \in \mathbb{N}}$, die in Abhängigkeit der gewählten Parameter (n für die Feinheit der Diskretisierung und k für die Anzahl der gesampelten Szenarien) mit Algorithmus 6.11 konstruiert wurde, so gilt folgende Konvergenzeigenschaft:

Theorem 6.12 *Es seien die Bedingungen (A1), (A2) und (A3) erfüllt, wobei in (A3) sämtliche Levelmengen $S_\alpha(\xi)$ wesentlich beschränkt und insbesondere für $\alpha = 0$ nichtleer seien. Des Weiteren sei $1 \leq r' < \infty$ sowie $(\xi_{\text{tr}}^{(n,k)})_{n,k \in \mathbb{N}}$ eine mit Algorithmus 6.11 konstruierte Folge von Szenariobäumen. Ist jeweils $\mathcal{D}^{(n+1)}$ eine Verfeinerung von $\mathcal{D}^{(n)}$ und gilt $P_\xi(\partial D_t) = 0$ für alle $D_t \in \mathcal{D}_t^{(n)}$ ($t = 1, \dots, T$) in der Folge der zugrunde liegenden Diskretisierungen, dann gilt für die Optimalwerte von (4.1)*

$$\lim_{n \rightarrow \infty} \left(\lim_{k \rightarrow \infty} v(\xi_{\text{tr}}^{(n,k)}) \right) = v(\xi) \quad (6.29)$$

mit Wahrscheinlichkeit 1.

Beweis: Unter Verwendung der Notation in Abschnitt 6.2.1 werde mit

$$\mathcal{S}^{(n)} := \left\{ \hat{\xi}_{D_1 \times \cdots \times D_T}^{(n)} : D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)} \right\} \quad (6.30)$$

die Menge der gewählten Szenarien der Diskretisierung $\mathcal{D}^{(n)}$ bezeichnet. Für jedes gewählte $\delta > 0$ gilt für hinreichend große $n \in \mathbb{N}$, d.h. bei hinreichend feiner Diskretisierung $\mathcal{D}^{(n)}$, für den durch (6.23) definierten diskreten Prozess

$$\hat{\xi}^{(n)} = \sum_{D_1 \times \cdots \times D_T \in \mathcal{D}^{(n)}} \hat{\xi}_{D_1 \times \cdots \times D_T}^{(n)} \cdot \mathbb{1}_{D_1 \times \cdots \times D_T}(\xi)$$

aufgrund von Proposition 6.8, dass $\|\xi - \hat{\xi}^{(n)}\|_r \leq \delta$, was die Anwendung des Stabilitätstheorems 4.1 erlaubt. Da Proposition 6.8 darüber hinaus die Adaptiertheit von $\hat{\xi}^{(n)}$ zu ξ sichert, erhält man wegen Proposition 6.7 die Konvergenz des Filtrationsabstands $D_f(\xi, \hat{\xi}^{(n)})$, was mit der Stabilitätsungleichung (4.10) zu einer Abschätzung

$$|v(\xi) - v(\hat{\xi}^{(n)})| \leq \varepsilon_n$$

führt, wobei $\varepsilon_n \rightarrow 0$ für $n \rightarrow \infty$ gilt.

Es sei nun $\xi_{\text{tr}}^{(n,k)} = \xi_{\text{tr}}^{(n,k)}(\omega^*)$ der durch die Schritte 2 bis 5 des Algorithmus 6.11 festgelegte (empirische) Prozess. Es genügt zu zeigen, dass für jedes $n \in \mathbb{N}$ ein Index $k(n)$ existiert, so dass für alle $k \geq k(n)$ und fast alle ω^*

$$|v(\hat{\xi}^{(n)}) - v(\xi_{\text{tr}}^{(n,k)}(\omega^*))| \leq \frac{1}{n}$$

gilt. Dann folgt aus der Dreiecksungleichung des Betrags unmittelbar

$$|v(\xi) - v(\xi_{\text{tr}}^{(n,k)})| \leq \varepsilon_n + \frac{1}{n} \quad (k \geq k(n))$$

mit Wahrscheinlichkeit 1 und damit die Behauptung.

Sei $n \in \mathbb{N}$ im Folgenden beliebig fixiert. Vergleicht man die diskreten Prozesse $\hat{\xi}^{(n)}$ und $\xi_{\text{tr}}^{(n,k)}$ ($k \in \mathbb{N}$), so stellt man fest, dass alle Prozesse identische Szenarien der Form (6.30) besitzen (diese hängen nur von der Diskretisierung und der nichtantizipativen Auswahl ab). Lediglich die Wahrscheinlichkeiten variieren. Wenn $p_{D_1 \times \dots \times D_T}^{(n)}$ und $p_{D_1 \times \dots \times D_T}^{(n,k)}$ die jeweilige Wahrscheinlichkeiten des Szenarios $\hat{\xi}_{D_1 \times \dots \times D_T}^{(n)} \in \mathcal{S}^{(n)}$ bezeichnen, so gilt

$$\begin{aligned} p_{D_1 \times \dots \times D_T}^{(n)} &= P_\xi(D_1 \times \dots \times D_T), \\ p_{D_1 \times \dots \times D_T}^{(n,k)} &= P^{(k)}(D_1 \times \dots \times D_T) \quad (k \in \mathbb{N}), \end{aligned}$$

wobei P_ξ die Verteilung des gegebenen Ausgangsprozess ξ und $P^{(k)}$ die empirischen Verteilungen bezeichnen. Mit Proposition 6.10 erhält man

$$p_{D_1 \times \dots \times D_T}^{(n,k)} \rightarrow p_{D_1 \times \dots \times D_T}^{(n)} \quad (k \rightarrow \infty)$$

mit Wahrscheinlichkeit 1. Anwendung von Proposition 6.9 impliziert daher die Existenz eines gemeinsamen Wahrscheinlichkeitsraums, auf dem

$$\|\hat{\xi}^{(n)} - \xi_{\text{tr}}^{(n,k)}\| \rightarrow 0 \quad \text{und} \quad D_f^\infty(\hat{\xi}^{(n)}, \xi_{\text{tr}}^{(n,k)}) \rightarrow 0 \quad (k \rightarrow \infty).$$

Da die wesentliche Beschränktheit der Lösungen des Optimierungsproblems für alle Eingangsprozessen in einer Umgebung von ξ vorausgesetzt wurde, sind auch die Lösungen bzgl. $\hat{\xi}^{(n)}$ und damit auch für $\xi_{\text{tr}}^{(n,k)}$ für hinreichend großes n wesentlich beschränkt. Damit gilt die Abschätzung (6.27) für den Filtrationsabstand und erneutes Anwenden der Stabilitätsungleichung (4.10) liefert

$$|v(\hat{\xi}^{(n)}) - \xi_{\text{tr}}^{(n,k)}| \rightarrow 0 \quad (k \rightarrow \infty)$$

und insbesondere

$$|v(\hat{\xi}^{(n)}) - \xi_{\text{tr}}^{(n,k)}| \leq \frac{1}{n}$$

für alle $k \geq k(n)$ mit Wahrscheinlichkeit 1. □

Generelle Interpretation

Algorithmus 6.11 beinhaltet in Verbindung mit dem Konvergenzsatz von Theorem 6.12 zwei wesentliche Punkte hinsichtlich des generellen Vorgehens zur Konstruktion eines approximativen Szenariobaums: Einerseits muss die Genauigkeit der Diskretisierung des Zustandsraums kontrolliert werden, andererseits muss für eine genügend große Anzahl gesampelter Szenarien zur Bestimmung der empirischen Verteilung gesorgt werden. Natürlich besteht hier eine klare Abhängigkeit. Je feiner die Diskretisierung gewählt wird, desto mehr Szenarien sind auch im Schritt 2 des Algorithmus nötig, um zu einer guten Näherung der Verteilung durch die empirische Schätzung zu gelangen. Ist also die Anzahl der verfügbaren Szenarien von Beginn an klein, ist es auch nicht sinnvoll, die Diskretisierungsgenauigkeit zu groß zu wählen.

Eine andere Frage betrifft die Diskretisierung selbst. Es ist zunächst nicht klar, in welcher Art und Weise der Träger Ξ partitioniert werden soll. Als zweckmäßig erscheint hier eine feinere Zerlegung immer dort, wo eine Konzentration an Information zu finden ist, also in Bereichen höherer Wahrscheinlichkeiten hinsichtlich der Verteilung von ξ .

Bemerkungen zu Algorithmen 6.2 und 6.5

Sowohl der Rückwärts- als auch der Vorwärtsalgorithmus zur Szenariobaumkonstruktion (Algorithmus 6.2 bzw. 6.5) können als spezielle heuristische Umsäzungen des generellen Vorgehens in Algorithmus 6.11 interpretiert werden. Folgt man den Ideen in Abschnitt 6.1, starten beide Algorithmen mit einer Anfangsapproximation ξ_f individueller Szenarien, für die nach Formel (6.1)

$$\|\xi - \xi_f\|_r \leq \varepsilon_{\text{appr}}$$

auf irgendeinem Wahrscheinlichkeitsraum gilt. Die Bestimmung eines derartigen Prozess ξ_f korrespondiert mit Schritt 2 von Algorithmus 6.11. Eine wichtige Rolle bei der Konstruktion der Szenariobäumen übernehmen die Parameter

$$\varepsilon_t \quad (t = 2, \dots, T) \quad \text{mit} \quad \sum_{t=2}^T \varepsilon_t \leq \varepsilon$$

in beiden Algorithmen. Sowohl der Vorwärts- als auch der Rückwärtsalgorithmus basieren auf einer stufenweisen Szenarioreduktion. Die Gestalt der individuellen Zerlegung der t -ten Komponente Ξ_t des Zustandsraums Ξ wird daher durch die in der jeweiligen Zeitperiode durchgeführte Szenarioreduktion bestimmt, die wiederum durch die Wahl von ε_t gesteuert wird. Dabei ist ε_t entscheidend für die Feinheit der Zerlegung verantwortlich. Kleinere Werte für ε_t bedeuten auch feinere Zerlegungen der jeweiligen Komponente. Das allgemeine Problem der optimalen Szenarioreduktion (5.9) berücksichtigt beides, den Normabstand von Szenarien zueinander sowie deren Wahrscheinlichkeiten. Das Ergebnis der

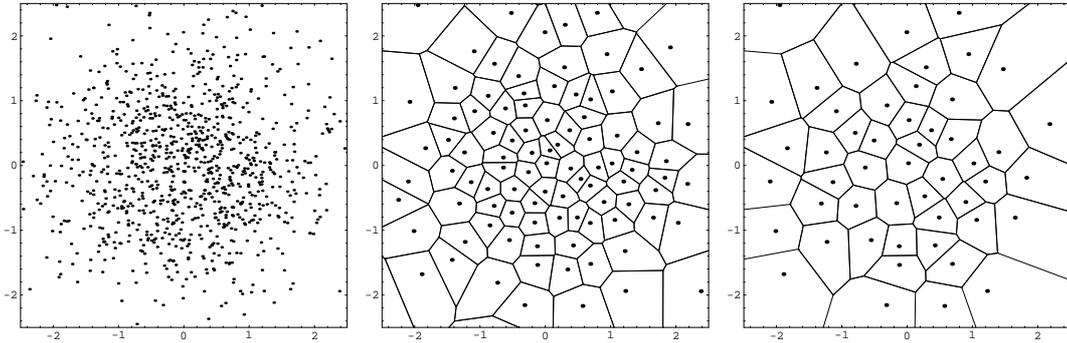


Abbildung 6.4: Illustration der Szenarioreduktion am Beispiel der 2-dimensionalen Standardnormalverteilung ausgehend von 1 000 generierten Szenarien (links) bei einer Reduktion auf 100 Szenarien (Mitte) und 50 Szenarien (rechts). Dargestellt sind die durch die Szenarioreduktion bestimmten Zerlegung der Ebene in Voronoi-Regionen bzgl. der Euklidischen Norm.

Szenarioreduktion ist daher eine von der Wahrscheinlichkeitsdichte abhängige endliche Aufteilung des Raums in Teilgebiete (sogenannte Voronoi-Regionen), erzeugt durch die aus der Reduktion hervorgehenden Szenarien. Die einzelnen Regionen sind dort besonders fein, wo sich die Ausgangsszenarien, d.h. die Wahrscheinlichkeiten konzentrieren. Abbildung 6.4 zeigt ein Beispiel für die Standardnormalverteilung im \mathbb{R}^2 . Eine solche Zerlegung wird auch Quantisierung genannt. Falls $\{\xi^1, \dots, \xi^N\}$ eine Menge von Punkten in \mathbb{R}^d bezeichnet, so sind die zugehörigen Voronoi-Regionen im Allgemeinen durch

$$V(\xi^j) := \left\{ \xi \in \mathbb{R}^d : \|\xi - \xi^j\| < \min_{k \in \{1, \dots, N\} \setminus \{j\}} \|\xi - \xi^k\| \right\}, \quad j = 1, \dots, N,$$

definiert. In [31] wird gezeigt, dass für strikt konvexe Normen (zum Beispiel p -Normen mit $1 < p < \infty$) der Abschluss dieser Mengen eine Überdeckung in \mathbb{R}^d definieren, deren Ränder $\partial V(\xi^j)$ Nullmengen jeder bezüglich des Lebesgue-Maßes absolut stetigen Verteilung sind. Darüber hinaus liefern die Konstruktionsmethoden 6.2 und 6.5 automatisch nichtantizipative Szenarien (vgl. Abschnitt 6.2.1) bzgl. des kartesischen Produkts der erzeugten Diskretisierungen von Ξ_t , da bei der rekursiven Anwendung der Szenarioreduktion jeweils in Stufe t nur die Szenarienvläufe bis zu diesem Zeitpunkt Berücksichtigung finden.

Innerhalb der Algorithmen 6.2 und 6.5 findet also heuristisch eine bzgl. der Konvergenztheorie geeignete simultane Diskretisierung, Auswahl nichtantizipativer Szenarien sowie Schätzung der Wahrscheinlichkeiten statt. Wählt man die Parameter ε_t für $t = 2, \dots, T$ innerhalb der Algorithmen 6.2 und 6.5 zur Konstruktion des Szenariobaums klein genug, so kann im Sinne von Theorem 6.12, bei gleichzeitig hinreichend großer Anzahl von Ausgangsszenarien, die Nähe der Optimalwerte bzgl. des originalen Eingangsprozesses und des Szenariobaumprozesses erwartet werden.

Kapitel 7

Anwendungen

In diesem Kapitel werden numerische Resultate für Anwendungen aus der Praxis zusammengestellt. Die umfangreichen Rechnungen umfassen sowohl ausgewählte Beispiele für die Generierung multivariater Szenariobäume für die Optimierung in der Energiewirtschaft und bei der Ertragssteuerung von Buchungsanfragen großer Airlines (Abschnitt 7.2) als auch eine Diskussion der verschiedenen Algorithmen zur Szenarioreduktion im folgenden Abschnitt.

7.1 Ergebnisse der Szenarioreduktion

Zum Testen der Methoden zur Szenarioreduktion aus Kapitel 5 wird auf ein bestehendes Beispiel zur Modellierung von Lastszenarien für ein stochastisches Optimierungsmodell zur Stromerzeugung zurückgegriffen (vgl. [19, 36]). Der

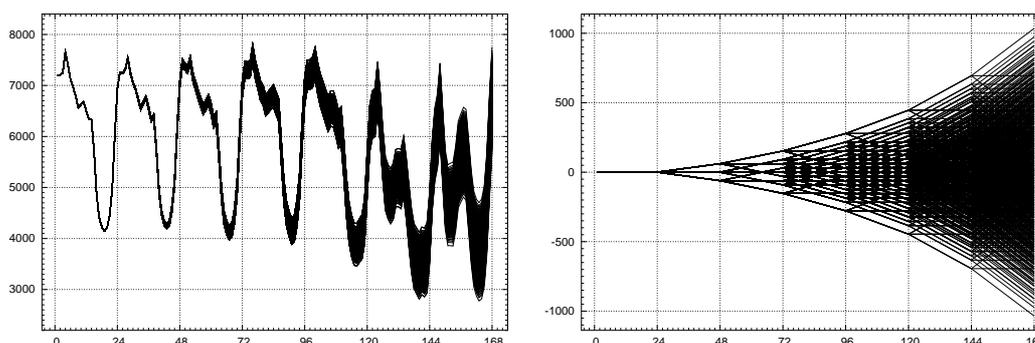


Abbildung 7.1: Simulierte Last-Szenarien für eine Woche sowie der konstruierte Szenariobaum mit Erwartungswert-bereinigten Daten.

dort beschriebene Szenariobaum basiert auf einem kalibrierten Zeitreihenmodell für die elektrische Last. Ausgehend von der Simulation einer großen Zahl von Realisierungen wurde, unter Verwendung des Mittelwerts und der Standardabweichungen, ein initialer 3-faltiger Szenariobaum konstruiert, welcher die

Anzahl Szenarien	Relativer $\overset{\circ}{\mu}_r$ -Abstand						
	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.522	0.646	0.684	0.696	0.687	0.682	0.668
10	0.419	0.536	0.589	0.577	0.582	0.556	0.535
20	0.323	0.420	0.469	0.472	0.466	0.431	0.395
50	0.230	0.305	0.335	0.337	0.301	0.256	0.210
100	0.169	0.220	0.242	0.222	0.180	0.133	0.094
150	0.137	0.178	0.185	0.156	0.114	0.077	0.049
200	0.117	0.148	0.143	0.112	0.076	0.045	0.025
300	0.094	0.102	0.085	0.057	0.032	0.016	0.008
400	0.072	0.067	0.049	0.028	0.013	0.006	0.002
500	0.050	0.039	0.024	0.012	0.005	0.002	0.001
600	0.028	0.018	0.009	0.004	0.001	0.000	0.000

Tabelle 7.1: Numerische Ergebnisse der optimalen Szenarioreduktion für $\overset{\circ}{\mu}_r$.

Anzahl Szenarien	Relativer $\hat{\mu}_r$ -Abstand						
	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
1	1.000	1.609	2.354	3.146	3.910	4.627	5.302
5	0.522	0.738	0.940	1.079	1.209	1.217	1.257
10	0.419	0.574	0.713	0.787	0.820	0.803	0.794
20	0.323	0.448	0.538	0.600	0.617	0.601	0.565
50	0.230	0.308	0.359	0.378	0.369	0.331	0.286
100	0.169	0.221	0.253	0.248	0.211	0.168	0.130
150	0.137	0.179	0.192	0.171	0.134	0.097	0.066
200	0.117	0.149	0.147	0.121	0.088	0.058	0.035
300	0.094	0.102	0.088	0.062	0.037	0.021	0.011
400	0.072	0.067	0.050	0.030	0.015	0.007	0.003
500	0.050	0.039	0.025	0.012	0.005	0.002	0.001
600	0.028	0.018	0.009	0.004	0.001	0.000	0.000

Tabelle 7.2: Numerische Ergebnisse der optimalen Szenarioreduktion für $\hat{\mu}_r$.

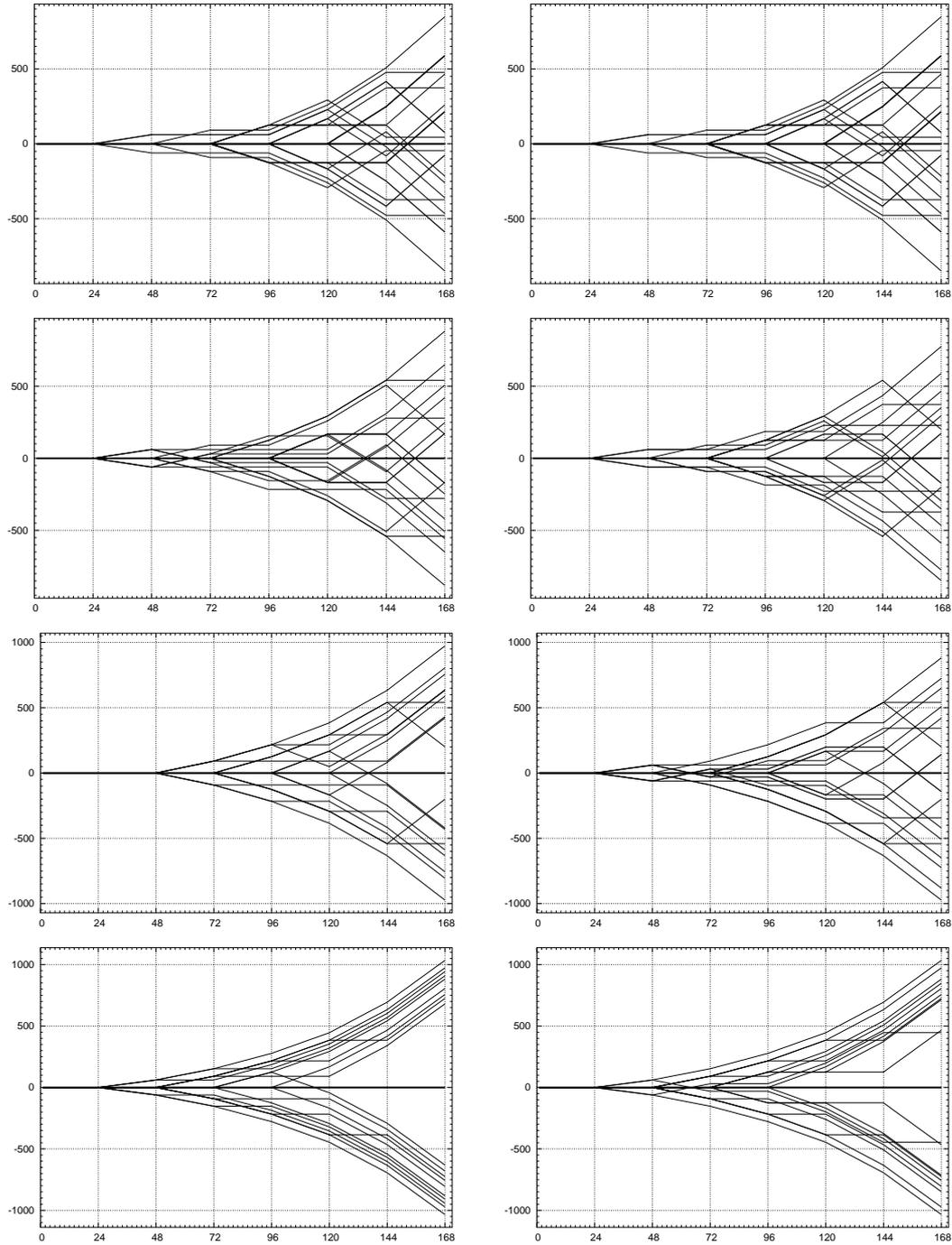


Abbildung 7.2: Reduzierte Szenariobäume mit jeweils 20 Szenarien ermittelt unter Verwendung von $\hat{\mu}_r$ (linke Spalte) und $\hat{\mu}_r$ (rechte Spalte) für $r = 1, 2, 4, 7$.

zunehmende Unsicherheit im Laufe des Planungszeitraum abbildet. Abbildung 7.1 zeigt die simulierten Szenarien des stochastischen Lastprozesses für eine Woche sowie den konstruierten Szenariobaum bei einer stündlichen Diskretisierung. Bei der Darstellung des Szenariobaums sind nur die Schwankungen um den empirischen Mittelwert herangezogen worden. Der Verzweigungsgrad ist jeweils 3, wobei alle 24 Stunden verzweigt wird. Das führt zu insgesamt $3^6 = 729$ Szenarien mit identischer Wahrscheinlichkeit. Der so konstruierte Szenariobaum repräsentiert eine diskrete Wahrscheinlichkeitsverteilung, die Eingang in ein 7-periodisches stochastisches Optimierungsproblem fand.

Während Tabelle 7.1 die Ergebnisse der optimalen Szenarioreduktion des Lastszenariobaums für den Vorwärtsalgorithmus 5.3 unter Verwendung des Kantorovich-Rubinstein-Abstandes $\overset{\circ}{\mu}_r$ zeigt, sind in Tabelle 7.2 die Resultate für denselben Algorithmus, aber unter Verwendung des Monge-Kantorovich-Funktional $\hat{\mu}_r$ zu sehen. Um den Einfluss des Parameters r genauer studieren zu können, wurden die Werte r für beide Rechnungen variiert. Beide Tabellen verwenden gewisse relative Abstände, die jeweils zwischen dem originalen Lastszenariobaum auf der einen Seite und einer Vielzahl reduzierter Szenariobäume auf der anderen Seite gebildet werden. Die relativen Abstände berechnen sich aus dem Quotienten des ermittelten Kantorovich-Rubinstein- bzw. Monge-Kantorovich-Abstands und dem bestmöglichen Kantorovich-Rubinstein-Abstand, aufgestellt vom originalen Szenariobaum zu einem seiner Szenarien versehen mit Wahrscheinlichkeit 1. Anders ausgedrückt, der relative Abstand ergibt sich, indem durch den $\overset{\circ}{\mu}_r$ -Abstand zwischen der Verteilung aller Szenarien und dem Dirac-Maß desjenigen Szenario, welches mit dem ersten Schritt des Vorwärtsalgorithmus gefunden wurde, dividiert wird.

Zur Bestimmung eines reduzierten Szenariobaums für $r = 1$, werden, auf einem PC mit 3 GHz Rechenleistung, nicht mehr als 10 Sekunden benötigt, wobei davon allein etwa 4 Sekunden auf die Berechnung der Kostenfunktion $c_r(\cdot, \cdot)$ entfallen. Für die Berechnungen mit $r > 1$ bei der Verwendung von $\overset{\circ}{\mu}_r$ müssen zusätzliche die reduzierten Kosten $\hat{c}_r(\cdot, \cdot)$ bestimmt werden, was zusätzlich rund 9 Sekunden erfordert.

Abbildung 7.2 zeigt die Struktur der reduzierter Szenariobäume sowohl in Abhängigkeit der verwendeten Abstände $\overset{\circ}{\mu}_r$ (linke Spalte) und $\hat{\mu}_r$ (rechte Spalte) als auch in Abhängigkeit von der Wahl des Parameters r . Alle Bilder zeigen das Ergebnis der Reduktion auf 20 Szenarien wobei von oben nach unten $r = 1$, $r = 2$, $r = 4$ und $r = 7$ gewählt wurde. Da Approximationen der Wahrscheinlichkeitsverteilung bezüglich $\overset{\circ}{\mu}_r$ bzw. $\hat{\mu}_r$ die Eigenschaft besitzen, dass die Momente r -ter Ordnung näherungsweise abgebildet werden (siehe Proposition 3.3), führt die Veränderung von r auch erwartungsgemäß zu einer unterschiedlichen Auswahl von Szenarien während des Reduktionsprozesses.

In Abbildung 7.3 ist das Verhalten der relativen Wahrscheinlichkeitsabstände bzgl. $\overset{\circ}{\mu}_r$ und $\hat{\mu}_r$ für verschiedene Parameter r dargestellt. Die abgebildeten Kur-

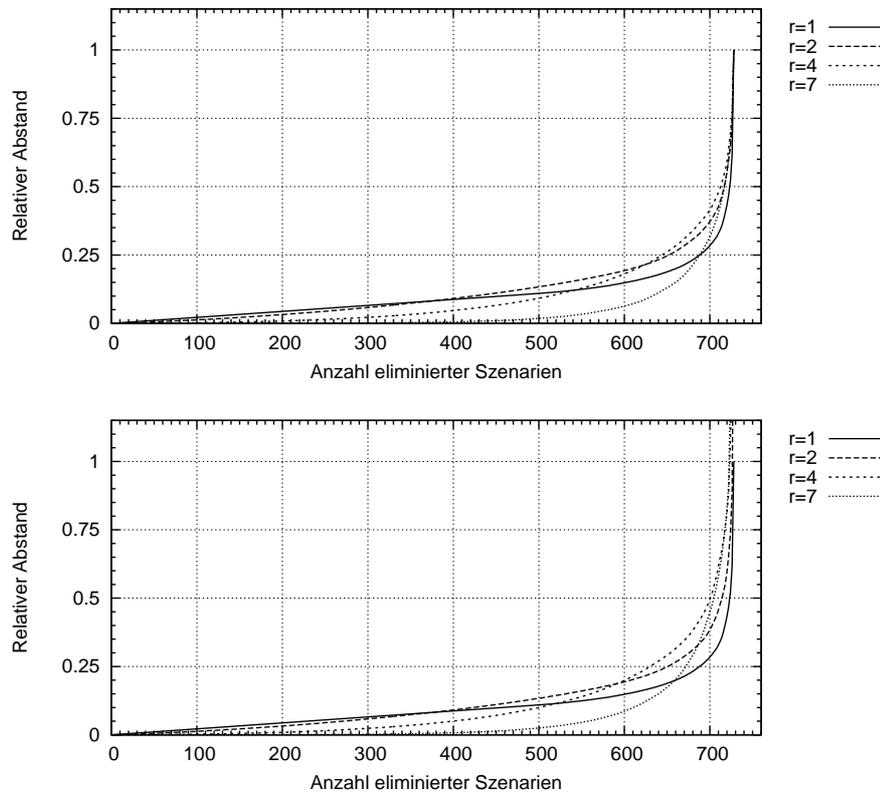


Abbildung 7.3: Verhalten des relativen μ_r -Abstands (oben) bzw. des relativen $\hat{\mu}_r$ -Abstands (unten) in Abhängigkeit von der Anzahl reduzierter Szenarien für $r = 1, 2, 4, 7$.

ven zeigen den zunehmenden relativen Abstand der ermittelten Verteilungen zur Originalverteilung in Abhängigkeit von der Anzahl der reduzierten Szenarien. Besonders auffällig ist, dass alle Kurven zu Beginn einen sehr schwachen Anstieg aufweisen. Das führt zu der wichtigen Erkenntnis, dass sich das Eliminieren von Szenarien, unabhängig von der konkreten Wahl des Wahrscheinlichkeitsabstands, zunächst nur sehr wenig auf die relative Genauigkeit auswirkt. Erst bei einer sehr starken Reduktion im Bereich von rund 600 eliminierten Szenarien nimmt der relative Fehler sprunghaft zu. Dieses Phänomen wird durch wachsende r noch verstärkt. In diesem Beispiel mit 729 Szenarien zeigt sich, dass etwa 100 Szenarien noch immer rund 80-90 Prozent der Verteilungsinformation enthalten können.

7.1.1 Bewertung der Reduktionsalgorithmen

Das Problem (5.2) der optimalen Szenarioreduktion beschreibt, wie bereits festgestellt wurde, ein schwer handhabbares kombinatorisches Optimierungsproblem, für das eine exakte Lösung bzw. eine ε -approximative Näherungslösung

entweder garnicht oder nur mit sehr hohem Aufwand bestimmt werden kann (vgl. Kapitel 5). Die heuristischen Algorithmen 5.3 und 5.4 liefern praktikable Ansätze zur näherungsweise Lösung des Problems und sind aufgrund ihrer Schnelligkeit auch bei sehr großen Szenarienmengen anwendbar. Weitgehend offen ist die Frage nach der Genauigkeit. Die Algorithmen selbst liefern keine unmittelbare Möglichkeit zur Berechnung unterer Schranken. Um dennoch Aussagen über die Approximationsgüte treffen zu können, zeigt Tabelle 7.3 einen

Anzahl Szenarien	Vorwärts		Rückwärts		CPLEX 10.0		
	$\hat{\mu}_1$	Zeit [s]	$\hat{\mu}_1$	Zeit [s]	$\hat{\mu}_1$	Schranke	Zeit [s]
1	1.000	0.00	1.119	0.14	1.000	1.000	36.28
5	0.559	0.01	0.552	0.14	0.548	0.545	72.00
10	0.451	0.03	0.462	0.14	0.445	0.441	114.39
15	0.402	0.04	0.412	0.13	0.394	0.390	782.34
20	0.369	0.05	0.375	0.14	0.358	0.355	4 751.55
30	0.315	0.06	0.317	0.13	0.303	0.301	3 278.65
40	0.275	0.08	0.277	0.13	0.263	0.260	10 445.90
50	0.244	0.10	0.245	0.13	0.237	0.235	245.22
60	0.221	0.12	0.221	0.11	0.219	0.217	66.81
70	0.203	0.12	0.201	0.11	0.200	0.200	13.79
80	0.187	0.14	0.184	0.10	0.182	0.182	14.92
90	0.173	0.15	0.171	0.10	0.171	0.171	9.60
100	0.160	0.16	0.160	0.09	0.160	0.160	9.84
125	0.132	0.18	0.132	0.07	0.132	0.132	9.56
150	0.104	0.19	0.104	0.05	0.104	0.104	9.59
200	0.048	0.20	0.048	0.03	0.048	0.048	9.35

Tabelle 7.3: Vergleich der Reduktionsalgorithmen mit CPLEX 10.0 am Beispiel des gekürzten Last-Szenariobaums mit $3^5 = 243$ Szenarien.

Vergleich mit Cplex, einem leistungsfähigen Löser für lineare und gemischt-ganzzahlige Optimierungsprobleme. Da das Lastbeispiel mit 729 Szenarien bereits zu groß ist, als dass es von Cplex bewältigt werden kann (bei der Reduktion auf beispielsweise 10 Szenarien bricht Cplex nach mehreren Stunden mit Speicherüberlauf ab), wurde für den Vergleich der Szenariobaum um eine Stufe gekürzt, so dass er insgesamt nur noch 243 Szenarien umfasst.

Für alle Berechnung wurde der Monge-Kantorovich-Abstand $\hat{\mu}_1$ herangezogen, wobei die berechneten Werte erneut die relativen Abstände repräsentieren. In der Tabelle sind die Ergebnisse und Rechenzeiten der beiden heuristischen Algorithmen 5.3 und 5.4 sowie die von Cplex gegenübergestellt, wobei für Cplex eine relative Genauigkeit von 0.01 gewählt wurde. Das führt im Ergebnis zu zwei

Werten für Cplex, zum einen zur berechneten Näherungslösung und zum anderen zu einer unteren Schranke für den tatsächlichen Optimalwert. Prinzipiell erlaubt Cplex auch höhere Genauigkeiten, allerdings steigt dann die Rechenzeit sprunghaft, häufig um ein Vielfaches, an. Eine höhere Genauigkeit muss aber nicht zwangsläufig zu einer besseren Lösung führen. Im Beispiel führte eine noch höhere Genauigkeit oft nur zu besseren unteren Schranken, so dass sich auf diese Weise lediglich die Optimalitätslücke verkleinerte. Aufgrund des enormen Mehraufwands an Rechenzeit, wurde in diesem Beispiel letztendlich auf genauere Berechnungen verzichtet.

7.2 Anwendung der Baumkonstruktion

In diesem Abschnitt wird anhand von ausgewählten Anwendungen die Einsatzvielfalt der Algorithmen 6.2 und 6.5 zur Szenariobaumkonstruktion in der Praxis dokumentiert. Es werden Beispiele aus dem Bereich der Optimierung in der Energiewirtschaft (Abschnitt 7.2.1 und 7.2.2) sowie zur Optimierung der Ertragssteuerung von Buchungsanfragen und Flugnetzen von Fluggesellschaften (Abschnitt 7.2.3) betrachtet.

7.2.1 Szenariobäume für EDF

Dieses Beispiel basiert auf Szenarien für ein Zweijahresmodell zur Optimierung des Einsatzfahrplans eines Kraftwerkverbunds zur Energieerzeugung, die vom französischen Energiekonzern EDF (Electricité de France) bereitgestellt wurden. Die Daten stellen endlich viele Realisierungen eines bivariaten stochastischen Prozesses dar, dessen Komponenten sich einerseits aus elektrischer Last und andererseits aus Zuflüssen zu hydraulischen Kraftwerken zusammensetzen. Jede der Realisierungen ist mit derselben Wahrscheinlichkeit versehen. Beide Komponenten des stochastischen Prozesses treten nur als rechte Seiten in den linearen Restriktionen des stochastischen Optimierungsmodells auf. Der Zeit-

Zufallsdaten	Unterteilungen	Zeitperioden
Elektrische Last	3 pro Tag	2 184
Wasser-Zufluss	wöchentlich	104

Tabelle 7.4: Unterteilung des Optimierungszeitraums von zwei Jahren.

	Anzahl
Szenarien	456
Zeitperioden	2 184
Knoten	995 449

Tabelle 7.5: Dimension der Ausgangsdaten.

raum von zwei Jahren ist in drei Zeitperioden pro Tag unterteilt, die für sich jeweils Stunden ähnlichen Bedarfs zusammenfassen. Die Tabellen 7.4 und 7.5 zeigen einen Überblick über die wichtigsten relevanten Eingangsparameter. Die nur wochenweise gegebenen Werte für die Zuflüsse wurden gleichmäßig auf die zugehörigen Zeitperioden verteilt.

In unterschiedlichen Testreihen wurden Szenariobäume unter Verwendung der beiden Algorithmen 6.2 und 6.5 zur Szenariobaumkonstruktion so generiert, dass Verzweigungen

- a) zu jeder Zeitperiode,
- b) nur einmal pro Tag bzw.
- c) nur zu Beginn einer jeden Woche

erlaubt waren (eine entsprechende Steuerung ist durch die jeweiligen Konstruktionsparameter innerhalb der Algorithmen möglich). Um wie bei der Szenarioreduktion einen relativen Wert für das Maß der Genauigkeit in Bezug auf die Verteilungen der Ausgangsdaten und der generierten Szenariobäume zu bekommen, werden analog relative Toleranzen

$$\varepsilon_{rel} := \frac{\varepsilon}{\varepsilon_{max}} \quad \text{bzw.} \quad \varepsilon_{t,rel} := \frac{\varepsilon_t}{\varepsilon_{max}} \quad (t = 2, \dots, T) \quad (7.1)$$

betrachtet. Der Wert für ε_{max} bezeichnet den kleinstmöglichen Verteilungsabstand (hier bzgl. ℓ_r), der durch die Ausgangsverteilung aller Eingangsszenarien sowie der Verteilung eines einzelnen Szenarien versehen mit der Wahrscheinlichkeit 1 erzielt werden kann. Da die zufallsbehafteten Daten in diesem Modell nur die rechten Seiten betreffen, genügt es hier $r = 1$ zu setzen.

Ergebnisse des Rückwärtsalgorithmus

Für die Rückwärtsvariante der Szenariobaumkonstruktion wurden die Parameter ε_t von Algorithmus 6.2 an den zugelassenen Verzweigungsstufen durch

$$\varepsilon_T = \varepsilon \cdot (1 - q), \quad q \in (0, 1) \quad \text{und} \quad \varepsilon_t = q \cdot \varepsilon_{t+1}, \quad t = T - 1, \dots, 2, \quad (7.2)$$

festgelegt. Der Parameter $q \in (0, 1)$ kontrolliert die Aufteilung der Gesamtgenauigkeit ε auf die einzelnen Verzweigungsstufen und generiert, in Übereinstimmung mit den sich von Schritt zu Schritt verkürzenden Szenarioabschnitten innerhalb des Algorithmus, eine absteigende Folge von Werten ε_t . Der Wert für q beeinflusst das Verzweigungsverhalten der entstehenden Szenariobäume. Numerische Erfahrungen zeigen, dass Werte nahe 1 sowohl zu mehr Szenarien als auch zu einer höheren Zahl an Verzweigungen führen. Wählt man dagegen q klein, erzielt man den umgekehrten Effekt. Insbesondere bei langen Zeithorizonten und in Hinblick auf eine sinnvolle Diskretisierung gemäß Abschnitt 6.2.1, scheint ein Wert nahe 1 sinnvoll. Für die EDF-Daten wurde mit $q = 0.95$ gerechnet.

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	442	995 449	584 270	151	172.86
0.20	456	429	995 449	371 046	150	129.11
0.30	456	417	995 449	268 201	146	117.42
0.40	456	405	995 449	193 014	135	110.83
0.50	456	393	995 449	140 536	115	106.30

Tabelle 7.6: Rückwärtskonstruktion ohne Verzweigungsrestriktion.

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	442	995 449	584 793	128	134.17
0.20	456	429	995 449	373 569	124	115.47
0.30	456	417	995 449	269 850	125	110.41
0.40	456	405	995 449	196 182	120	107.4
0.50	456	393	995 449	144 009	110	104.93

Tabelle 7.7: Rückwärtskonstruktion mit täglicher Verzweigung.

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	442	995 449	589 575	88	118.47
0.20	456	429	995 449	397 047	83	110.65
0.30	456	416	995 449	293 403	86	108.40
0.40	456	405	995 449	219 714	83	106.15
0.50	456	393	995 449	170 520	81	105.16

Tabelle 7.8: Rückwärtskonstruktion mit wöchentlicher Verzweigung.

Die Tabellen 7.6 und 7.8 zeigen die Ergebnisse der Testreihen a) bis c) für verschiedene vorgegebene Approximationsparameter (ersten Spalte). Die zweite und dritte Spalte vergleicht die Dimension der Eingangsdaten mit denen der konstruierten Bäume hinsichtlich Anzahl der Szenarien und Knoten. Die vorletzte Spalte enthält Angaben über die tatsächliche Anzahl der beobachteten Verzweigungsstufen in den resultierenden Szenariobäumen. Des Weiteren geben die Tabellen Aufschluss über die Rechenzeiten auf einem Standard-PC. Alle aufgelisteten Rechenzeiten beinhalten rund 100 Sekunden für die Bestimmung der Szenarienabstände, die jeweils vor der eigentlichen Ausführung des Algorithmus berechnet wurden. Trotz der großen Datenmengen und des langen Zeithorizonts erzeugt der Konstruktionsalgorithmus die Szenariobäume sehr schnell und das weitgehend unabhängig von der vorgegebenen Toleranz. Die gewählte Toleranz wirkt sich jedoch erwartungsgemäß auf die Anzahl der Verzweigungen und auch

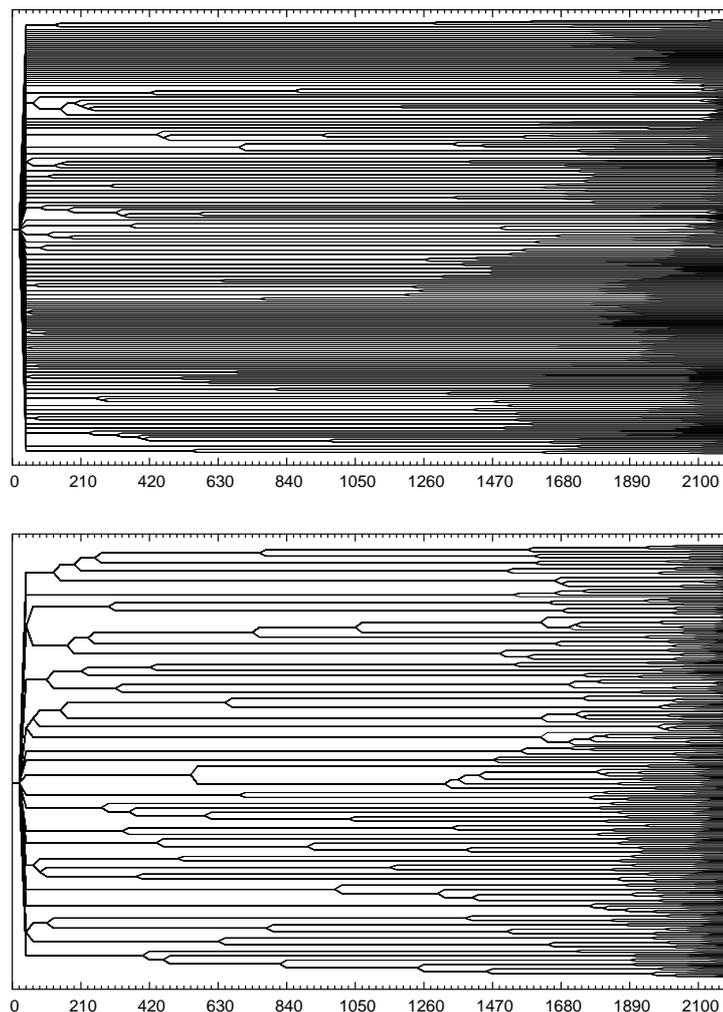


Abbildung 7.4: Mit dem Rückwärtsalgorithmus erzeugte Szenariobäume für $\varepsilon_{rel} = 0.2$ (oben) und $\varepsilon_{rel} = 0.5$ sowie wöchentliche Verzweigung.

auf die Anzahl der Knoten aus. In Abbildung 7.4 ist die Struktur zweier konstruierter Szenariobäume mit unterschiedlicher Größe und jeweils wöchentliche Verzweigungsstruktur dargestellt.

Ergebnisse des Vorwärtsalgorithmus

Analog zur Rückwärtsvariante der Szenariobaumkonstruktion wurden Testreihen mit dem Vorwärtsalgorithmus 6.5 zur Bestimmung von Szenariobäumen, ausgehend von den bivariaten EDF-Datenszenarien, aufgestellt. Da die in den Einzelschritten des Algorithmus durchgeführte Szenarioreduktion nur auf die jeweilige Zeitperiode ausgerichtet ist und damit alle Zeitperioden prinzipiell gleich behandelt werden können, ist eine exponentielle Aufteilung der Gesamtteranz

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	378	995 449	743 087	129	108.11
0.20	456	305	995 449	529 994	162	109.15
0.30	456	216	995 449	289 324	161	114.18
0.40	456	145	995 449	138 175	121	134.11
0.50	456	93	995 449	67 696	84	202.42

Tabelle 7.9: Vorwärtskonstruktion ohne Verzweigungsrestriktion.

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	380	995 449	739 545	101	106.72
0.20	456	309	995 449	521 871	131	107.33
0.30	456	217	995 449	299 520	137	108.99
0.40	456	144	995 449	139 236	108	115.95
0.50	456	92	995 449	64 569	74	149.43

Tabelle 7.10: Vorwärtskonstruktion mit täglicher Verzweigung.

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.10	456	389	995 449	746 613	49	106.53
0.20	456	300	995 449	509 103	57	106.84
0.30	456	228	995 449	310 653	64	107.59
0.40	456	163	995 449	151 809	69	109.78
0.50	456	92	995 449	60 501	46	119.12

Tabelle 7.11: Vorwärtskonstruktion mit wöchentlicher Verzweigung.

(7.2) wie beim Rückwärtsalgorithmus nicht zweckmäßig, wenngleich eine gewisse variable Aufteilung auch sinnvoll sein kann. Im Fall der Vorwärtskonstruktion erfolgte eine lineare Festlegung der individuellen Parameter ε_t durch

$$\varepsilon_t = \frac{\varepsilon}{T} \left[1 + \bar{q} \left(\frac{1}{2} - \frac{t}{T} \right) \right], \quad t = 2, \dots, T, \quad (7.3)$$

wobei der Wert $\bar{q} \in [0, 1]$. Aufgrund der Linearität sind die Auswirkungen auf die Diskretisierungen weniger stark als vorher. Dennoch zeigen numerische Erfahrungen eine hohe Sensibilität bzgl. \bar{q} bei der Vorwärtskonstruktion. In den hier durchgeführten Berechnungen wurde schließlich $\bar{q} = 0.6$ gewählt.

Die Tabellen 7.9 bis 7.11 zeigen die Ergebnisse der Konstruktionen mit Algorithmus 6.5. Wie zuvor beziehen sich die Tabellen auf die verschiedenen Testrei-

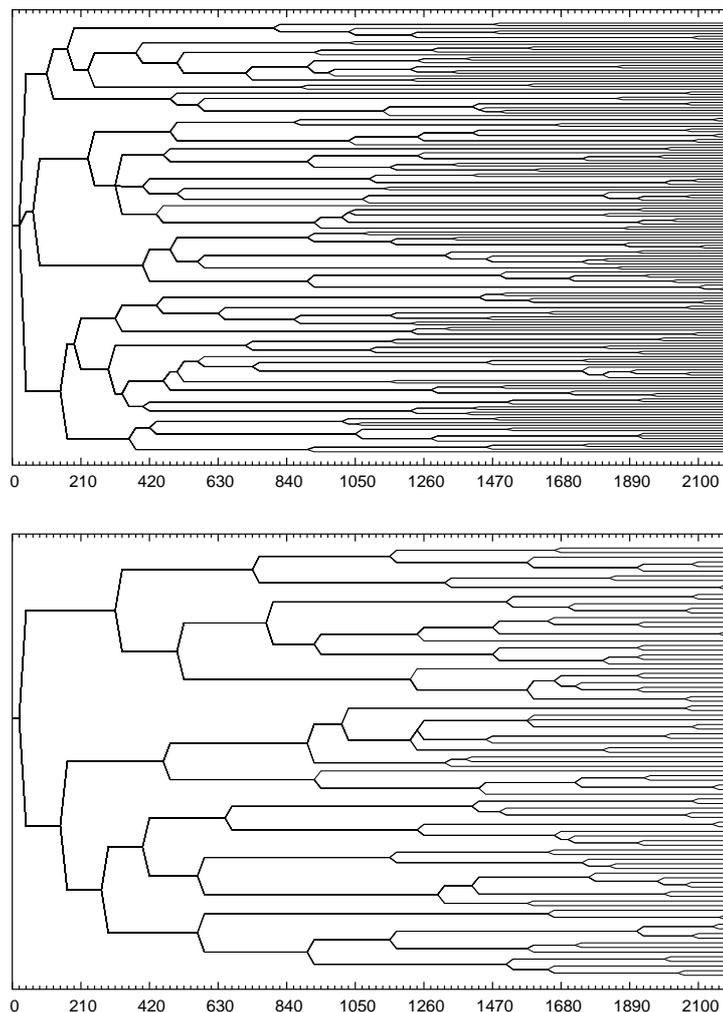


Abbildung 7.5: Mit dem Vorwärtsalgorithmus erzeugte Szenariobäume für Toleranzen $\varepsilon_{rel} = 0.4$ (oben) und $\varepsilon_{rel} = 0.5$ sowie wöchentlicher Verzweigung.

hen a) bis c). Die Ergebnisse weisen gegenüber der Rückwärtsvariante ähnliche Merkmale auf. Dennoch gibt es im direkten Vergleich mit Algorithmus 6.2 auch Unterschiede. Offensichtlich generiert der Vorwärtsalgorithmus bei Vorgabe ähnlicher Parameter (insbesondere bei der Wahl relativer Toleranzen größer als 30%) Bäume mit deutlich geringerer Knotenzahl. Eine Erklärung dafür liefert die Tatsache, dass die Herleitung der generellen Fehlertoleranzen in Kapitel 6.1 im Fall des Vorwärtsalgorithmus, im Gegensatz zum Rückwärtsalgorithmus, ohne Verwendung der Dreiecksungleichung möglich ist (vgl. Beweise der Theoreme 6.1 und 6.4), was sich insbesondere bei wachsender Toleranz deutlicher auswirkt. Schließlich erlaubt Abbildung 7.5 noch einen Blick auf die Struktur der Szenariobäume mit 40 bzw. 50 Prozent gewählter Genauigkeitstoleranz bei erneut wöchentlicher Verzweigung.

7.2.2 Konstruktion von Last/Preis-Szenarien

Das zweite Anwendungsbeispiel für die Szenariobaumkonstruktion kommt ebenfalls aus der Energiewirtschaft. Zur Sicherung ihrer Wettbewerbsfähigkeit nutzen auch kleinere Marktteilnehmer wie Stadtwerke zunehmend die Beschaffungsmöglichkeiten der deutschen Strombörse bzw. die Handelsmöglichkeiten mit derivativen Energieprodukten. Damit setzen sie sich in vollem Umfang dem Preisrisiko durch hochvolatile Spotpreise aus, verfügen jedoch nicht über die personelle und finanzielle Ausstattung großer Handelshäuser zur Abwicklung ihrer Handelsaktivitäten. Daher ist insbesondere die Verwendung eines stochastischen energiewirtschaftlichen Optimierungsmodells, welches alle relevanten Beschaffungsmöglichkeiten der Stadtwerke abbildet und die neu entstandenen Risiken durch die stochastischen Datenprozesse im liberalisierten Strommarkt berücksichtigt, notwendig geworden.

Das in [22] beschriebene Optimierungsmodell ist speziell auf die Bedürfnisse eines deutschen Stadtwerkes zugeschnitten, deren Aufgabe die Versorgung einer mittleren Großstadt mit Strom und Wärme ist. Der Wärmebedarf wird dabei komplett durch eine eigene Kraft-Wärme-Kopplungsanlage gedeckt, während zur Deckung des Bedarfs an elektrischer Energie Stromprodukte externer Anbieter zugekauft werden müssen. Der Planungszeitraum beträgt ein Kalenderjahr und wurde in Einzelstunden zerlegt. Das risikoorientierte Modell basiert auf Spotpreis- und Lastgangsszenarien sowie Szenarien für die Wärmelast, die mit Hilfe von Monte-Carlo-Simulationen auf Basis geeigneter statistischer Modelle generiert wurden und Ausgangspunkt für die Szenariobaumkonstruktion waren.

Zufallsdaten	Unterteilungen	Zeitperioden
Elektrische Last	24 pro Tag	8 760
Spotpreise	24 pro Tag	8 760
Wärmelast	24 pro Tag	8 760

Tabelle 7.12: Unterteilung des Optimierungszeitraums von einem Jahr.

	Anzahl
Szenarien	100
Zeitperioden	8 760
Knoten	875 901

Tabelle 7.13: Dimension der Ausgangsdaten.

Die Tabellen 7.12 und 7.13 zeigen alle wichtigen Parameter der Ausgangsdaten, mit denen die Szenariobäume konstruiert wurden. Die Vorgabe bei der Konstruktion des Baums war dabei die, dass nur zu Beginn eines jeden Monats verzweigt werden durfte. Das heißt, maximal 12 Verzweigungsstufen wurden zugelassen. Die Szenariobäume wurden in dieser Anwendung mit dem Vorwärt-

salgorithmus zur Baumkonstruktion, d.h. mit Algorithmus 6.5, generiert. Wie im vorherigen Beispiel wurden die Parameter zur Steuerung der Konstruktion über eine linearen Abstiegsfunktion gewählt. Zusätzlich wurde hier auch die Abhängigkeit des Modell-Parameters r (vgl. Kapitel 4.1) berücksichtigt, so dass die Beziehung (7.3) die allgemeinere r -abhängige Form

$$\varepsilon_t^r = \frac{\varepsilon^r}{T} \left[1 + \bar{q} \left(\frac{1}{2} - \frac{t}{T} \right) \right], \quad t = 2, \dots, T,$$

annimmt. In diesem Beispiel wurde für \bar{q} der Wert 0.2 gewählt, was zu einer leicht fallenden Folge von Werten ε_t führt. Da die Zufallsdaten mit elektrischer Last und Wärmelast einerseits rechten Seiten des Optimierungsproblems bilden, aufgrund der stochastischen Spotpreise aber andererseits auch die Kosten in der Zielfunktion betreffen, wurde $r = 2$ gemäß Auswahl (4.2) festgelegt. Der Übergang zu relativen Toleranzen erfolgt genau wie bisher, d.h. unter Verwendung von Formel (7.1).

ε_{rel}	Szenarien		Knoten		Stufen	Zeit [s]
	initial	Baum	initial	Baum		
0.20	100	100	875 901	775 992	4	24.53
0.25	100	100	875 901	752 136	5	24.54
0.30	100	100	875 901	719 472	7	24.55
0.35	100	97	875 901	676 416	8	24.61
0.40	100	98	875 901	645 672	10	24.64
0.45	100	96	875 901	598 704	10	24.75
0.50	100	95	875 901	565 800	9	24.74
0.55	100	88	875 901	452 184	10	24.75
0.60	100	87	875 901	337 728	11	25.89

Tabelle 7.14: Vorwärtskonstruktion mit monatlicher Verzweigungsstruktur.

Tabelle 7.14 zeigt die Resultate einer Reihe von Berechnungen mit dem Vorwärtsalgorithmus zur Baumkonstruktion bzgl. unterschiedlicher relativer Genauigkeiten. Die Rechnungen spiegeln im Wesentlichen die Ergebnisse des vorherigen Abschnitts 7.2.1 wider. Einer der erwarteten Effekte ist: Je größer die relativen Toleranzen gewählt werden desto mehr verringert sich die Anzahl der Knoten in den konstruierten Szenariobäumen. Gleichzeitig vergrößert sich die Anzahl der Verzweigungsstufen. Der Algorithmus benötigt für die Berechnung rund 25 Sekunden, wobei der wesentliche Aufwand erneut bei der Berechnung der Szenarienabstände liegt. Aufgrund der geringeren Anzahl von Szenarien liegen der Rechenzeiten damit insgesamt deutlich unter denen für die Berechnung der Last-/Zufluss-Szenariobäume für EDF (Tabellen 7.9–7.11).

Die Bilder in Abbildung 7.6 zeigen die Struktur zweier generierter Szenariobäume mit unterschiedlicher Verzweigungsdichte und Anzahl von Knoten.

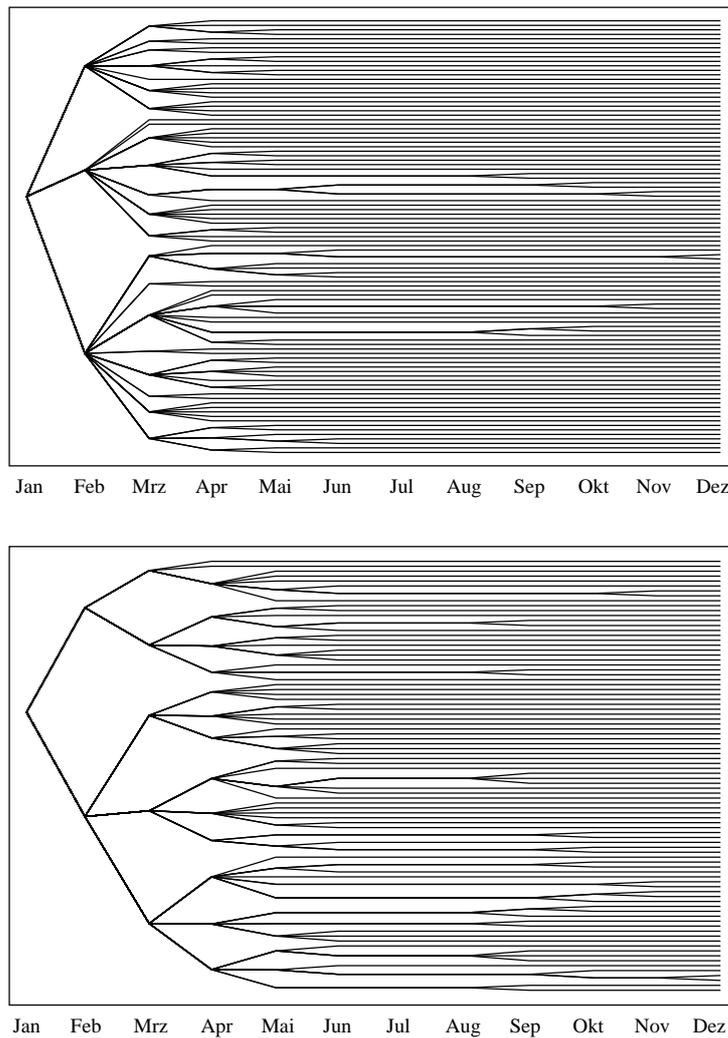


Abbildung 7.6: Mit dem Vorwärtsalgorithmus erzeugte Szenariobäume für $\varepsilon_{rel} = 0.40$ (oben) und $\varepsilon_{rel} = 0.55$ sowie monatlicher Verzweigung.

Der erste Szenariobaum umfasst noch insgesamt 98 Szenarien bei 645 672 Knoten, der zweite nur noch 88 Szenarien bei insgesamt 452 184 Knoten. Eine Beobachtung ist die, dass in sämtlichen Rechnungen, d.h. auch bei starker Reduktion, die Anzahl 12 der möglichen Verzweigungen nicht ausgeschöpft wird. Insbesondere in den Sommermonaten ist der Verzweigungsgrad sehr gering. Dies lässt sich möglicherweise auf die stark reduzierte und gleichbleibend geringe Wärmelast in den Sommermonaten zurückführen. Dagegen sind sowohl die Schwankungen als auch der Betrag in den ersten Monaten des Jahres sehr hoch, was die starke Verzweigung der Szenariobäume in diesem Zeitraum erklärt. Ein Jahresprofil der drei Komponenten ist in Abbildung 7.7 dargestellt.

Ein weiterer Effekt resultiert aus der Wahl des Parameters r . Wenn auch die qualitativen Ergebnisse in beiden Anwendungen aus dem Energiebereich

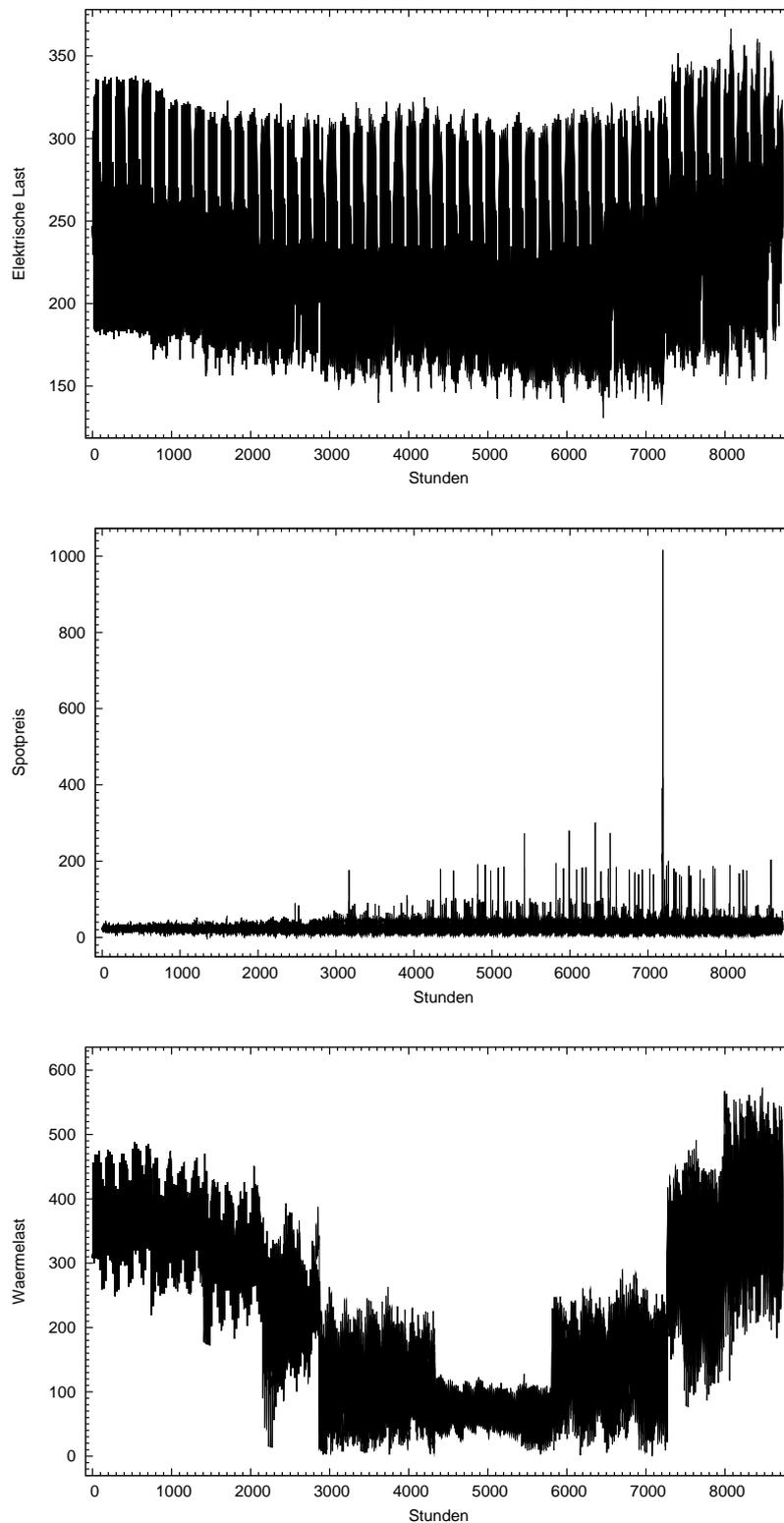


Abbildung 7.7: Komponentenprofile des trivariaten Szenariobaumprozesses für elektrische Last (oben), Spotpreise (Mitte) und Wärmelast (unten).

übereinstimmen, so wirkt sich die Verwendung von $r = 2$ gegenüber $r = 1$ auf die Rechnungen wie folgt aus. Vergleicht man die (prozentuale) Verringerung der Anzahl von Knoten bei der Wahl von $r = 1$ bzw. $r = 2$ bei jeweils gleicher relativer Genauigkeit ε_{rel} (siehe zum Beispiel Tabelle 7.11 für $r = 1$ und Tabelle 7.14 für $r = 2$), so erkennt man eine scheinbare Verbesserung des Verhältnis, d.h. einen höheren Faktor der Reduktion, bei Verwendung von $r = 1$. Die Ursache ist eine Verschiebung des Verhältnis zwischen ε und ε_{max} mit wachsendem r bei der Betrachtung relativer Toleranzen in (7.1). Das liegt darin begründet, dass der berechnete Wert für ε_{max} in Abhängigkeit von r , d.h. bei Verwendung des L_r -Abstands, mit wachsendem r immer kleiner wird. Der für die Stabilitätsaussage in Theorem 4.1 relevante absolute L_r -Abstand wird von dem ungünstigeren Verhältnis jedoch nicht beeinflusst.

Eine andere Konsequenz des Parameters r resultiert aus folgender Überlegung. Je größer das r gewählt wird, desto weniger spielt die Summe der individuellen Werte ε_t eine Rolle. Vielmehr wird das Maximum dieser Werte bei wachsendem r für die Gesamtgenauigkeit ausschlaggebend. Die Wahl des Wertes r beeinflusst damit wie bei der Szenarioreduktion im Allgemeinen auch die Auswahl von Szenarien innerhalb des Konstruktionsprozesses.

7.2.3 Szenariobäume für Buchungsklassen

Das abschließende Beispiel kommt aus einer Anwendung im Bereich Airline-Revenue-Management, d.h. der Optimierung der Ertragssteuerung von Flugbuchungsanfragen in Flugnetzen einer größeren Fluggesellschaft. Die Ertragssteuerung (Revenue-Management) setzt sich im Wesentlichen aus den Instrumenten Preisdifferenzierung (einschließlich Gestaltung der Vertragsbedingungen), Überbuchung und Preis-Mengen-Steuerung zusammen. Die Modellierung in [60] besteht in der Berechnung von optimalen Buchungskontingenten (Protection-Level) für die einzelnen Preis- oder Buchungsklassen aus denen verschachtelte Buchungsschranken (Nested-Booking-Limits) für das Buchungssystem abgeleitet werden. Die besondere Komplexität der Aufgabenstellung besteht u.a. in der Betrachtung eines gesamten Flugnetzes einer Fluggesellschaft (zum Beispiel für einen Tag) und den Unsicherheiten bei der Nachfrage, d.h. dem stochastischen Charakter der Buchungen und Stornierungen. Das Flugnetz besteht aus Reisewegen, die neben dem Start- und Zielflughafen sowie sämtlicher Zwischenlandungen auch Datum und Zeit des Abflugs vom Startflughafen beinhalten. Jede der Teilstrecken (Legs) eines Reiseweges kann dabei verschiedenen Reisewegen zugeordnet sein und selbst wiederum in Beförderungsklassen (Compartments) unterteilt werden.

Der Buchungszeitraum selbst wird in mehrere Buchungsintervalle unterteilt. Die Beschreibung der Buchungsnachfragen und der Stornierungen erfolgt entsprechend der Buchungsintervalle über Szenarien. Jede Komponente enthält dabei einen kompletten Datensatz für die Buchungsnachfrage und Stornierun-

gen aller Reisewege, Buchungsklassen und darüber hinaus für verschiedene Verkaufsregionen. Das Ergebnis ist ein hochgradig multivariater stochastischer Prozess in Form eines Szenariobaums. Die Entscheidungsvariablen des Optimierungsproblems setzen sich aus den jeweiligen Buchungskontingenten für das gesamte Flugnetz, den Buchungsklassen und Verkaufsregionen innerhalb der Buchungsintervalle zusammen. Die (deterministischen) Entscheidungen für das erste Buchungsintervall werden für den tatsächlichen Buchungsprozess verwendet. Für die folgenden Buchungsintervalle $t = 2, \dots, T$ (Data-Collection-Points, DCP's) wird auf dem bisherigen Buchungsverlauf basierend, schrittweise eine Re-Optimierung mit neu generierten Szenarien durchgeführt.

Numerische Ergebnisse

Für das betrachtete Optimierungsmodell zur Maximierung des Ertrags wurde in [60] ein zunächst vergleichsweise kleines Flugnetz herangezogen. Zur Lösung des Optimierungsproblems wurden Szenariobäume mit dem Verfahren von Algorithmus 6.5 mit unterschiedlicher Größe erzeugt. Tabelle 7.15 zeigt die Cha-

Reise- wege	Buchungs- klassen	Vekaufs- regionen	Teil- strecken	Beförderungs- klassen	Buchungs- intervalle
12	6	1	6	2	14

Tabelle 7.15: Dimension des Flugnetzes.

DCP	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Tage	182	126	84	56	35	21	14	10	7	5	3	2	1	0

Tabelle 7.16: Buchungsintervalle (DCP) in Abhängigkeit der Tage vor Abflug.

rakteristika des betrachteten Flugnetzes. Es besteht aus 6 Teilstrecken und 12 möglichen Reisewegen. Betrachtet werden verschiedene Buchungs- und Beförderungsklassen sowie Buchungsintervalle. In Tabelle 7.16 ist für jedes einzelne Buchungsintervall die Ablauffrist vor einem Abflug in Tagen zu sehen. Die kürzeren Buchungsintervalle zum Ende des Buchungszeitraum sind dadurch motiviert, dass zum einen die Information über den bisherigen Buchungsverlauf zunimmt und zum anderen die höherwertige Nachfrage zum Ende des Buchungszeitraums auftritt.

Der Buchungsprozess kann als nichthomogener Poissonprozess modelliert werden, der sich durch Samplings der Gamma-Verteilung erzeugen lässt (mehr Details sind in [60] zu finden). Auf diese Weise wurden für das Modell Eingangsszenarien mit 100, 200, 300 sowie 400 Szenarien bestimmt, die daraufhin der Szenariobaumkonstruktion mit Algorithmus 6.5 unterzogen wurden. Abschließend wurde das gemischt ganzzahlige Optimierungsmodell (unter Verwendung

von CPLEX) gelöst. Die Ergebnisse der Rechnungen sind in Tabelle 7.17 zu sehen. Besonders auffällig ist der beachtliche Zuwachs an Rechenzeit bei der

Beispiel	Sampling Szenarien	Szenariobaum Knoten	Szenarien	Optimalwert (Ertrag)	Rechenzeit [h:min:s]
1	100	1 018	92	210 393	0:03:29.8
2	100	1 022	93	210 128	0:03:38.4
3	100	1 016	90	210 271	0:02:44.1
4	200	2 006	181	210 001	0:10:03.4
5	200	2 017	183	214 477	0:11:49.9
6	200	2 025	183	210 485	0:11:29.6
7	300	3 020	272	210 517	0:28:41.8
8	300	2 998	271	210 436	0:24:18.6
9	300	3 008	273	210 327	0:29:58.6
10	400	4 007	364	209 916	1:05:31.3
11	400	4 045	367	210 108	3:23:55.7
12	400	4 018	363	210 470	0:39:42.3

Tabelle 7.17: Resultate der Optimierung für verschiedene Samplings.

Verwendung von größeren Szenariobäumen mit mehr Szenarien und vor allem mehr Knoten. So ergibt sich eine Rechenzeit von über 3 Stunden im Beispiel 11. Die starke Abhängigkeit der Rechenzeiten von der Anzahl der Knoten ist in erster Linie auf den gemischt-ganzzahligen Charakter des Optimierungsmodells zurückzuführen. Andererseits zeigen die Resultate hinsichtlich des Optimalwertes keine signifikanten Unterschiede bei Verwendung von deutlich mehr Szenarien. Die Verwendung von 100 Ausgangsszenarien liefert bereits hinreichend gute Resultate in vergleichsweise kurzer Rechenzeit. Das unterstreicht die Bedeutung der Szenariobaumkonstruktion auch unter dem Gesichtspunkt der Reduktion der Dimension des Optimierungsproblems und der damit verbundenen numerischen Behandelbarkeit von mehrstufigen stochastischen Optimierungsproblemen.

DCP	2	3	4	5	6	7	8
$\varepsilon_{t,rel}$	0.107	0.069	0.045	0.029	0.019	0.012	0.008

DCP	9	10	11	12	13	14
$\varepsilon_{t,rel}$	0.005	0.003	0.002	0.001	0.001	0.001

Tabelle 7.18: Wahl der relativen Toleranzen zur Szenariobaumkonstruktion.

Bei der Konstruktion der Szenariobäume wurde aufgrund der unterschiedlichen Länge der Buchungsintervalle eine Anpassung der individuellen Toleranzparameter ε_t vorgenommen. Entsprechend der über den Buchungszeitraum

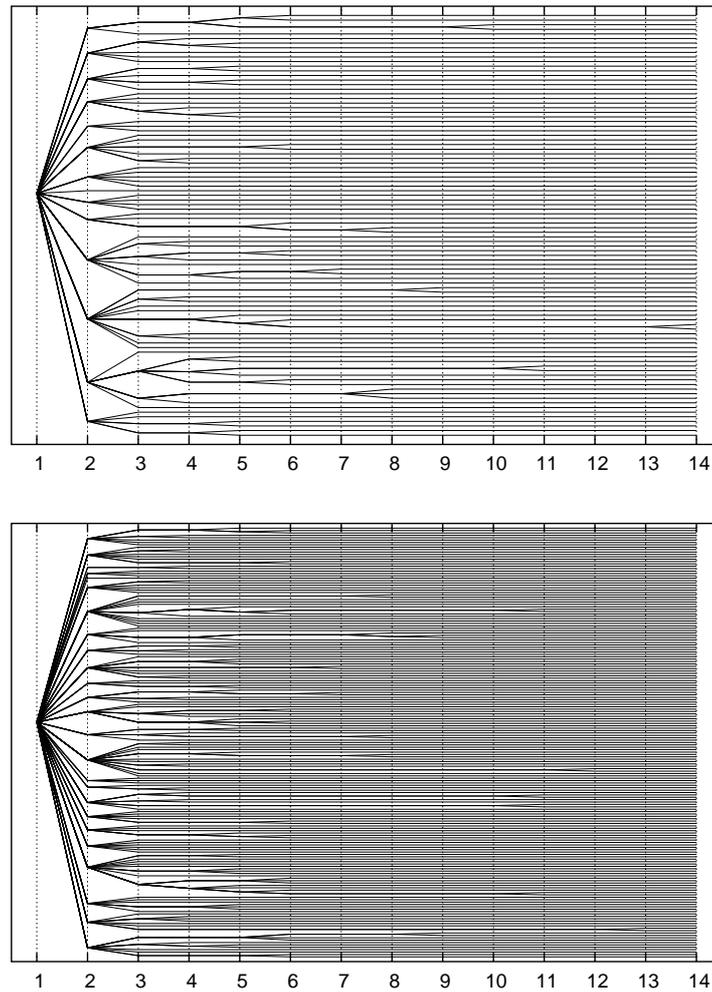


Abbildung 7.8: Mit dem Vorwärtsalgorithmus erzeugte Szenariobäume für 100 (oben) und 200 Ausgangsszenarien bei Verwendung von $\varepsilon_{rel} = 0.3$ (Szenariobäume für Beispiel 1 bzw. 4 in Tabelle 7.17).

kürzer werdenden Buchungsintervalle wurden individuelle relative Toleranzen gemäß Tabelle 7.18 festgelegt. Exemplarisch sind die für Beispiel 1 und 4 (siehe Tabelle 7.17) konstruierten Szenariobäume in Abbildung 7.8 dargestellt.

7.2.4 Schlussbemerkung

Sowohl der Vorwärtsalgorithmus 6.5 als auch der Rückwärtsalgorithmus 6.2 zur Konstruktion von Szenariobäumen für mehrstufige stochastische Optimierungsprobleme erlauben über die Steuerungsparameter ε_t ($t = 2, \dots, T$) eine gezielte Einflussnahme auf Form und Größe der Szenariobäume. Die Ergebnisse der Anwendungsbeispiele spiegeln die Flexibilität der Algorithmen bei der Generierung unterschiedlicher Strukturen wider.

Eine wichtige Bemerkung in diesem Zusammenhang betrifft noch einmal die Konvergenzresultate des letzten Kapitels. Vor dem Hintergrund dieser Resultate sollte die Wahl der Parameter ε_t erfolgen. Eine sehr hohe Genauigkeit durch sehr kleine Werte mag in Bezug auf den Verteilungsabstand (siehe Proposition 6.3 bzw. 6.6) günstig sein, muss aber nicht zwangsläufig zu besseren Ergebnissen hinsichtlich der Approximation des Optimierungsproblems führen. Im Gegenteil, sind die Ausgangsszenarien fest vorgegeben und weisen sie formal keine Baumstruktur auf, würde eine zu hohe Genauigkeit, d.h. eine zu geringe ε -Toleranz, keinen Sinn bezüglich des Informationsabstands ergeben. Das Resultat wären Szenarien die wenig oder garnicht verzweigen. Geht man von einem wirklich mehrstufigen Ausgangsproblem aus, d.h. die Filtration des Originalproblems ist nicht trivial, wäre der Filtrationsabstand dann sogar besonders groß. Das heißt, für ein endliches Sampling von Szenarien führt eine beliebige Verfeinerung der Diskretisierung bei gleichzeitig festgehaltener Szenarienzahl im Allgemeinen zur Divergenz des Filtrationsabstands, wenn (wie bei stetigen Prozessen üblicherweise) die gesampelte Szenarien wirklich individuell sind. Damit kann die Konvergenz der Optimalwerte aufgrund von Beispiel 4.1.4 nicht länger erwartet werden. Das steht allerdings nicht im Widerspruch zum generellen Konvergenzresultat, bedeutet aber, dass die Grenzwerte von (6.29) im Konvergenztheorem 6.12 im Allgemeinen nicht vertauschbar sind.

Will man das Resultat von Theorem 6.12 also richtig anwenden, muss man Diskretisierungsgenauigkeit und Anzahl der Szenarien immer im Zusammenhang sehen. Entscheidet man sich für eine Genauigkeit, d.h. innerhalb der Algorithmen für gewisse ε -Toleranzen, muss man gleichzeitig die Bereitstellung einer hinreichend großen Zahl von Szenarien gewährleisten können. Wie groß diese Zahl sein muss, hängt auch von gewissen Regularitätseigenschaften des Eingangsdatenprozesses ab. Ist auf der anderen Seite die Zahl der Szenarien von vornherein stark beschränkt, sollten die ε -Toleranzen innerhalb der Konstruktionsalgorithmen nicht zu klein gewählt werden, um dennoch brauchbare Resultate erzielen zu können.

Literaturverzeichnis

- [1] Barty, K.: Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique, Thèse de Doctorat, École Nationale des Ponts et Chaussées, 2004.
- [2] Birge, J.R.; Louveaux, F.: Introduction to Stochastic Programming, Springer, New York, 1997.
- [3] Boylan, E.S.: Equiconvergence of martingales, *The Annals of Mathematical Statistics* 42 (1971), 552–559.
- [4] Brockwell, P.J.; Davies, R.A.: Introduction to Time Series and Forecasting, Springer, New York 1996.
- [5] Burger, M.; Klar, B.; Müller, A.; Schindlmayr, G.: A spot market model for pricing derivatives in electricity markets, in *Journal of Quantitative Finance* 4 (2004), 109–122.
- [6] Casey, M.; Sen, S.: The scenario generation algorithm for multistage stochastic linear programming, in *Mathematics of Operations Research* 30 (2005), 615–631.
- [7] Charikar, M.; Guha, S.; Tardos, E.; Shmoys, D. B.: A constant-factor approximation algorithm for the k -median problem, *Journ. Comp. Syst. Sciences* 65 (2002), 129–149.
- [8] Chiralaksanakul, A.; Morton, D.P.: Assessing policy quality in multi-stage stochastic programming, *Stochastic Programming E-Print Series* 12–2004 (<speps.org>).
- [9] Consigli, G.; Dempster, M.A.H.: Dynamic stochastic programming for dynamic asset-liability management, *Annals of Operations Research* 81 (1998), 131–161.
- [10] Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.: Introduction to Algorithms, The Massachusetts Institute of Technology, 1990.

- [11] Corvera Poiré, X.: Model Generation and Sampling Algorithms for Dynamic Stochastic Programming, PhD Thesis, Department of Mathematics, University of Essex, 1995.
- [12] Cotter, K.D., Convergence of information, random variables and noise, *Journal of Mathematical Economics* 16 (1987), 39–51.
- [13] Dantzig, G.B.: Linear programming under uncertainty, *Management Science* 1 (1955), 197–206.
- [14] Dantzig, G.B.: Linear Programming and Extensions, Princeton University Press, Princeton 1963.
- [15] Dentcheva, D.; Henrion, R.; Ruszczyński, A.: Stability and Sensitivity of Optimization Problems with First Order Stochastic Dominance Constraints, *SIAM Journal on Optimization* 18 (2007), 322–337.
- [16] Dempster, M.A.H.: Sequential importance sampling algorithms for dynamic stochastic programming, *Zap. Nauchn. Semin. POMI* 312 (2004), 94–129.
- [17] Dudley, R.M.: Real Analysis and Probability, Chapman & Hall, New York 1989.
- [18] Dupačová, J.; Consigli, G.; Wallace, S.W.: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.
- [19] Dupačová, J.; Gröwe-Kuska, N.; Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming Ser. A* 95 (2003), 493–511.
- [20] Edirisinghe, N.C.P.: Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation, *Annals of Operations Research* 85 (1999), 103–127.
- [21] Eichhorn, A.; Römisch, W.: Polyhedral risk measures in stochastic programming, in *SIAM Journal on Optimization* 16 (2005), 69–95.
- [22] Eichhorn, A.; Römisch, W.; Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, *IEEE St. Petersburg Power Tech* 2005.
- [23] Evstigneev, I.: Measurable selection and dynamic programming, *Mathematics of Operations Research* 1 (1976) 267–272.
- [24] Fabian, M.; Habala, P.; Hájek, P.; Montesinos Santalucia, V.; Pelant, J.; Zizler, V.: Functional Analysis and Infinite-Dimensional Geometry, CMS Books in Mathematics, Springer, New York 2001.

- [25] Fetter, H.: On the continuity of conditional expectations, *Journal of Mathematical Analysis and Applications* 61 (1977), 227–231.
- [26] Fleten, S.-E.; Wallace, S.W.; Ziemba, W.T.: Hedging electricity portfolios via stochastic programming, *Decision Making under Uncertainty: Energy and Power* (C. Greengard and A. Ruszczyński Eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer, New York 2002, 71–93.
- [27] Fortet, R.; Mourier, E.: Convergence de la répartition empirique vers la répartition théorique, *Ann. Sci. Ecole Norm. Sup.* 70 (1953), 266–285.
- [28] Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming Ser. B*, 75 (1996), 277–293.
- [29] Gaivoronski, A.; Pflug, G.Ch.: Value-at-Risk in portfolio optimization: properties and conditional approach, in *Journal of Risk* 7 (2005), 1–31.
- [30] Givens, C.R.; Shortt, R.M.: A class of Wasserstein metrics for probability distributions, *Michigan Mathematical Journal* 31 (1984), 231–240.
- [31] Graf, S.; Luschgy, H.: Foundations of Quantization for Probability Distributions, *Lecture Notes in Mathematics* Vol. 1730, Springer, New York 2000.
- [32] Gröwe-Kuska, N.; Heitsch, H.; Römisch, W.: Modellierung stochastischer Datenprozesse für Optimierungsmodelle der Energiewirtschaft, *IT-Lösungen für die Energiewirtschaft in liberalisierten Märkten*, VDI-Berichte 1647, VDI-Verlag, Düsseldorf 2001, 69–78.
- [33] Gröwe-Kuska, N.; Heitsch, H.; Römisch, W.: Scenario reduction and scenario tree construction for power management problems, *IEEE Bologna Power Tech Proceedings* (Borghetti, A., Nucci, C. A., Paolone, M. eds.), 2003.
- [34] Gröwe-Kuska, N.; Kiwiel, K. C.; Nowak, M. P.; Römisch, W.; Wegner, I.: Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation, in: *Decision Making under Uncertainty: Energy and Power* (C. Greengard and A. Ruszczyński Eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer-Verlag, New York 2002, 39–70.
- [35] Gröwe-Kuska, N.; Römisch, W.: Stochastic unit commitment in hydro-thermal power production planning, Chapter 30 in *Applications of Stochastic Programming* (Wallace, S. W., Ziemba W. T., eds.), MPS-SIAM Series in Optimization, 2005.

- [36] Heitsch, H.; Römisch, W.: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.
- [37] Heitsch, H.; Römisch, W.: Generation of multivariate scenario trees to model stochasticity in power management, *IEEE St. Petersburg Power Tech* 2005.
- [38] Heitsch, H.; Römisch, W.: Scenario tree modelling for multistage stochastic programs, *Mathematical Programming*, erscheint 2008.
- [39] Heitsch, H.; Römisch, W.: Stability and scenario trees for multistage stochastic programs, Preprint 324, DFG Research Center MATHEON "Mathematics for key technologies", 2006 und eingereicht in *Stochastic Programming - The State of the Art* (G. Dantzig, G. Infanger eds.).
- [40] Heitsch, H.; Römisch, W.: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.
- [41] Heitsch, H.; Römisch, W.; Strugarek, C.: Stability of multistage stochastic programs, *SIAM Journal on Optimization* 17 (2006), 511–525.
- [42] Henrion, R.: Some remarks on value-at-risk optimization, *Management Science and Engineering Management* 1 (2006), 111–118.
- [43] Henrion, R.; Küchler, C; Römisch, W.: Scenario reduction in stochastic programming with respect to discrepancy distances, in *Stochastic Programming E-Print Series (SPEPS)* 13 (2006), erscheint in *Computational Optimization and Applications*.
- [44] Henrion, R.; Römisch, W.: Hölder and Lipschitz Stability of Solution Sets in Programs with Probabilistic Constraints, *Mathematical Programming* 100 (2004), 589–611.
- [45] Higle, J.L.; Rayco, B.; Sen, S.: Stochastic scenario decomposition for multistage stochastic programs, eingereicht in *Annals of Operations Research*.
- [46] Hochreiter, R.: Computational Optimal Management Decisions – The case of Stochastic Programming for Financial Management, Ph.D. Thesis, University of Vienna, 2005.
- [47] Hochreiter, R.; Pflug, G. Ch.: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* 152 (2007), 257–272.
- [48] Hochreiter, R.; Pflug, G. Ch.; Wozabal, D.: Multi-stage stochastic electricity portfolio optimization, in liberalized energy markets, *System Modeling and Optimization* 199 (2006), 219–226.

- [49] Hoover, D.N.: Convergence in distribution and Skorokhod convergence for the general theory of processes, *Probability Theory and Related Fields* 89 (1991), 239-259.
- [50] Høyland, K.; Wallace, S. W.: Generating scenario trees for multi-stage decision problems, *Management Science* 47 (2001), 295–307.
- [51] Høyland, K.; Kaut, M.; Wallace, S.W.: A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* 24 (2003), 169–185.
- [52] Infanger, G.: Planning under Uncertainty – Solving Large-Scale Stochastic Linear Programs, Boyd& Fraser Publ. Comp., 1994.
- [53] Kall, P.; Mayer, J.: Stochastic Linear Programming, Springer, 2005.
- [54] Kall, P.; Wallace, W.: Stochastic Programming, Wiley, New York, 1994.
- [55] Kaut, M.; Wallace, S.W.: Evaluation of scenario-generation methods for stochastic programming, *Stochastic Programming E-Print Series* 14-2003 (<speps.org>).
- [56] Krasenbrink, B.: Integrierte Jahresplanung von Elektrizitätserzeugung und -handel, ABEV Band 81, Klinkenberg Verlag, Achen, 2002.
- [57] Kuhn, D.: Generalized bounds for convex multistage stochastic programs, *Lecture Notes in Economics and Mathematical Systems* Vol. 548, Springer, Berlin, 2005.
- [58] Magnanti, T.L., Wolsey, L.A.: Optimal Trees, *Network Models* (Nemhauser, G.L., Rinnooy Kan, A.H.G. eds.), Handbooks in Operations Research and Management Science, Volume 7, Elsevier, Amsterdam 1995, 503–616.
- [59] Möller, A.; Römisch, W.; Weber, K.: A new approach to O&D revenue management based on scenario trees, *Journal of Revenue and Pricing Management* 3 (2004), 265–276.
- [60] Möller, A.; Römisch, W.; Weber, K.: Airline network revenue management by multistage stochastic programming, in *Computational Management Science*, erscheint.
- [61] Neveu, J.: Note on the tightness of the metric on the set of complete sub σ -algebras of a probability space, *The Annals of Mathematical Statistics*, 43 (1972), 1369–1371.
- [62] Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods, *CMBS-NSF Regional Conference Series in Applied Mathematics* Vol. 63, SIAM, Philadelphia 1992.

- [63] Olsen, P.: Multistage stochastic programming with recourse as mathematical programming in an L_p -space, *SIAM Journal Control and Optimization* 14 (1976), 528-537.
- [64] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs, *Mathematics of Operations Research* 30 (2005), 245–256.
- [65] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Mathematical Programming* Ser. B (erscheint).
- [66] Pflug, G. Ch.: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* 89 (2001), 251–271.
- [67] Pflug, G. Ch.: Stochastic Optimization and statistical inference, in: *Stochastic Programming* (Ruszczynski, A., Shapiro, A. eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 427–482.
- [68] Pflug, G. Ch.: A value-of-information approach to measuring risk in multiperiod economic activity, in *Journal of Banking & Finance* 30 (2006), 695–715.
- [69] Rachev, S. T.: *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.
- [70] Rachev, S. T.; Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.
- [71] Rachev, S. T.; Rüschendorf, L.: *Mass Transportation Problems, Vol. I and II*, Springer, Berlin 1998.
- [72] Rachev, S. T.; Schief, A.: On L_p -minimal metrics, *Probability and Mathematical Statistics* 13 (1992), 311–320.
- [73] Rachev, S. T.; Shortt, R.-M.: Duality theorems for Kantorovich-Rubinstein and Wasserstein functionals, *Dissertationes Mathematicae* 299 (1990).
- [74] Robinson, S. M.: Some continuity properties of polyhedral multifunctions, *Mathematical Programming Study* 14 (1981), 206–214.
- [75] Rockafellar, R. T.: Integral functionals, normal integrands and measurable selections, in: *Nonlinear Operators and the Calculus of Variations* (J. P Gossez et al. eds.), Lecture Notes in Mathematics, Vol. 543, Springer, Berlin 1976, 157–207.

- [76] Rockafellar, R. T.; Wets, R. J-B: Nonanticipativity and \mathcal{L}^1 -martingales in stochastic optimization problems, *Mathematical Programming Study* 6 (1976), 170-187.
- [77] Rockafellar, R. T.; Wets, R. J-B: Variational Analysis, Springer-Verlag, Berlin, 1998.
- [78] Römisch, W.: Stability of stochastic programming problems, in: *Stochastic Programming* (Ruszczynski, A., Shapiro, A. eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.
- [79] Römisch, W.; Schultz, R.: Stability analysis for stochastic programs, *Annals of Operations Research* 30 (1991), 241–266.
- [80] Römisch, W.; Schultz, R: Multistage stochastic integer programs: An introduction, *Online Optimization of Large Scale Systems* (M. Grötschel, S.O. Krumke, J. Rambau eds.), Springer-Verlag, Berlin 2001, 579—598.
- [81] Römisch, W.; Wets, R. J-B: Stability of ε -approximate solutions to convex stochastic programs, *SIAM Journal on Optimization* 18 (2007), 961–979.
- [82] Rogge, L.: Uniform inequalities for conditional expectations, *Annals of Probability* 2 (1974), 486–489.
- [83] Ruszczyński, A.; Shapiro, A. (Eds.): Stochastic Programming, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.
- [84] Schultz, R.: Rates of convergence in stochastic programs with complete integer recourse, *SIAM Journal on Optimization* 6 (1996), 1138–1152.
- [85] Sen, S.; Yu, L.; Genc, T.; A stochastic programming approach to power portfolio optimization, *Operations Research* 54 (2006), 55–72.
- [86] Schmöller, H.K.; Modellierung von Unsicherheiten bei der mittelfristigen Stromerzeugungs- und Handelsplanung, Aachener Beiträge zur Energieversorgung, Band 103, Klinkenberg Verlag 2005.
- [87] Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems, *Math. Meth. Oper. Res.* 58 (2003), 57–68.
- [88] Shapiro, A.: On complexity of multistage stochastic programs, *Operations Research Letters* 34 (2006), 1–8.
- [89] Shapiro, A.: Stochastic programming approach to optimization under uncertainty, *Mathematical Programming* 112 (2007), 183–220.

- [90] Stern, B.: Kraftwerkseinsatz und Stromhandel unter Berücksichtigung von Planungsunsicherheiten, Dissertation IAEW; RWTH Aachen, Aachener Beiträge zur Energieversorgung, Bd 78, Aachen 2001.
- [91] Vazirani, V.V.: Approximation Algorithms, Springer, Berlin 2001.
- [92] Walkup, D.; Wets, R.J-B: Lifting projections of convex polyhedra, *Pacific Journal of Mathematics* 28 (1969), 465–475.
- [93] Wallace, S.W.; Ziemba, W.T. (Eds.): Applications of Stochastic Programming, MPS-SIAM Series in Optimization, 2005.